

Received 24 July 2022, accepted 15 August 2022, date of publication 18 August 2022, date of current version 29 August 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3199753

RESEARCH ARTICLE

Moving Pedestrian Localization and Detection With Guided Filtering

KAHLIL MUCHTAR^{1,2}, (Senior Member, IEEE), **AL BAHRI**^{1,2}, (Member, IEEE),
MAYA FITRIA^{1,2}, (Member, IEEE), **TJENG WAWAN CENGGORO**^{3,5}, (Member, IEEE),
BENS PARDAMEAN^{4,5}, (Member, IEEE), **ADHIGUNA MAHENDRA**^{6,7},
MUHAMMAD RIZKY MUNGgaran⁶, AND **CHIH-YANG LIN**⁸, (Senior Member, IEEE)

¹Department of Electrical and Computer Engineering, Universitas Syiah Kuala, Banda Aceh, Aceh 23111, Indonesia

²Telematics Research Center (TRC), Universitas Syiah Kuala, Banda Aceh, Aceh 23111, Indonesia

³Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta 11480, Indonesia

⁴Computer Science Department, BINUS Graduate Program-Master of Computer Science Program, Bina Nusantara University, Jakarta 11480, Indonesia

⁵Bioinformatics and Data Science Research Center, Bina Nusantara University, Jakarta 11480, Indonesia

⁶Nodeflux, Jakarta 12730, Indonesia

⁷Faculty of Engineering and Information Technology, Swiss German University, Tangerang 15143, Indonesia

⁸Department of Electrical Engineering, Yuan Ze University, Taoyuan City 32003, Taiwan

Corresponding author: Kahlil Muchtar (kahlil@unsyiah.ac.id)


This work was supported by the Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia under the 2022 Applied Research Grant (Hibah Penelitian Terapan Kompetitif Nasional) 98/UN11.2.1/PT.01.03/DPRM/2022.

ABSTRACT Detecting a moving pedestrian is still a challenging task in a smart surveillance system due to dynamic scenes. Locating and detecting the moving pedestrian simultaneously influences the development of an integrated but low-resource smart surveillance system. This paper proposes a novel approach to locating and detecting moving pedestrians in a video. Our proposed method first locates the region of interest (ROI) using a background subtraction algorithm based on guided filtering. This novel background subtraction algorithm allows our method to also filter unexpected noises at the same time, which could benefit the performance of our proposed method. Subsequently, the pedestrians are detected using YOLOv2, YOLOv3, and YOLOv4 within the provided ROI. Our proposed method resulted in more processing frames per second compared with previous approaches. Our experiments showed that the proposed method has a competitive performance in the CDNET2014 dataset with a fast-processing time. It costs around ~50 fps in CPU to classify moving pedestrians and maintain a highly accurate rate. Due to its fast processing, the proposed approach is suitable for IoT or smart surveillance device which has limited resource.

INDEX TERMS Moving object analysis, pedestrian localization and detection, convolutional neural network (CNN), integrated surveillance system, YOLO.

I. INTRODUCTION

Intelligent video surveillance systems currently play a crucial role in monitoring and evaluating human activity in public areas. This is especially true for pedestrian detection systems, which have been one of the most common subjects in various areas over the last decade. However, developing a robust pedestrian detection system is challenging due to many aspects such as illumination, cluttered background, and variations in pedestrian sizes.

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

Specifically, to address the background challenge, one of the promising approaches is integrating a robust background (BG) model into the pedestrian detection system. In developing the BG model, it is important to notice Stauffer and Grimson's work [1], which largely influenced the field. The Stauffer-Grimson BG model is based on a Gaussian Mixture Model (GMM) that was fitted to the pixel values distribution over time. After the fitting process, the model is able to decide if the incoming pixels belong to the BG, based on the probability of the MoG identified at each pixel location. The BG model in their work has inspired the development of many variants of BG models, such as the texture-based

approach [2], improved GMM [3], [4], and other novel approaches [5], [6], [7], [8]. Despite its popularity, the GMM-based model like the Stauffer-Grimson model still has a problem in dealing with a noisy background. Fig. 1 depicts the problem that is caused by shadow as the noise in the background. The changes in illumination can also introduce noises that lead to undesired results [9].



FIGURE 1. Illustration of GMM drawbacks in CDNET 2014 – pedestrians' dataset [10].

Ultimately, the ROI provided by the BG subtraction algorithm is required for building a complete pedestrian detection system. The recent trend suggests the use of a deep-learning-based detection algorithm for pedestrian detection [11], [12], [13], [14], although other types of computer vision techniques can also be employed [15], [16], [17]. Because of the sequential integration between the BG model and the detection system, the problem caused by the BG model can affect the overall performance of the pedestrian detection system. Thus, the aforementioned problem of the BG model needs to be addressed. At the same time, the whole pedestrian detection system needs to be fast enough to meet the requirement of real-time applications.

To answer the challenges that have been mentioned, this paper proposes an integration of a BG subtraction algorithm based on guided filtering [18] and a pedestrian detection algorithm using YOLO [19]. The key idea is to develop a fast pedestrian detection system that is robust to noisy frames. The guided filtering part allows the proposed system to generate an ROI while filtering the noises in the incoming frames. Subsequently, the detection process can be executed at a fast speed with YOLOv3. Unlike the previously prevalent approaches that used variants of R-CNN [20], [21], which are two-stage deep-learning-based object detectors, YOLOv3 is a one-stage object detector. This allows YOLOv3 to have a faster inference time, with the speed at about 45 FPS (frame per second). Interestingly, the accuracy of YOLOv3 is not significantly compromised. It is also worth noting that YOLOv3 excels at detecting large objects in the PASCAL dataset [22], which contains a large number of objects that are classified as pedestrians. Thus, YOLOv3 is a natural choice for a real-time pedestrian detection system.

In summary, the main contributions of this study are:

- To develop a novel and robust BG subtraction algorithm using guided filtering and texture-based modeling. This algorithm is expected to be robust against noise with the use of guided filtering.

- To apply YOLOv3 to detect pedestrians based on the ROI that is provided by the BG subtraction algorithm. This ultimately leads to a fast and accurate pedestrian detection system for real-time applications.

The remainder of this paper is organized as follows. The previous works related to this study are presented in Section 2. The details of the proposed BG subtraction model are provided in Section 3. The experiment result is presented and discussed in Section 4. Finally, the study is concluded in Section 5.

II. RELATED WORKS

A. BG SUBTRACTION FOR FINDING ROI

BG subtraction is a standard way of detecting ROI from BG to find objects in successive frames. The BG subtraction is used in moving pedestrian detection fields to find the probable pedestrian areas (ROI) prior to detecting the real pedestrian object in a surveillance camera [23], [24]. The common method is based on color and texture features [1], [3], [25], which can utilize either pixel-based or block-based processing. To successfully detect the ROI from BG, one of the most prevalent approaches is to use an edge-aware filtering technique. This technique has a unique trait that can also filter the noise in the frames while detecting ROI. Recently, edge-aware filtering has been applied in many applications, for example, in the study by Wang *et al.* [26] and Munadi *et al.* [27]. Currently, the bilateral filter [28] and anisotropic diffusion [29] are the most popular variant of edge-aware filters. Despite their popularity, these two filters require a relatively high computational cost. To alleviate this issue, the guided filter [18] was developed, which is increasingly being applied in many fields, such as image fusion, image matting, up-sampling, etc. [30], [31]. Due to its non-approximate implementation, the guided filter is more preferred than other filters. Compared to other filters, the guided filter is able to generate a filtered image with improved quality at a faster runtime due to its invariant filter size [30].

B. MOVING PEDESTRIAN DETECTION VIA HANDCRAFTED-FEATURE-BASED TECHNIQUES

Before the advent of deep learning, the solution for object detection, including pedestrian detection, relies on the use of handcrafted features that are subsequently fed into a detector algorithm. Specifically for pedestrian detection, the most frequently employed features are the Histogram of Oriented Gradients (HOG) [32], [33], Haar-like features [34], [35], Viola-Jones features [36], texture features [37], and Local Binary Patterns (LBP) [38]. Since pedestrians are typically moving, Spatio-temporal features are also commonly used for pedestrian detection [39], [40]. It is also beneficial to use the handcrafted features for an intermediary BG subtraction process within a detection pipeline, as demonstrated by Kanagamalliga and Vasuki [41]. Recently, Kim *et al.* [60] focused on integrating the teacher-student concept into the standard random forest (RF) to create a novel fast pedestrian

detection algorithm for surveillance cameras that can be run on a low-level computer device.

C. MOVING PEDESTRIAN DETECTION VIA DEEP LEARNING TECHNIQUES

Like in most computer vision tasks, deep learning has also emerged as a preference for pedestrian detection. The deep learning algorithm in a pedestrian detection system is usually treated as a feature extractor, whose features are processed by another algorithm for the detection task. For instance, in the study proposed by Chahyati *et al.* [42] and Zhang *et al.* [43], the features were extracted from a deep learning algorithm named Faster R-CNN. Similarly, Li *et al.* [44] used the features from a fully convolutional network (FCN). Not only improve pedestrian detection system performance, but the utilization of deep learning can also alleviate challenging problems such as detecting pedestrians from an occluded image [13].

III. METHODS

Unlike previous approaches, our proposed method integrates a guided-filtering-based BG subtraction algorithm prior to a deep learning algorithm for pedestrian detection. The motivation for incorporating guided filtering is to relieve unwanted noises in the images that can degrade the accuracy of pedestrian detection. Fig. 2 depicts the outline of the proposed system pipeline. Firstly, the image is inputted into the BG subtraction algorithm to generate a bitmap that discriminates foreground and background. The foreground can be thought of as the promising part of the image containing pedestrians that eventually are detected by the subsequent pedestrian detector. Therefore, to eliminate unnecessary computation, the input image is cropped to the smallest part of the image that contains all foreground. Afterward, this cropped image is fed into a deep learning algorithm for the final pedestrian detection. Because of the elimination of unnecessary computation, our proposed approach is guaranteed to run faster than the typical deep-learning-based pedestrian detection system. In the next two subsections, the detail of this pipeline is elaborated. Specifically, the first subsection covers the details of the BG subtraction algorithm and the second subsection covers the detail of the deep learning algorithm for the final pedestrian detection.

