ISSN: 2278-4632
Vol-14, Issue-1, January: 2024

FUSION OF MULTI-INTENSITY IMAGE FOR DEEP LEARNING-BASED HUMAN AND FACE DETECTION

Kasidi Sumith Reddy[1], DR. M. Ramchander[2]
[1]MCA Student, Chaitanya Bharathi Institute of Technology (A), Gandipet, Hyderabad, Telangana State, India.
[2]Assistant Professor, Department of MCA, Chaitanya Bharathi Institute of Technology (A),Gandipet, Hyderabad, Telangana State, India.

ABSTRACT: For ordinary IR-illuminators in nighttime surveillance systems, insufficient illumination may cause misdetection for faraway objects while excessive illumination leads to over-exposure of nearby object. To overcome these two problems, we use the MI3 image dataset, which is established by multi-intensity IR-illumination (MIIR), as our benchmark dataset for modern object detection methods. We first provide complete annotations for the MI3 as its current ground-truth is incomplete. Then, we use these multi-intensity illuminated IR videos to evaluate several widely used object detectors, i.e., SSD, YOLO, Faster R-CNN, and Mask R-CNN, by analyzing the effective range of different illumination intensities. By including a tracking scheme, as well as developing of a new fusion method for different illumination intensities to improve the performance, the proposed approach may serve as a new benchmark of face and object detection for a wide range of distances.

Keywords *–SSD, YOLO, Faster R-CNN, and Mask R-CNN*

1. INTRODUCTION

In nighttime video surveillance, difficulties usually arise from the variation of environmental light. It is hard to detect invaders at far distance under poor lighting conditions, while it is also hard to recognize objects at near distance due to overexposure under strong light. To help solving both the underexposure and overexposure problems simultaneously, multi-intensity IR-illuminator is developed in [1] to provide periodically varying illumination intensity. Subsequently, Chan et al. [2] established the MI3 database, which contains brightness-varying video sequences of several indoor and outdoor scenes. Two kinds of ground-truths are provided, i.e., people counting and the labeling of foreground image pixels, which do not include any bounding box information. Although MI3 exhibits promising results, they still require strong assumptions, e.g., no foreground in the first 100 frames. In addition, the foreground ground-truths provided in MI3 dataset often merge multiple objects together, e.g., a bag cannot be separated from the person carrying it, while some ground-truths are incomplete or questionable.

ISSN: 2278-4632
Vol-14, Issue-1, January: 2024

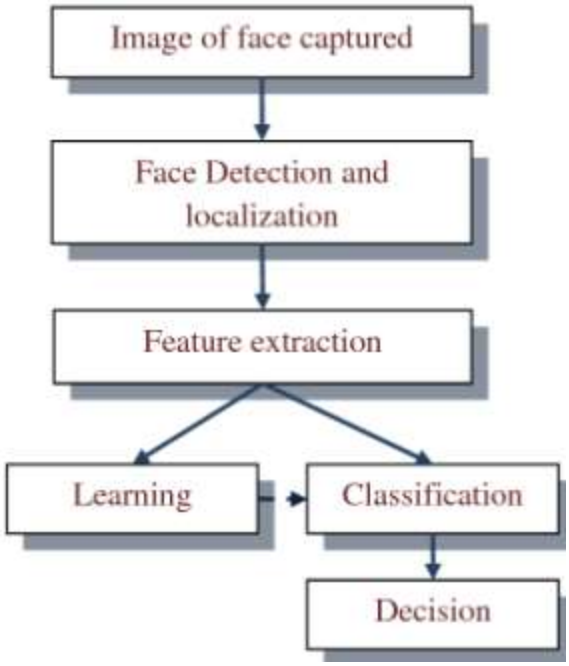

Fig.1: Example figure

In [3]–[5], Gaussian Mixture Model (GMM) is employed for foreground (object) detection in multi-intensity IR videos. However, such approach is usually incapable of dealing with complicated foreground reliably. On the other hand, these previous works only demonstrate qualitatively that better image quality of far (near) objects can be captured with high (low) intensity levels with multi-intensity illumination. Accordingly, quantitative evaluation of such complementary effect among videos of different illumination intensities, called channels, will also be developed in this paper. Following the evergrowing trends of exploring deep learning for object detection, we will adopt MI3 as a benchmark dataset to evaluate such object detectors for selected scenes and illuminations.