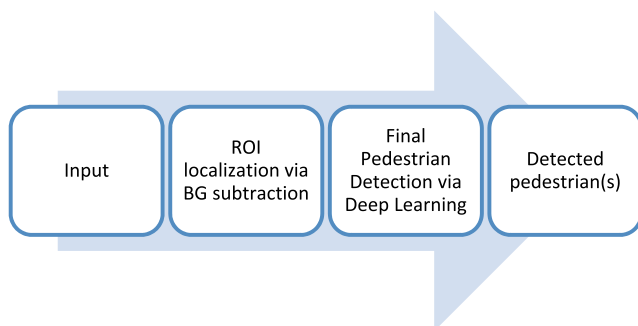


FIGURE 2. The pipeline of our proposed method.

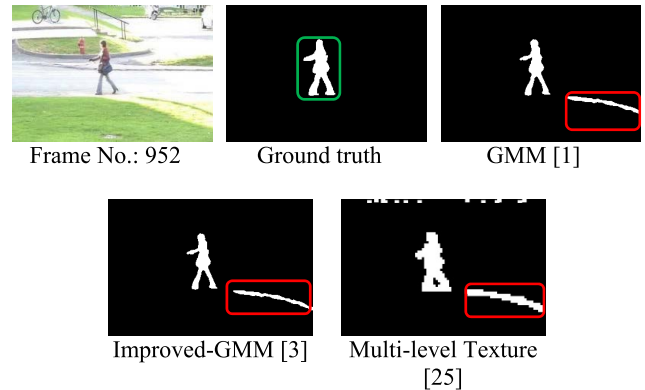


FIGURE 3. The observed failure of several BG subtraction methods at frame 952 in the CDNET 2014 – Pedestrians dataset.

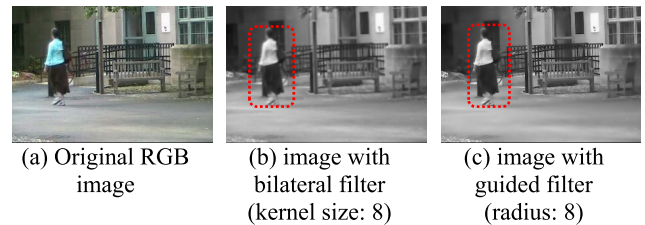


FIGURE 4. The qualitative comparison between bilateral and guided filter.

A. ROI LOCALIZATION VIA BG SUBTRACTION

In brief, our proposed BG subtraction algorithm combines guided filtering and the improved version of the multi-level texture BG subtraction algorithm proposed by Yeh *et al.* [25]. This proposal was motivated by our observation of the existing BG subtraction methods' behavior, including the multi-level texture method that naturally identifies occurring noises as foreground. The observation is presented in Fig. 3. In the figure, the correct foreground region is marked in green, and the incorrectly identified foreground is in red. Because BG subtraction is an integral part of our proposed pedestrian detection method, this flaw may introduce performance degradation to the overall detection system. For this reason, this paper proposes to infuse guided filtering into a BG subtraction algorithm in a pedestrian detection system.

The first process of our proposed BG subtraction is to apply a guided filter to an input frame I . This process generates a grayscale image I_{guided} , whose noises have been filtered. Afterward, I_{guided} is fed into the multi-level texture BG model [25] to produce a bitmap that separates foreground and background. In this study, the BG model is applied to I_{guided} for each 4×4 -pixel block, as suggested by Yeh *et al.* [25]. Subsequently, the final binary bitmap BM_{guided} is obtained by calculating the average of each block and comparing each pixel value with its corresponding mean block. A pixel in BM_{guided} is set to 1 if the I_{guided} pixel at the same location is greater than the mean value, and vice versa. In the first frame, the BM_{guided} 's blocks are stacked to obtain the initial BG model BM_{mod} . The initial model is subsequently updated for each incoming frame to get a more accurate representation of the true BG. The same post-processing steps

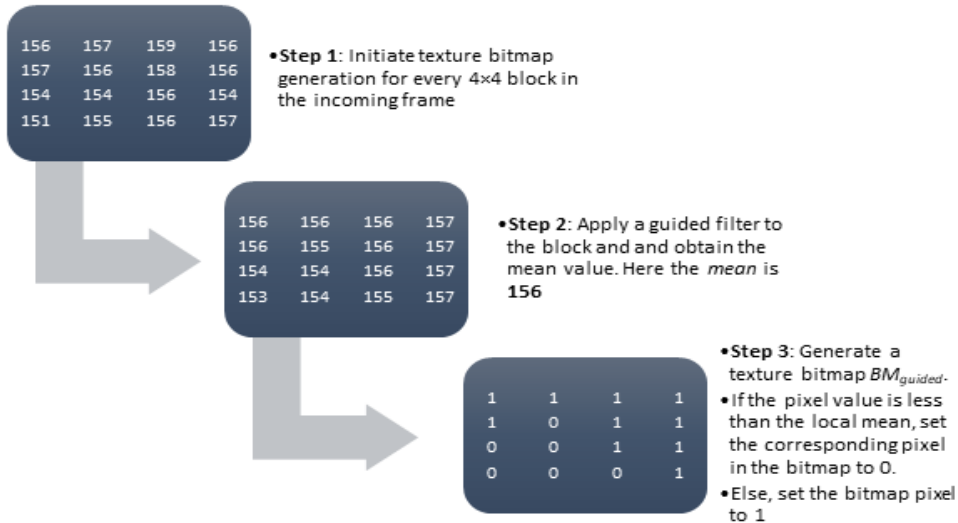


FIGURE 5. An illustration of texture bitmap generation.

(connected component and morphology), update rule and learning rate that was suggested by Yeh et al. [25] are used. Finally, to determine whether the blocks in a new incoming frame are foreground or background, they are compared to the BM_{mod} . The details of this comparison process are elaborated in subsection 3.1.2.

B. GUIDED FILTER AS PRE-PROCESSING

Since the guided filter [18] directly inspired our algorithm, this subsection provides a brief review of the edge-preserving property and its formula. The guided filter algorithm assumes that an edge-preserving smoothing filter can be learned via a linear model of the filtered image I_{guided} from a guidance image G within a window w_n that surrounds a pixel n . This can be formally expressed as follows:

$$I_{guided} = a_n G_p + b_n, \quad \forall p \in w_n \quad (1)$$

where a_n and b_n are constants that have a unique value for each window in the image. The value can be obtained analytically by framing the case as an optimization problem to minimize the squared error between I_{guided} and I as well as an $L2$ regularization on a_n , which is formally expressed as follows:

$$E(a_n, b_n) = \sum_{p \in w_n} ((a_n G_p + b_n - I_p)^2 + \varepsilon a_n^2) \quad (2)$$

where ε is a parameter to adjust the effect of the regularization term. To estimate a_n and b_n , a linear model as in equations (3) and (4) are utilized.

$$a_n = \frac{(1/|w|) \sum_{p \in w_n} G_p I_p - \bar{G}_n \bar{I}_n}{\sigma_n^2 + \varepsilon} \quad (3)$$

$$b_n = \bar{I}_n - a_n \bar{G}_n \quad (4)$$

where \bar{G}_n and \bar{I}_n are respectively the local means in a window w centered at pixel n of the G and I value, $|w|$ is the window's

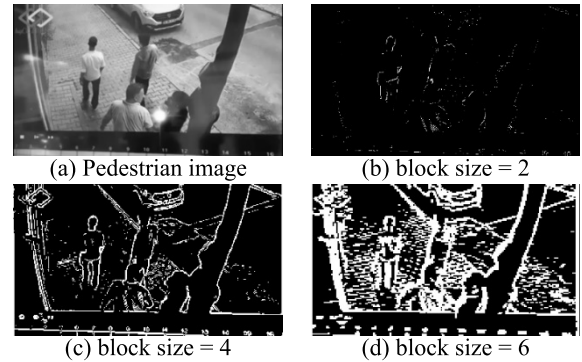


FIGURE 6. The generated bitmap from the proposed texture descriptor with different block sizes.

size, and σ_n^2 is the variance of G in w_n . The guided filtering process was implemented using OpenCV. Fig. 4 shows the comparison between bilateral and guided filters. As clearly shown in the figure that the guided filter is able to preserve the interesting area while smoothing the remaining regions.

C. TEXTURE-BASED BG MODELING WITH GUIDED FILTERING

The texture-based BG modeling in this study was applied to each non-overlapping 4×4 block in the incoming frame. Firstly, each block is filtered by a guided filter. Afterward, a texture bitmap is generated by thresholding each pixel value with respect to the local mean of the block. If the pixel value is less than the mean, then the corresponding pixel in the texture bitmap is set to 0. Else, the bitmap pixel is set to 1. This process is illustrated in Fig. 5.

To justify the choice of 4×4 block size, an example of the generated bitmap with different block sizes using the proposed texture descriptor is visualized in Fig. 6. The block size of 2×2 failed to identify most of the important textures (Fig. 6(b)). Meanwhile, the 6×6 block size captures excessive

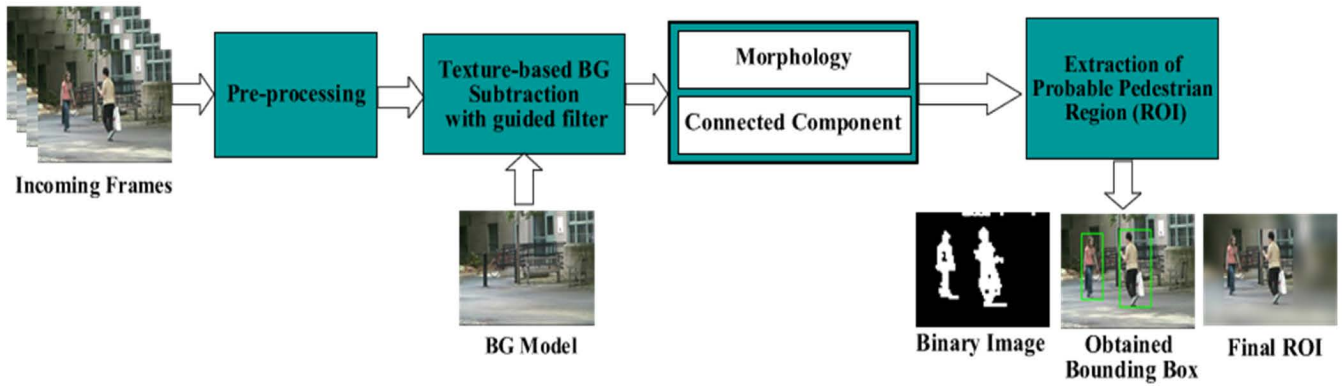


FIGURE 7. The pipeline of the proposed BG modeling.

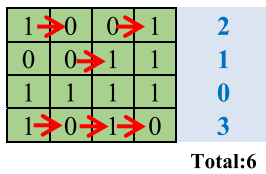


FIGURE 8. The generated bitmap from the proposed texture descriptor with different block sizes.

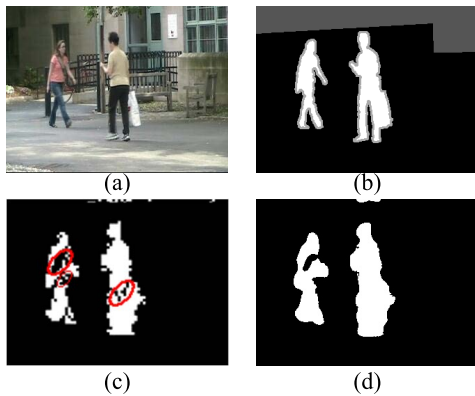


FIGURE 9. The illustration of improved results from the proposed method (CDNET 2014 – backdoor, frame no. 1860). (a) The RGB image and Guidance G , (b) the ground truth, (c) Yeh et al. [25], and (d) our final improvement result.

texture (Fig. 6(d)). The 4×4 block size generates the bitmap with the most perfectly balanced texture among the tested block size (Fig. 6(c)).

Motivated by PBAS [50], the texture information is used in this study to model the observed blocks by adding the neighborhood information. The improved BG subtraction method is divided into two parts, namely the initial improvement, and final improvement.

1) INITIAL IMPROVEMENT

The first modification was made to Yeh et.al.’s updating mechanism. In contrast to this method, which only updates the BG model of the observed block, the proposed idea incorporates the adjacent blocks by checking for similarity before

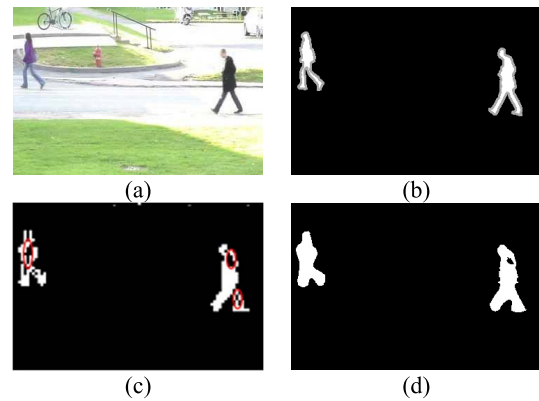


FIGURE 10. The illustration of improved results from the proposed method (CDNET 2014 – pedestrians, frame no. 582). (a) The RGB image and Guidance G , (b) the ground truth, (c) Yeh et al. [25], and (d) our final improvement result.

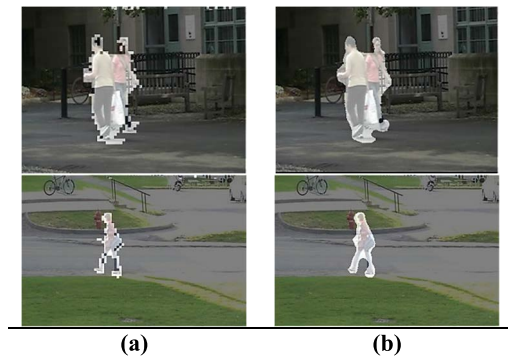


FIGURE 11. The comparison of overlay results between block textured-based by Yeh et al. [25], and our proposed work (backdoor - frame no. 1890, and pedestrians - frame no. 965, respectively). (a) Yeh et al. [25], and (b) our proposed result.

updating the adjacent BG models. More specifically, different from [50] which selects and updates randomly neighboring pixels, the proposed step is first check the similarity of the observed BM_{model} with $BM_{adjacent_model}$ via hamming distance. If the similarity of models exceeds the $TH_{adjacent}$, then all $BM_{adjacent_model}$ is replaced by its current binary bitmap.

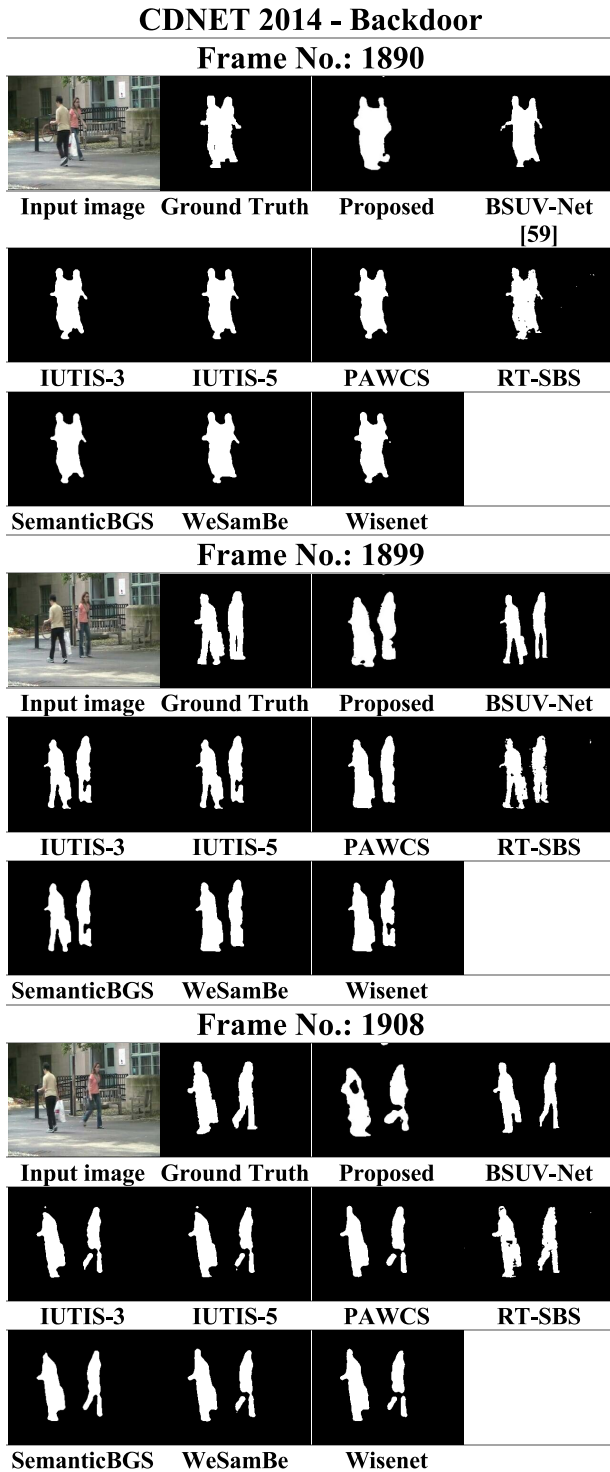


FIGURE 12. The comparison between our proposed approach and the top-10 SOTA methods from CDNET 2014-Backdoor (supervised methods are not compared).

2) FINAL IMPROVEMENT

The bit transition is estimated by Yeh, *et al.* [25] to determine the mode of a block. When a block is complex, the upper level is used (2 or 3-bits mode, instead of 1-bit mode). The result of the first phase in the proposed approach of this study

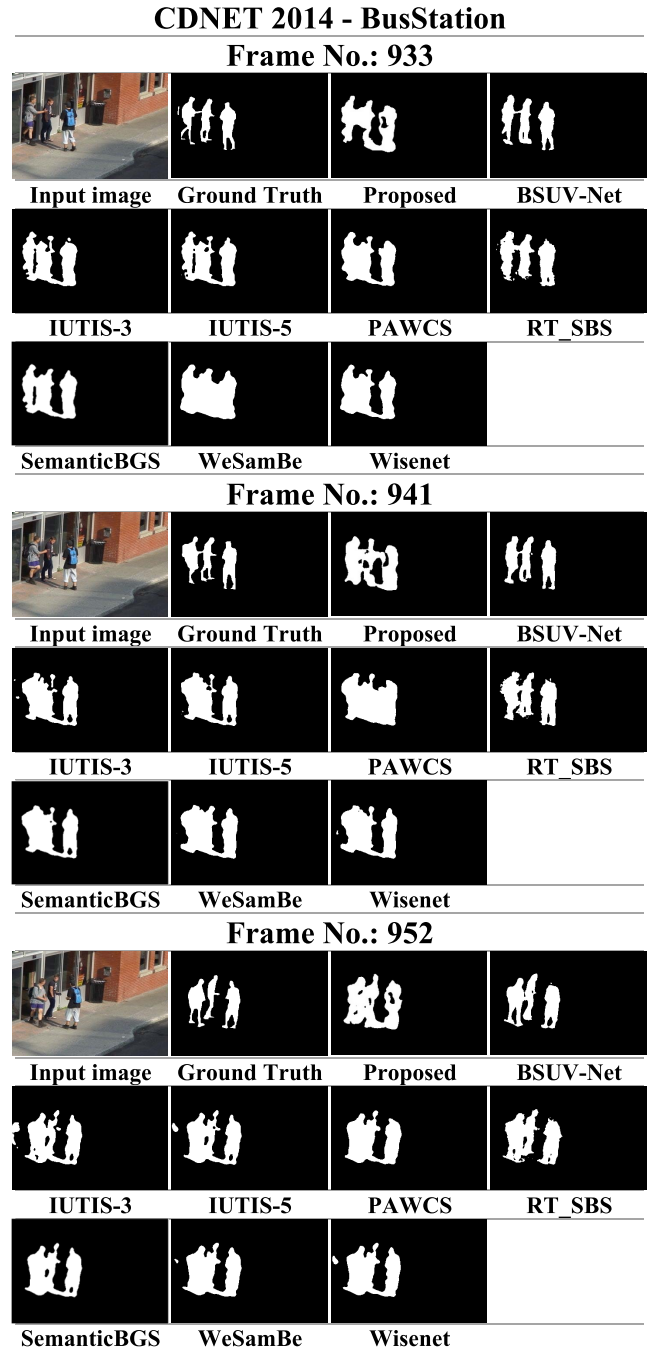


FIGURE 13. The comparison between our proposed approach and the top-10 SOTA methods from CDNET 2014-busStation (supervised methods are not compared).

allows for further development by using bit transition to check the complexity of the observed block and adjacent blocks. Fig. 8 shows an example of accumulating bit transition of a block. Note that, the higher the total transition, the more complex block, and texture information will be.