2. LITERATURE REVIEW

MI3: Multi-intensity infrared illumination video database:

Vision-based video surveillance systems have gained increasing popularity. However, their functionality is substantially limited under nighttime conditions due to the poor visibility caused by improper illumination. Equipped on night vision cameras, ordinary infrared (IR) illuminators of fixed-intensity usually lead to the imaging problem of overexposure (or underexposure) when the object is too close to (or too far from) the camera. To overcome this limitation, we use a novel multi-intensity IR illuminator to extend the effective range of distance of camera surveillance, and establish in this paper the MI3 (Multi-Intensity Infrared Illumination) database based on such an illuminator. The database contains intensity varying video sequences of several indoor and outdoor scenes. Ground truths including people counting and foreground labelling are provided for different research usages. Performances of related algorithms are tested for demonstration and evaluation.

Intelligent nighttime video surveillance using multi-intensity infrared illuminator:

In nighttime video surveillance, the image details of far objects are often hard to be identified due to poor illumination conditions while the image regions of near objects may be whitened due to overexposure. To alleviate the two problems simultaneously for nighttime video surveillance, we adopt a new multi-intensity infrared illuminator as a supportive light source to provide multiple illumination levels periodically. By using the illuminator with multiple degrees of illumination power, both far and near objects can be clearly captured. For automatic

ISSN: 2278-4632
Vol-14, Issue-1, January: 2024

detection of foreground objects at different distances in the image sequences captured with the multi-intensity infrared illuminator, two foreground object detection methods are proposed in this paper. Experimental results show that the two methods both achieve >90% accuracy in average in foreground object detection while giving different computational complexities.

Robust license plate detection in nighttime scenes using multiple intensity IR-illuminator:

The functionality of video surveillance is significantly degraded by the low illumination and poor visibility under the nighttime environment. However, the demand for nighttime surveillance is no less than the daytime one because of the high incidence of accidents during night. The Infrared (IR) light source with fixed intensity works for only certain distance, resulting in the defect of underexposure/overexposure due to the object being too far from/close to the light source. In this paper an innovative idea is brought up that we use a multiple intensity IR-illuminator to enhance the effective distance of license plate detection. Based on the stroke width of the license ID, license plates are detected in the images under different illuminations and then the results are integrated into a synthesized high dynamic range image, in which the license plate regions and the background scene can be better visualized. Experimental results show that the proposed approach can effectively enlarge the monitored area in both depth and width, as well as enhance the security level of nighttime video surveillance.

A novel video summarization method for multi-intensity illuminated infrared videos:

In nighttime video surveillance, proper illumination plays a key role for the image quality. For ordinary IR-illuminators with fixed intensity, faraway objects are often hard to identify due to insufficient illumination while nearby objects may suffer from over-exposure, resulting in image foreground/background of poor quality. In this paper we proposed a novel video summarization method which utilizes a novel multi-intensity IR-illuminator to generate images of human activities with different illumination levels. By adopting GMM-based foreground extraction procedure for images acquired for each illumination level, foreground objects with most plausible quality can be selected and merged with a preselected representation for still background. The result brings out a reasonable video summary for moving foreground, which is generally unachievable for nighttime surveillance videos.

Faster R-CNN: Towards realtime object detection with region proposal networks:

State-of-the-art object detection networks depend on region proposal algorithms to hypothesize object locations. Advances like SPPnet and Fast R-CNN have reduced the running time of these detection networks, exposing region proposal computation as a bottleneck. In this work, we introduce a Region Proposal Network (RPN) that shares full-image convolutional features with the detection network, thus enabling nearly cost-free region proposals. An RPN is a fully convolutional network that simultaneously predicts object bounds and objectness scores at each position. The RPN is trained end-to-end to generate high-quality region proposals, which are used by Fast R-CNN for detection. We further merge RPN and Fast R-CNN into a single network by sharing their convolutional features---using the recently popular terminology of

ISSN: 2278-4632
Vol-14, Issue-1, January: 2024

neural networks with 'attention' mechanisms, the RPN component tells the unified network where to look. For the very deep VGG-16 model, our detection system has a frame rate of 5fps (including all steps) on a GPU, while achieving state-of-the-art object detection accuracy on PASCAL VOC 2007, 2012, and MS COCO datasets with only 300 proposals per image. In ILSVRC and COCO 2015 competitions, Faster R-CNN and RPN are the foundations of the 1st-place winning entries in several tracks.