If the observed block is regarded as BG, the complexity of current adjacent blocks is considered to be identified if it is an FG block. If both conditions are met (the adjacent blocks are all complex and FG blocks), the label of the observed block

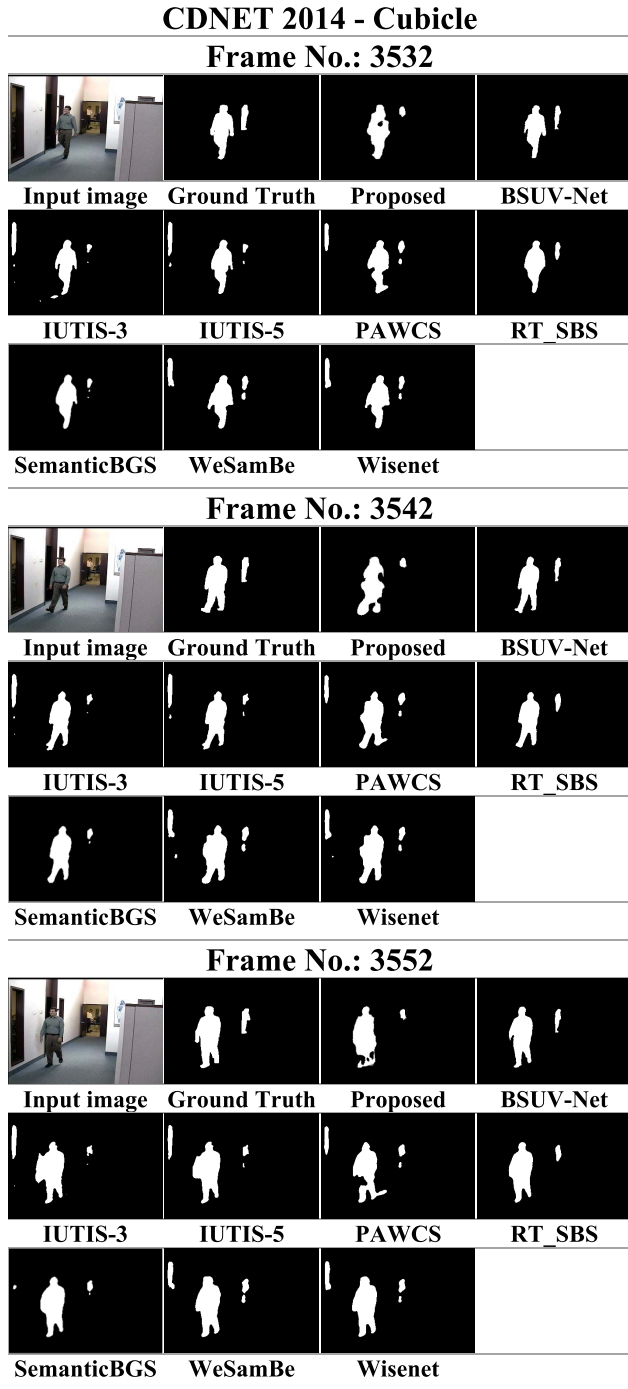


FIGURE 14. The comparison between our proposed approach and the top-10 SOTA methods from CDNET 2014-Cubicle (supervised methods are not compared).

is changed from BG to FG, as shown in Eq. (5).

$$\begin{aligned}
 & \text{Replace } (BM_{obs}) \\
 & = \begin{cases} BM_{out} = FG, & \text{if } (BM_{complex} \text{ AND } BM_{out_adj_FG} = true) \\ BM_{out} = BG, & \text{else} \end{cases}
 \end{aligned}
 \tag{5}$$

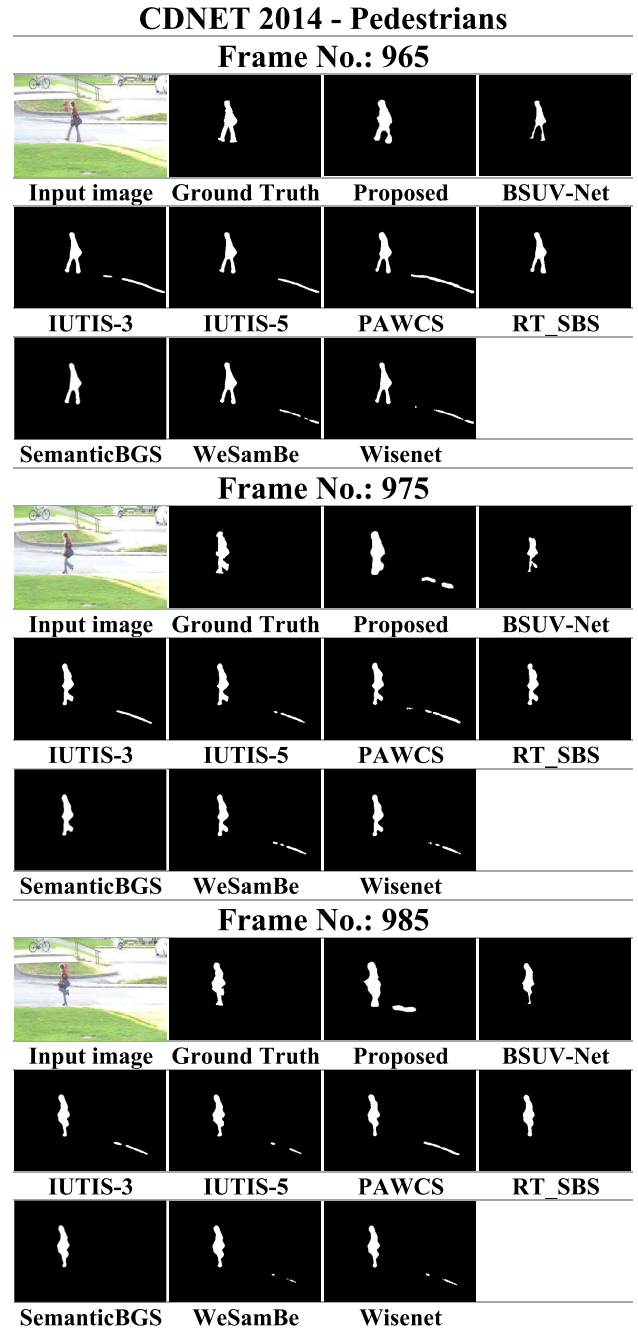


FIGURE 15. The comparison between our proposed approach and the top-10 SOTA methods from CDNET 2014-Pedestrians (supervised methods are not compared).

Once the binary mask is obtained, the guided feathering is finally performed to further improve the results. To be specific, a binary mask b_{output} is filtered under the guidance of G (see section III.b). The parameters are $r = 5$, and $\epsilon = 0.2^2$ for the guided filter. Where the r and ϵ are radius and epsilon, respectively. As shown in Fig. 9-10, the fragment issues highlighted in red (in the inner region of a detected object) can be alleviated.

Furthermore, more detailed comparative evaluations are shown in the subsequent results. It aims to overlay the

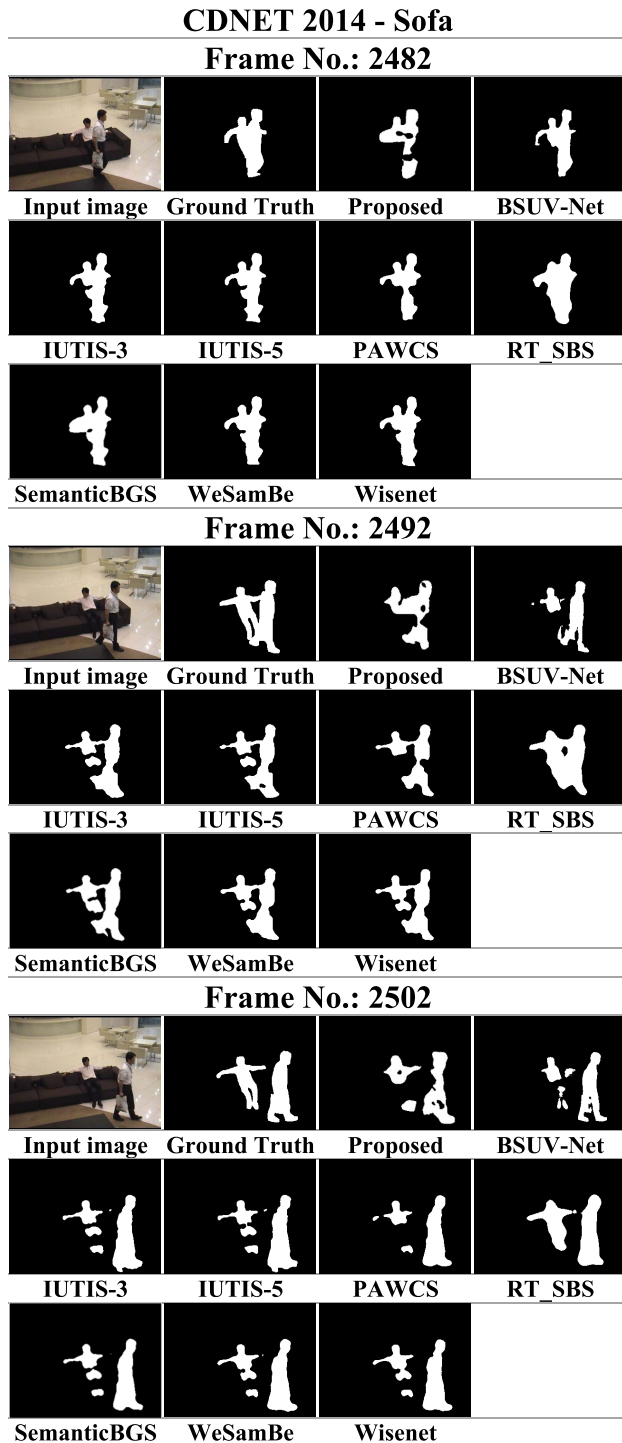


FIGURE 16. The comparison between our proposed approach and the top-10 SOTA methods from CDNET 2014-Sofa (supervised methods are not compared).

obtained final binary mask over an RGB color image. This verifies that the proposed method is able to improve the previous work, especially for optimizing the block textured-based approach. It is noteworthy that the block textured-based is selected to guarantee the initial moving object detection is executed rapidly prior to inference through a deep learning pipeline.