3. METHODOLOGY

Many deep learning-based schemes have been developed for object detection in recent years, which significantly push forward the state-of-the-art. In general, object detectors can be categorized into two-stage detectors and single-stage detectors. The former adopt selective search to generate region proposals as in Faster R-CNN, while Mask R-CNN added a branch from Faster R-CNN to achieve promising results of instance segmentation and object detection. On the other hand, single-stage object detectors such as YOLO and SSD do not have a region cropping module. They are simpler and faster than two-stage detectors, but have trailed behind in detection accuracy.

Disadvantages:

1. For ordinary IR-illuminators in nighttime surveillance system, insufficient illumination may cause misdetection
2. For faraway object while excessive illumination leads to over-exposure of nearby object.

In this paper, we will consider single-stage detectors such as SSD and YOLOv4, and two-stage ones such as Faster RCNN and Mask R-CNN, in the experiments. As different applications use infrared images in quite different ways, it is not possible to establish a universal IR dataset; therefore, credibly pretrained model of the above detectors are experimented on the MI3 dataset to setup a baseline for quantitative evaluation of the effect of adopting the multi-intensity illumination. For example, examination of confidence value of deep learning-based object detection may suggest the number of illumination intensities required for object detection for an extended range of distance. Moreover, we may also identify an effective range wherein reasonable detection results can be achieved with one or more illumination intensities of the multi-intensity IR illuminator

Advantages:

1. A tracking method is presented for refining face detection results to increase the F-measure of face detection.
2. A fusion method is proposed to effectively merge information obtained from multiple channels to achieve higher accuracy in object/face detection.

ISSN: 2278-4632
Vol-14, Issue-1, January: 2024

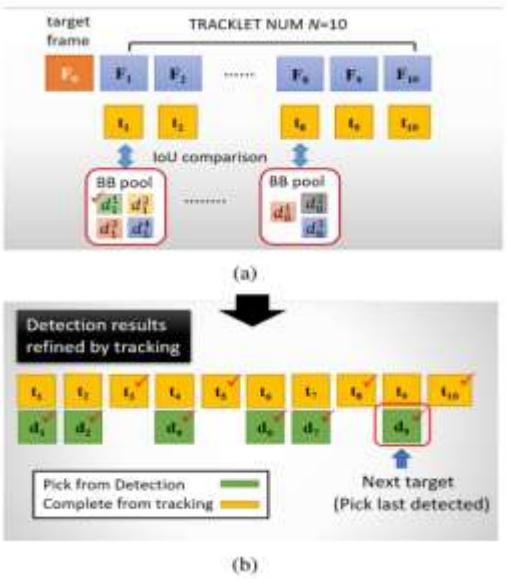

Fig.2: System architecture

MODULES:

In this project we have designed following modules

- Data exploration: using this module we will load data into system
- Processing: Using the module we will read data for processing
- Splitting data into train & test: using this module data will be divided into train & test
- Model generation: Building the model in colab - YOLOV5 - YoloV8 - YoloV3 - MaskRCNN - FasterRCNN - SSD
- User signup & login: Using this module will get registration and login
- User input: Using this module will give input for prediction
- Prediction: final predicted displayed

4. IMPLEMENTATION

YOLOV5 – YOLO is an acronym that stands for You Only Look Once. We are employing Version 5, which was launched by Ultralytics in June 2020 and is now the most advanced object identification algorithm available. It is a novel convolutional neural network (CNN) that detects objects in real-time with great accuracy. This approach uses a single neural network to process the entire picture, then separates it into parts and predicts bounding boxes and probabilities for each component. These bounding boxes are weighted by the expected probability. The method “just looks once” at the image in the sense that it makes predictions after only one forward propagation run through the neural network. It then delivers detected items after non-max suppression (which ensures that the object detection algorithm only identifies each object once).