TABLE 1. FPS comparison of the previous approaches and our proposed approach in a Non-GPU environment.¹

	FRAME PER SECOND (FPS)	RESOLUTION AND CONFIGURATION
SEMANTICBGS [51]	~7 FPS	320x240 (CPU)
IUTIS-3 [52]	~10 FPS	320x240 (CPU)
WISENETMD [53]	12 FPS	320x240 (CPU)
PAWCS [54]	~27 FPS	320x240 (CPU)
WESAMBE [55]	~2 FPS	320x240 (CPU)
SUBSENSE [56]	~30 FPS	320x240 (CPU)
RT-SBS [6]	25 FPS	320x240 (CPU)
GMM [1]	~21 FPS	320x240 (CPU)
IMPROVED GMM [3]	~49 FPS	320x240 (CPU)
OUR PROPOSED APPROACH	~55 FPS	320x240 (CPU)

D. FINAL PEDESTRIAN DETECTION VIA DEEP LEARNING

To detect pedestrians from the previously generated ROI by our BG model, a deep-learning-based pedestrian detector is utilized, especially with the model based on Convolutional Neural Networks (CNN). It is currently the most popular model to solve many computers vision tasks, including image classification and object detection. In its simplest form, CNN consists of convolutional, pooling, and fully-connected (FC) layers. The core of CNN is the convolutional layer, which applies a fixed-size kernel to the input matrix via a convolution process and sends the output matrix to the next layer. To allow non-linear mapping, a non-linear activation function is applied after each layer. CNN has been observed to generate more optimized features than hand-crafted features, which leads to better performance.

In particular, the model by YOLOv2 [46], YOLOv3 [19], and YOLOv4 [58] are used in this work, which demonstrated robust performance for object detection. YOLO is a type of CNN specially engineered for fast object detection with competitive accuracy. It is an improved model from the previous versions of YOLOv1 [45]. At its core, YOLO is a single regression model that fully connected and explains its fast inference compared to other object detection methods that are typically multi regression models. The single regression model is achieved by framing detection as regression of bounding boxes for each $S \times S$ grid in the input image. If a target bounding box is centered at a certain grid, the representation of that grid is utilized for detecting the bounding box. Each grid can detect a fixed number of bounding boxes B , along with the corresponding confidence score, which is calculated as $P_r(object) * IOU_{pred}^{truth}$. The confidence score can

¹source from CDNET2014 website: <http://jacarini.dinf.usherbrooke.ca/>



FIGURE 17. The final pedestrian detection through YOLOv2, YOLOv3, and YOLOv4 (CDNET2014-backdoor).



FIGURE 18. The final pedestrian detection through YOLOv2, YOLOv3, and YOLOv4 (CDNET2014-busStation).

be interpreted as the probability that the detected bounding box is correct, which is measured by the Intersection over Union (IOU) of the predicted box compared to intersecting target boxes. If no target boxes intersect the detected box, its confidence score is set to 0.

To get a better detection in the second version, the size of the anchor boxes was set with the size obtained from the training dataset via k-means clustering. Because YOLOv2, YOLOv3, and YOLOv4 include “person” as one of the object categories, it can be employed as a pedestrian detector.

Given its fast inference time, these methods are suitable to be used as the detector of our proposed method that is designed for real-time application.

IV. RESULTS AND DISCUSSION

This section aims to present and discuss our simulation result as well as its evaluation. The simulation hardware is a PC with an Intel i7-7700HQ processor, 16 GB of memory, and NVIDIA GeForce GTX 1050 Ti 4 GB. It is noteworthy that the comparison between the previous works against the

CDNET 2014 – Cubicle

Frame No.: 3532



Frame No.: 3542



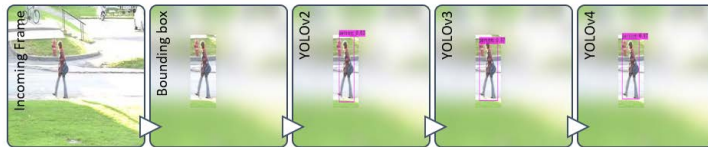
Frame No.: 3552



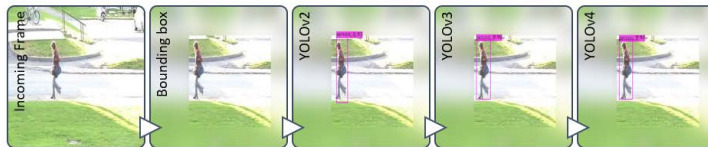
FIGURE 19. The final pedestrian detection through YOLOv2, YOLOv3, and YOLOv4 (CDNET2014-cubicle).

CDNET 2014 – Pedestrians

Frame No.: 965



Frame No.: 975



Frame No.: 985

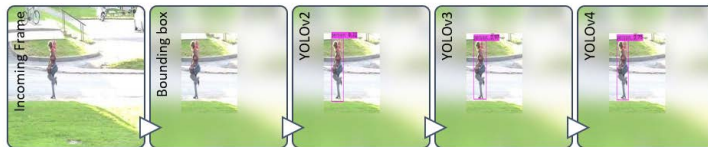


FIGURE 20. The final pedestrian detection through YOLOv2, YOLOv3, and YOLOv4 (CDNET2014-pedestrians).

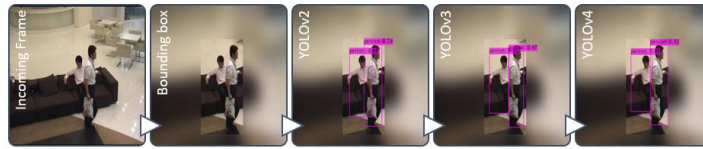
proposed work is evaluated in terms of speed since the main objective of our proposed approach is to be deployed in real-time or edge devices. In our simulation, the proposed approach can achieve around 55 frames per second, the fastest among the previous approaches that have been considered in this study. The comparison of our proposed approach to the previous approaches is presented in Table 1. All approaches were performed without GPU acceleration, following the typical protocol of BG modeling studies.

A. QUALITATIVE COMPARISON FOR PEDESTRIAN DETECTION

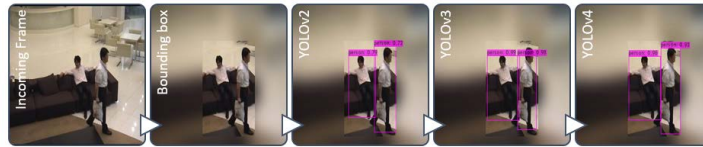
The comparison discussed in this section was evaluated on the CDNET2014 dataset [10], which introduces several challenging pedestrian scenes. In this paper, five representative scenarios, which are backdoor, busStation, cubicle, pedestrians, and sofa, respectively are selected. Notably, these videos contain difficult challenges such as shadow, and occlusion. In Fig. 12-16, the comparison of the BG subtraction result

CDNET 2014 – Sofa

Frame No.: 2482



Frame No.: 2492



Frame No.: 2502

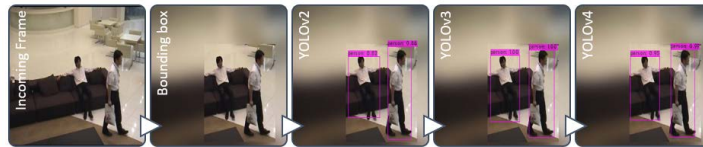


FIGURE 21. The final pedestrian detection through YOLOv2, YOLOv3, and YOLOv4 (CDNET2014-sofa).

TABLE 2. Pedestrian detection: comparison of quantitative measurements (CDNET2014 Dataset “Backdoor”, Frame No. 1890, 1899, and 1908).

	Wisenet	WeSamBE	SemanticBGS	RT-SBS	PAWCS	IUTIS-5	IUTIS-3	BSUV-Net	Proposed
Specificity	0.9975	0.9967	0.9994	0.9988	0.9974	0.9996	0.9994	0.9997	0.9884
FPR	0.0025	0.0033	0.0006	0.0012	0.0026	0.0004	0.0006	0.0002	0.0116
FNR	0.0210	0.0172	0.0215	0.0221	0.0210	0.0223	0.0222	0.0284	0.0116
PWC	2.0917	1.8210	1.9696	2.0792	2.1078	2.0298	2.0277	2.5442	2.0928
Precision	0.9785	0.9722	0.9944	0.9889	0.9775	0.9957	0.9944	0.9974	0.8959
F1 Score	0.8940	0.9099	0.8986	0.8928	0.8931	0.8953	0.8956	0.8679	0.8949

TABLE 3. Pedestrian detection: comparison of quantitative measurements (CDNET2014 Dataset “busStation”, Frame No. 933, 941, and 952).

	Wisenet	WeSamBE	SemanticBGS	RT-SBS	PAWCS	IUTIS-5	IUTIS-3	BSUV-Net	Proposed
Specificity	0.9513	0.9429	0.9580	0.9667	0.9485	0.9572	0.9568	0.9827	0.9594
FPR	0.0487	0.0571	0.0420	0.0333	0.0515	0.0428	0.0432	0.0173	0.0406
FNR	0.0027	0.0015	0.0027	0.0021	0.0038	0.0042	0.0048	0.0045	0.0087
PWC	4.7855	5.4545	4.1624	3.2928	5.1447	4.3765	4.4753	2.0309	4.6363
Precision	0.5948	0.5640	0.6299	0.6852	0.5799	0.6204	0.6158	0.8020	0.5774
F1 Score	0.7355	0.7144	0.7617	0.8033	0.7193	0.7484	0.7424	0.8648	0.6920

TABLE 4. Pedestrian detection: comparison of quantitative measurements (CDNET2014 Dataset “Cubicle”, Frame No. 3532, 3542, and 3552).

	Wisenet	WeSamBE	SemanticBGS	RT-SBS	PAWCS	IUTIS-5	IUTIS-3	BSUV-Net	Proposed
Specificity	0.9886	0.9874	0.9991	0.9992	0.9896	0.9888	0.9868	0.9999	0.9953
FPR	0.0114	0.0126	0.0009	0.0008	0.0104	0.0112	0.0132	0.0001	0.0047
FNR	0.0078	0.0061	0.0122	0.0123	0.0113	0.0118	0.0126	0.0106	0.0109
PWC	1.7981	1.7519	1.2334	1.2267	2.0316	2.1591	2.4215	0.9998	1.4778
Precision	0.8348	0.8250	0.9843	0.9852	0.8403	0.8292	0.8043	0.9978	0.9096
F1 Score	0.8568	0.8635	0.8897	0.8907	0.8345	0.8231	0.8044	0.9124	0.8566

from our proposed approach and the top-10 SOTA methods from CDNET 2014 are visually presented. The proposed approach removes the incorrectly detected regions caused by the previously mentioned challenges. It also eliminates noises better than the multi-level texture approach. As the result,

our proposed method provides a tighter ROI, which leads to faster detection of the subsequent deep learning process. In Fig. 17-21, the detection result from YOLO given the ROI from the BG model are visualized. As specifically highlight that YOLO can accurately detect persons in the given frames,

TABLE 5. Pedestrian detection: comparison of quantitative measurements (CDNET2014 Dataset "Pedestrians", Frame No. 965, 975, and 985).

	Wisenet	WeSamBE	SemanticBGS	RT-SBS	PAWCS	IUTIS-5	IUTIS-3	BSUV-Net	Proposed
Specificity	0.9965	0.9970	0.9998	0.9997	0.9920	0.9960	0.9943	0.9999	0.9927
FPR	0.0035	0.0030	0.0002	0.0003	0.0080	0.0040	0.0058	0.0001	0.0073
FNR	0.0046	0.0041	0.0041	0.0045	0.0042	0.0041	0.0038	0.0087	0.0004
PWC	0.7921	0.6956	0.4216	0.4691	1.1922	0.7851	0.9344	0.8530	0.7586
Precision	0.8453	0.8683	0.9896	0.9849	0.7106	0.8328	0.7738	0.9958	0.7335
F1 Score	0.8217	0.8437	0.8986	0.8858	0.7588	0.8280	0.8029	0.7659	0.8366

TABLE 6. Pedestrian detection: comparison of quantitative measurements (CDNET2014 Dataset "Sofa", Frame No. 2482, 2492, and 2502).

	Wisenet	WeSamBE	SemanticBGS	RT-SBS	PAWCS	IUTIS-5	IUTIS-3	BSUV-Net	Proposed
Specificity	0.9950	0.9939	0.9935	0.9830	0.9972	0.9927	0.9922	0.9982	0.9900
FPR	0.0050	0.0061	0.0065	0.0170	0.0028	0.0073	0.0078	0.0018	0.0100
FNR	0.0205	0.0182	0.0173	0.0073	0.0292	0.0198	0.0199	0.0325	0.0297
PWC	2.2882	2.1823	2.1454	2.1901	2.8789	2.4392	2.4918	3.0777	3.6275
Precision	0.9483	0.9401	0.9306	0.8535	0.9684	0.9285	0.9227	0.9792	0.8643
F1 Score	0.8764	0.8854	0.8847	0.8915	0.8336	0.8705	0.8678	0.8203	0.7596

TABLE 7. Pedestrian detection: comparison of average quantitative measurements and frame rates.

	Wisenet	WeSamBE	SemanticBGS	RT-SBS	PAWCS	IUTIS-5	IUTIS-3	BSUV-Net	Proposed
Avg. Specificity	0.9858	0.9836	0.9900	0.9895	0.9849	0.9869	0.9859	0.9961	0.9851
Avg. FPR	0.0142	0.0164	0.0100	0.0105	0.0151	0.0131	0.0141	0.0039	0.0149
Avg. FNR	0.0113	0.0094	0.0116	0.0097	0.0139	0.0125	0.0127	0.0169	0.0123
Avg. PWC	2.3511	2.3811	1.9865	1.8516	2.6710	2.3580	2.4701	1.9011	2.5186
Avg. Precision	0.8403	0.8339	0.9057	0.8996	0.8153	0.8413	0.8222	0.9544	0.7961
Avg. F1 Score	0.8369	0.8434	0.8666	0.8728	0.8079	0.8330	0.8226	0.8463	0.8080

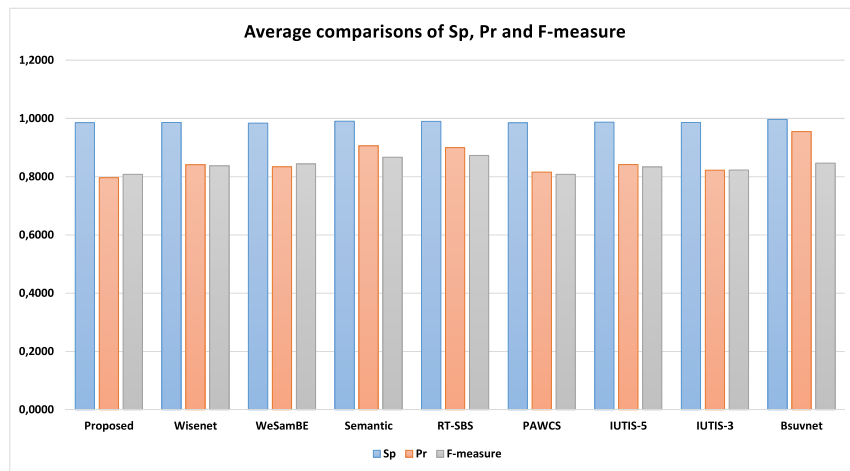


FIGURE 22. A group bar chart of average comparisons of Sp, Pr and F-measure (higher scores are better).

so that demonstrates the suitability of YOLO as the pedestrian detector in our proposed pipeline.

B. QUANTITATIVE COMPARISON FOR PEDESTRIAN DETECTION

The performance of our proposed approach compared to the previous approaches is qualitatively evaluated. The quantitative evaluation in this study is based on pixel-wise binary measurements with the following metrics: Specificity (Sp), False Positive Rate (FPR), False Negative Rate (FNR),

Percentage of Wrong Classifications (PWC), Precision (Pr), and F1 score [59]. In the case of a BG model assessment, specificity measures the number of background pixel which was correctly classified; FPR measures the ratio of background pixels misclassified as foregrounds; FNR measures the ratio of foreground pixels misclassified as backgrounds; PWC measures the overall misclassification rate; precision measures the number of foreground pixel which was correctly classified; F1 measures the harmonic mean of precision and recall (1 - FNR).

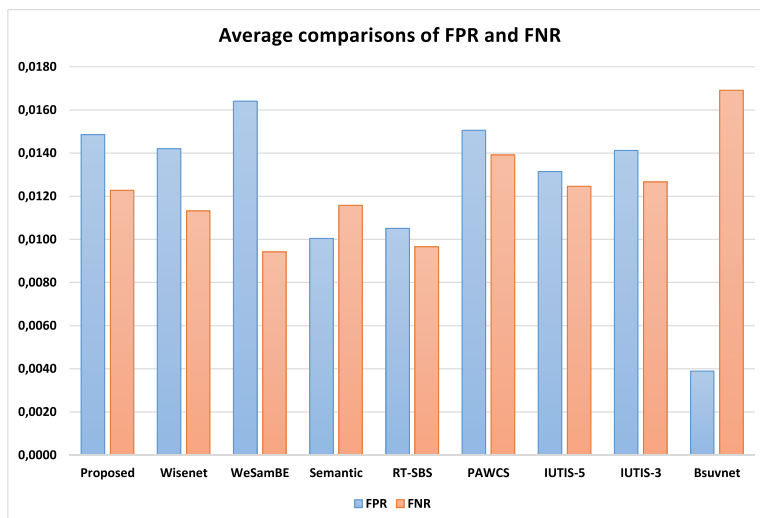


FIGURE 23. A group bar chart of average comparisons of FPR and FNR (lower scores are better).

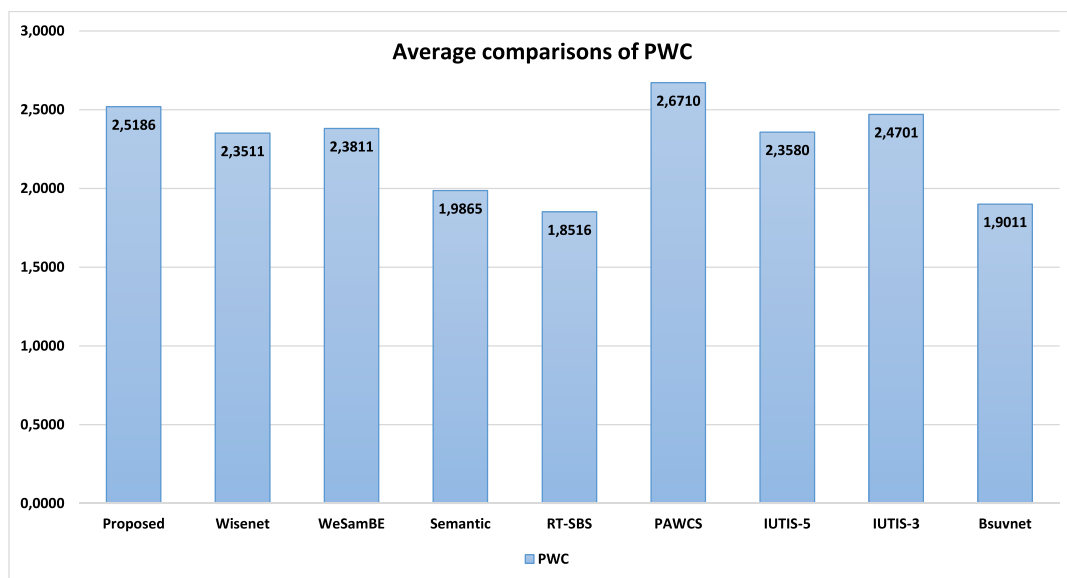


FIGURE 24. A group bar chart of average comparisons of PWC (lower scores are better).

The results of the quantitative evaluation are presented in Tables 2 to Table 7. The best performance in these tables is highlighted in red, while the second best is highlighted in blue. As visualized in Fig. 22 through 24, in general, our proposed method can be very competitive and applicable for extracting moving regions. It is noteworthy that the proposed approach aims to localize ROI prior to pedestrian classification through YOLO. Therefore, as shown in Fig. 25, the whole pipeline can execute the incoming frames faster than full-frame processing. It allows the proposed pipeline to be applied in a real-time environment.