YoloV8 - Ultralytics YOLOv8 is the latest version of the YOLO object detection and image segmentation model. As a cutting-edge, state-of-the-art (SOTA) model, YOLOv8 builds on the success of previous versions, introducing new features and improvements for enhanced performance, flexibility, and efficiency. YOLOv8 is designed with a strong focus on speed, size, and accuracy, making it a compelling choice for various vision AI tasks. It outperforms previous versions by incorporating innovations like a new backbone network, a new anchor-free split head, and new loss functions. These improvements enable YOLOv8 to deliver superior results, while maintaining a compact size and exceptional speed. Additionally, YOLOv8 supports a full range of vision AI tasks, including detection, segmentation, pose estimation, tracking, and classification. This versatility allows

ISSN: 2278-4632
Vol-14, Issue-1, January: 2024

users to leverage YOLOv8's capabilities across diverse applications and domains.

YoloV3 – YOLOv3 (You Only Look Once, Version 3) is a real-time object detection algorithm that identifies specific objects in videos, live feeds, or images. The YOLO machine learning algorithm uses features learned by a deep convolutional neural network to detect an object. Versions 1-3 of YOLO were created by Joseph Redmon and Ali Farhadi, and the third version of the YOLO machine learning algorithm is a more accurate version of the original ML algorithm.

MaskRNN - Mask R-CNN is a state of the art model for instance segmentation, developed on top of Faster R-CNN. Faster R-CNN is a region-based convolutional neural networks [2], that returns bounding boxes for each object and its class label with a confidence score. To understand Mask R-CNN, let's first discus architecture of Faster R-CNN that works in two stages:

Stage1: The first stage consists of two networks, backbone (ResNet, VGG, Inception, etc..) and region proposal network. These networks run once per image to give a set of region proposals. Region proposals are regions in the feature map which contain the object.

Stage2: In the second stage, the network predicts bounding boxes and object class for each of the proposed region obtained in stage1. Each proposed region can be of different size whereas fully connected layers in the networks always require fixed size vector to make predictions. Size of these proposed regions is fixed by using either RoI pool (which is very similar to MaxPooling) or RoIAlign method.

FasterRCNN – Faster R-CNN is a deep convolutional network used for object detection, that appears to the user as a single, end-to-end, unified network. The network can accurately and quickly predict the locations of different objects.

SSD – SSD uses a matching phase while training, to match the appropriate anchor box with the bounding boxes of each ground truth object within an image. Essentially, the anchor box with the highest degree of overlap with an object is responsible for predicting that object's class and its location.