From the tables, it can be seen that the specificity value of the proposed approach yields a slightly similar value to the best specificity value, namely BSUV-NET, in most cases.

The FPR and FNR values obtained in the proposed approach are not the best for almost all representative scenarios, but in “Backdoor” and “Pedestrian” scenarios as the best FNR among other approaches. The percentage of wrong classification (PWC) shows that the approach generates the average value that marginally not different to the other approaches. For the obtained precision, our proposed approach yields the second highest score for the “cubicle” scenario, while in the other scenarios almost achieve the lowest score of precision value. Moreover, the F1 values of proposed approach are marginally similar to the highest score.

Table 7 represented the average value of quantitative measurements of all pedestrian scenes. From the table can it be seen that the best performance is provided by the

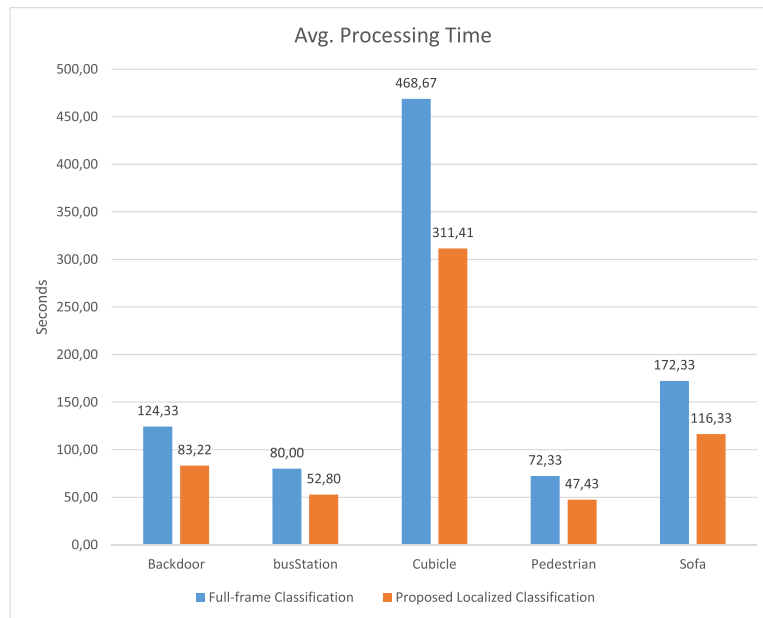


FIGURE 25. Average comparisons of processing time in CPU (frames per second) between full-frame and proposed localized classification (lower scores are better).

BSUV-NET approach. However, our proposed pedestrian detection pipeline obtained slightly identical values to the BSUV-NET in terms of specificity and FNR.

V. CONCLUSION

The advantages of our proposed pedestrian detection pipeline based on a robust ROI localization, have been comprehensively evaluated. The results of this study suggest that the robust ROI localization with a guided-filtering-based BG model contributes to the rapid and accurate pedestrian detection in our pipeline. Moreover, the guided filter allows our proposed method to be robust against various complex challenges caused by noises, which are not adequately handled by the previous approaches. The future works is to evaluate more pre-processing steps on the robust BG subtraction methods. In addition, comprehensive experiments through edge environments will enable this work to be applied in the product-ready and real-time environment.

REFERENCES

- [1] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Fort Collins, CO, USA, Jun. 1999, pp. 246–252.
- [2] C.-Y. Lin, K. Mughtar, W.-Y. Lin, and Z.-Y. Jian, "Moving object detection through image bit-planes representation without thresholding," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 4, pp. 1404–1414, Apr. 2020.
- [3] Z. Zivkovic, "Improved adaptive Gaussian mixture model for background subtraction," in *Proc. ICPR*, Cambridge, U.K., 2004, pp. 28–31.
- [4] Z. Zivkovic and F. van der Heijden, "Efficient adaptive density estimation per image pixel for the task of background subtraction," *Pattern Recognit. Lett.*, vol. 27, no. 7, pp. 773–780, May 2006.
- [5] O. Barnich and M. Van Droogenbroeck, "ViBe: A universal background subtraction algorithm for video sequences," *IEEE Trans. Image Process.*, vol. 20, no. 6, pp. 1709–1724, Jun. 2011.
- [6] A. Cioppa, M. V. Droogenbroeck, and M. Braham, "Real-time semantic background subtraction," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Abu Dhabi, United Arab Emirates, Oct. 2020, pp. 3214–3218, doi: 10.1109/ICIP40778.2020.9190838.
- [7] S. S. Sengar and S. Mukhopadhyay, "Moving object detection using statistical background subtraction in wavelet compressed domain," *Multimedia Tools Appl.*, vol. 79, pp. 5919–5940, Dec. 2019.
- [8] Y. Xu, H. Ji, and W. Zhang, "Coarse-to-fine sample-based background subtraction for moving object detection," *Optik*, vol. 207, pp. 164–195, Apr. 2020.
- [9] W. Kim, "Background subtraction with variable illumination in outdoor scenes," *Multimedia Tools Appl.*, vol. 77, no. 15, pp. 19439–19454, Aug. 2018.
- [10] Y. Wang, P.-M. Jodoin, F. Porikli, J. Konrad, Y. Benezeth, and P. Ishwar, "CDnet 2014: An expanded change detection benchmark dataset," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2014, pp. 387–394.
- [11] A. Brunetti, D. Buongiorno, G. F. Trotta, and V. Bevilacqua, "Computer vision and deep learning techniques for pedestrian detection and tracking: A survey," *Neurocomputing*, vol. 300, pp. 17–33, Jul. 2018.
- [12] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 21–37.
- [13] W. Ouyang and X. Wang, "A discriminative deep model for pedestrian detection with occlusion handling," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Providence, RI, USA, Jun. 2012, pp. 3258–3265.
- [14] S. Paisitkriangkrai, C. Shen, and A. van den Hengel, "Pedestrian detection with spatially pooled features and structured ensemble learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 6, pp. 1243–1257, Jun. 2016.
- [15] W. Guo, Y. Xiao, and G. Zhang, "Multi-scale pedestrian detection by use of AdaBoost learning algorithm," in *Proc. Int. Conf. Virtual Reality Vis., Shenyang, China*, Aug. 2014, pp. 266–271.
- [16] Y.-L. Hou and G. K. H. Pang, "People counting and human detection in a challenging situation," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 41, no. 1, pp. 24–33, Jan. 2011.
- [17] Y. Xia, H. Ruimin, W. Zhongyuan, and L. Tao, "Moving foreground detection based on spatio-temporal saliency," *J. Comput. Sci. Issues*, vol. 10, no. 1, p. 79, 2013.
- [18] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1397–1409, Jun. 2013.
- [19] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.
- [20] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 580–587.
- [21] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

- [22] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL visual object classes (VOC) challenge," *Int. J. Comput. Vis.*, vol. 88, no. 2, pp. 303–338, Jun. 2010.
- [23] E. N. Kajabad and S. V. Ivanov, "People detection and finding attractive areas by the use of movement detection analysis and deep learning approach," in *Proc. 8th Int. Young Scientist Conf. Comput. Sci.*, 2019, pp. 1–11.
- [24] Q. Peng, W. Luo, G. Hong, M. Feng, Y. Xia, L. Yu, X. Hao, X. Wang, and M. Li, "Pedestrian detection for transformer substation based on Gaussian mixture model and YOLO," in *Proc. 8th Int. Conf. Intell. Hum.-Mach. Syst. Cybern. (IHMSC)*, Aug. 2016, pp. 562–565.
- [25] C.-H. Yeh, C.-Y. Lin, K. Muchtar, and L.-W. Kang, "Real-time background modeling based on a multi-level texture description," *Inf. Sci.*, vol. 269, pp. 106–127, Jun. 2014.
- [26] Y. Wang, S. Piérard, S.-Z. Su, and P.-M. Jodoin, "Improving pedestrian detection using motion-guided filtering," *Pattern Recognit. Lett.*, vol. 96, pp. 106–112, Sep. 2017.
- [27] K. Munadi, K. Muchtar, N. Maulina, and B. Pradhan, "Image enhancement for tuberculosis detection using deep learning," *IEEE Access*, vol. 8, pp. 217897–217907, 2020.
- [28] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. IEEE Int. Conf. Comput. Vis.*, Jan. 1998, pp. 839–846.
- [29] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 7, pp. 629–639, Jul. 1990.
- [30] C. C. Pham and J. W. Jeon, "Efficient image sharpening and denoising using adaptive guided image filtering," *IET Image Process.*, vol. 9, no. 1, pp. 71–79, 2015.
- [31] R. Rajendran, S. Paramathma Rao, S. S. Aгаian, and K. Panetta, "A versatile edge preserving image enhancement approach for medical images using guided filter," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2016, pp. 2341–2346.
- [32] C. Q. Lai and S. S. Teoh, "A review on pedestrian detection techniques based on histogram of oriented gradient feature," in *Proc. IEEE Student Conf. Res. Develop.*, Dec. 2014, pp. 1–6.
- [33] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2005, pp. 886–893.
- [34] S. Zhang, C. Bauckhage, and A. B. Cremers, "Informed Haar-like features improve pedestrian detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 947–954.
- [35] M. Oren, C. Papageorgiou, P. Sinha, E. Osuna, and T. Poggio, "Pedestrian detection using wavelet templates," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 1997, pp. 193–199.
- [36] P. Viola, M. J. Jones, and D. Snow, "Detecting pedestrians using patterns of motion and appearance," *Int. J. Comput. Vis.*, vol. 63, no. 2, pp. 153–161, 2005.
- [37] D. M. Gavrilu and S. Munder, "Multi-cue pedestrian detection and tracking from a moving vehicle," *Int. J. Comput. Vis.*, vol. 73, no. 1, pp. 41–59, Jun. 2007.
- [38] X. Wang, T. X. Han, and S. Yan, "An HOG-LBP human detector with partial occlusion handling," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep. 2009, pp. 32–39.
- [39] Y. Jiang, J. Wang, Y. Liang, and J. Xia, "Combining static and dynamic features for real-time moving pedestrian detection," *Multimedia Tools Appl.*, vol. 78, no. 3, pp. 3781–3795, Feb. 2019.
- [40] K. Zhao, J. Deng, and D. Cheng, "Real-time moving pedestrian detection using contour features," *Multimedia Tools Appl.*, vol. 77, no. 23, pp. 30891–30910, Dec. 2018.
- [41] S. Kanagamalliga and S. Vasuki, "Contour-based object tracking in video scenes through optical flow and Gabor features," *Optik*, vol. 157, pp. 787–797, Mar. 2018.
- [42] D. Chahyati, M. I. Fanany, and A. M. Arymurthy, "Tracking people by detection using CNN features," *Proc. Comput. Sci.*, vol. 124, pp. 167–172, Jan. 2017.
- [43] L. Zhang, L. Lin, X. Liang, and K. He, "Is faster R-CNN doing well for pedestrian detection?" in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 443–457.
- [44] C. Li, X. Wang, and W. Liu, "Neural features for pedestrian detection," *Neurocomputing*, vol. 238, pp. 420–432, May 2017.
- [45] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.
- [46] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, San Juan, PR, USA, Jul. 2017, pp. 7263–7271.
- [47] G. Chen, Y. Ding, J. Xiao, and T. X. Han, "Detection evolution with multi-order contextual co-occurrence," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 1798–1805.
- [48] P. Singh, B. B. V. L. Deepak, T. Sethi, and M. D. P. Murthy, "Real-time object detection and tracking using color feature and motion," in *Proc. Int. Conf. Commun. Signal Process. (ICCSPP)*, Melmaruvathur, India, Apr. 2015, pp. 1236–1241.
- [49] M. Szarvas, A. Yoshizawa, M. Yamamoto, and J. Ogata, "Pedestrian detection with convolutional neural networks," in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2005, pp. 224–229.
- [50] M. Hofmann, P. Tiefenbacher, and G. Rigoll, "Background segmentation with feedback: The pixel-based adaptive segmenter," in *Proc. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, 2012, pp. 38–43.
- [51] M. Braham, S. Piérard, and M. Van Droogenbroeck, "Semantic background subtraction," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 4552–4556.
- [52] S. Bianco, G. Ciocca, and R. Schettini, "Combination of video change detection algorithms by genetic programming," *IEEE Trans. Evol. Comput.*, vol. 21, no. 6, pp. 914–928, Dec. 2017.
- [53] S.-H. Lee, G.-C. Lee, J. Yoo, and S. Kwon, "WisenetMD: Motion detection using dynamic background region analysis," *Symmetry*, vol. 11, no. 5, p. 621, May 2019.
- [54] P.-L. St-Charles, G.-A. Bilodeau, and R. Bergevin, "Universal background subtraction using word consensus models," *IEEE Trans. Image Process.*, vol. 25, no. 10, pp. 4768–4781, Oct. 2016.
- [55] S. Jiang and X. Lu, "WeSamBE: A weight-sample-based method for background subtraction," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 9, pp. 2105–2115, Sep. 2018.
- [56] P.-L. St-Charles, G.-A. Bilodeau, and R. Bergevin, "SuBSENSE: A universal change detection method with local adaptive sensitivity," *IEEE Trans. Image Process.*, vol. 24, no. 1, pp. 359–373, Jan. 2015.
- [57] M. O. Tezcan, P. Ishwar, and J. Konrad, "BSUV-Net: A fully-convolutional neural network for background subtraction of unseen videos," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2020, pp. 2763–2772, doi: 10.1109/WACV45572.2020.9093464.
- [58] C.-Y. Wang, A. Bochkovskiy, and H.-Y.-M. Liao, "Scaled-YOLOv4: Scaling cross stage partial network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 13029–13038.
- [59] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, Jun. 2006.
- [60] S. Kim, S. Kwak, and B. C. Ko, "Fast pedestrian detection in surveillance video based on soft target training of shallow random forest," *IEEE Access*, vol. 7, pp. 12415–12426, 2019.



KAHLIL MUCHTAR (Senior Member, IEEE) received the B.S. degree in informatics from the School for Engineering of PLN's Foundation (STT-PLN), Jakarta, Indonesia, in 2007, the M.S. degree in computer science and information engineering from Asia University, Taichung, Taiwan, in 2012, and the Ph.D. degree in electrical engineering from the National Sun Yat-sen University (NSYSU), Kaohsiung City, Taiwan. From 2018 to 2020, he was involved in a startup company as an AI Research Scientist at Nodeflux, Jakarta. From 2019 to 2021, he was appointed as the Chairperson of the Telematics Research Center (TRC), Universitas Syiah Kuala, Banda Aceh, Indonesia. Since October 2021, he has been appointed as the Head of the Computer Engineering Bachelor Program. He is currently an Assistant Professor with the Department of Electrical, and Computer Engineering, Universitas Syiah Kuala. His research interests include computer vision and image processing. He received the 2014 IEEE GCCE Outstanding Poster Award and IICM Taiwan 2017 The Best of Ph.D. Dissertation Award. He served as the Publication Chair for IEEE ICILTICs 2017 and 2018. In 2021, he served as the General Chair for the IEEE IC-COSITE.



AL BAHRI (Member, IEEE) received the bachelor's and master's degrees from the School of Electrical Engineering and Informatics, Bandung Institute of Technology (Institut Teknologi Bandung), in 2014 and 2018, respectively. Since March 2018, he has been with the Department of Electrical Engineering, Faculty of Engineering, Syiah Kuala University. He had conducted some research, such as Question Bank of SeaCyberclass in 2013, the Geography Information System of School and Hospital in Bandung, and the Implementation of Shuffling, Fisher Yate Algorithm, in an e-learning system. His research interests include digital media technology, game, and computer.



she has been a Lecturer and a member of the Electrical and Computer Engineering Department, Universitas Syiah Kuala. Her research interests include human-computer interaction, interactive visualization, and software engineering.

MAYA FITRIA (Member, IEEE) received the bachelor's degree in computer science from the Universitas Indonesia (UI), in 2012, and the Master of Science degree in computer engineering, in 2016. In 2013, she continued her study with the Computer Engineering Department, University of Duisburg-Essen, Germany, with expertise in the field of interactive system and visualization. During the study, she got a support from DAAD-LPSDM Aceh Scholarship. Since 2017,



TJENG WAWAN CENGGORO (Member, IEEE) received the bachelor's degree in information technology from STMIK Widya Cipta Dharma, and the master's degree in information technology from Bina Nusantara University. He is currently an AI Researcher whose focus is in the development of deep learning algorithms for application in computer vision, natural language processing, and bioinformatics. He is also a Certified Instructor with the NVIDIA Deep Learning Institute. He led

several research projects that utilize deep learning for computer vision, which is applied to indoor video analytics and plant phenotyping. He has published over 20 peer-reviewed publications and reviewed for prestigious journals, such as *Scientific Reports* and IEEE Access. He also holds two copyrights for AI-based video analytics software.



BENS PARDAMEAN (Member, IEEE) received the bachelor's degree in computer science and the master's degree in computer education from California State University, Los Angeles, CA, USA, and the Ph.D. degree in informative research from the University of Southern California (USC). He has over 30 years of global experience in information technology, bioinformatics, and education. After successfully leading the Bioinformatics Research Interest Group, he currently holds a

dual appointment as the Director of the Bioinformatics and Data Science Research Center (BDSRC), and an Associate Professor of computer science with Bina Nusantara University, Jakarta, Indonesia.



ADHIGUNA MAHENDRA received the M.S. degree in computer vision and robotics from the University of Heriot-Watt, U.K., in 2008, and the Ph.D. degree in machine learning and computer vision from the Universite de Dijon, France, in 2012. He is currently a Lecturer of data science and enterprise architecture in the Master of Information Technology with Swiss German University. He is also a Lecturer of business analytics in the MBA Program at Central Queensland University.

He is also active in the industry with over 20 years of experience building intelligent systems based on AI and machine learning for global and national companies in Europe, Singapore, and Indonesia in verticals, such as oil and gas, industrial automation, aviation, logistics, smart-city, and B2B eCommerce. He is also the Chief of AI and product with Nodeflux, a leading AI Vision company in Indonesia, leading the implementation of AI products, such as the video analytics platform, biometrics eKYC platform, and retail visual analytics platform from the product design, algorithm development, operationalization (MLOps), and commercialization. He has publications in SPIE and IEEE and served as a Reviewer for the International Conference on Engineering and Information Technology for Sustainable Industry (ICONETSI), in 2020. He received the Best Lecturer Award from Swiss German University in 2018.



MUHAMMAD RIZKY MUNGGARAN received the M.T. degree in business intelligence from the School of Electronic Engineering and Informatics, Institute of Technology Bandung, in August 2016. His current research interests include information retrieval, artificial intelligence, computer vision, and business intelligence. In career and experience, he has been active in several machine learning projects related for over seven years.

He built machine learning systems on healthy systems using signal processing, temporal data analysis, text mining, and vision on hand writing recognition, face recognition using biometrics. He also have intellectual property rights (HAKI) for an application to reads and extracts the values of archive graphs image (Graph Digitizer) was joined development with the Indonesian Agency for Meteorological, Climatological and Geophysics (Badan Meteorologi, Klimatologi, and Geofisika or simply BMKG) in 2020. He focus on computer vision and artificial intelligence and has been joining with Nodeflux since six years ago, a start-up company who provides video analytics platform, biometrics eKYC platform, and retail visual analytics platform from the product design, algorithm development, operationalization (MLOps), and leading AI Vision company in Indonesia.



CHIH-YANG LIN (Senior Member, IEEE) received the Ph.D. degree in computer science and information engineering from the National Chung Cheng University, Chiayi, Taiwan, in 2006. He was with the Advanced Technology Center, Industrial Technology Research Institute, Taiwan, from 2007 to 2009. He was a Postdoctoral Fellow with the Institute of Information Science, Academia Sinica, in 2009. He joined Asia University, Taichung, Taiwan, in 2010, where he is

currently an Assistant Professor and then became an Associate Professor, in 2013. He was the Chair of the Department of Bioinformatics and Medical Engineering, from August 2014 to January 2017. He is also a Professor with the Department of Electrical Engineering, Yuan-Ze University, Taoyuan, Taiwan. He has authored or coauthored over 100 articles and patents. His research interests include computer vision, machine learning, image processing, and the design of surveillance systems. He has served as a Session Chair, Publication Chair, or Workshop Organizer on many international conferences, including AHFE, ICCE, ACCV, IEEE Multimedia Big Data, ACM IH&MMSec, APSIPA, and CVGIP.

...