5. EXPERIMENTAL RESULTS

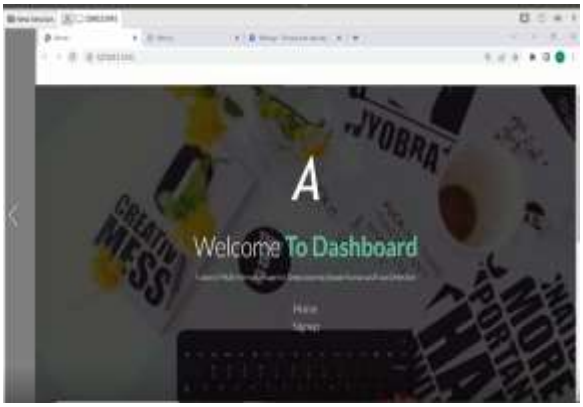

Fig.3: Home screen

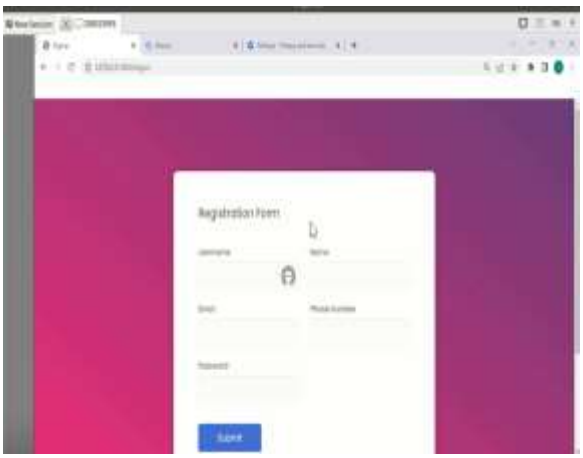

Fig.4: User registration

ISSN: 2278-4632
Vol-14, Issue-1, January: 2024

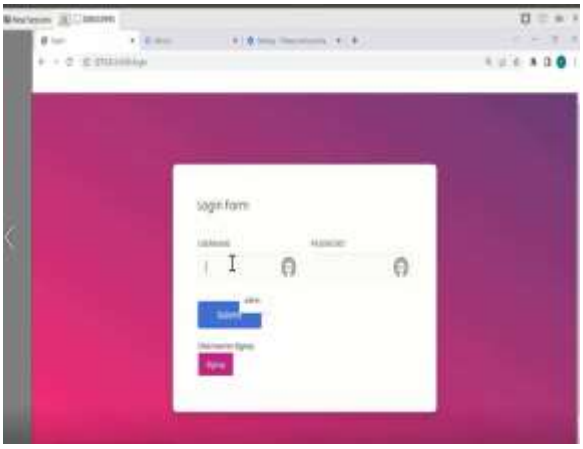

Fig.5: User login

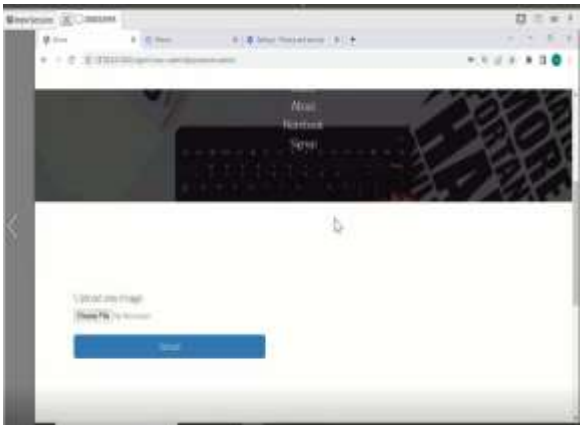

Fig.6: Main page

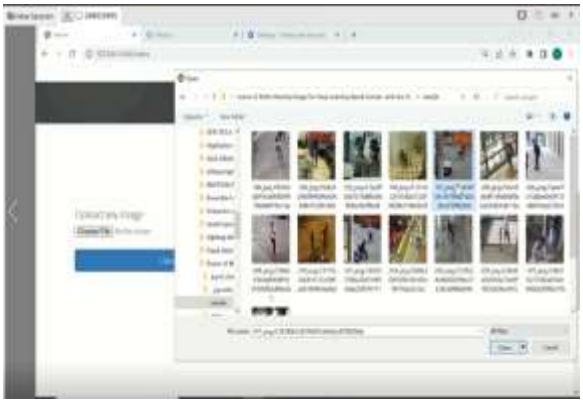

Fig.7: User input

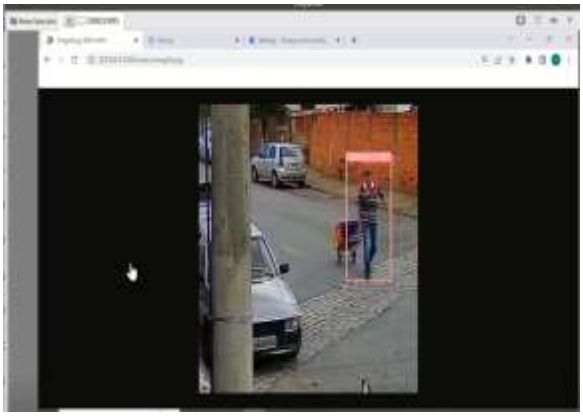

Fig.8: Prediction result

6. CONCLUSION

This work evaluates state-of-the-art human and face detectors and reports their performances on an existing multi-intensity IR illumination dataset, with complete annotations also established for the dataset. To that end, a baseline approach is proposed, which is based on pre-trained CNN detectors, a recently proposed tracker, and simple fusion scheme to take advantage of the complementary effect among different illumination intensities. While satisfactory detection and tracking results are demonstrated in this paper for some simple scenes, further improvements for more complicated datasets, better fusion methods, as well as a systematic way of determining relevant parameters, such as batch size or learning rate for training a specific CNN model, are currently under investigation.

REFERENCES

[1] W. Teng, ‘‘A new design of ir illuminator for nighttime surveillance,’’ M.S. thesis, Dept. Comput. Sci., Nat. Chiao Tung Univ., Hsinchu, Taiwan, 2010.

[2] C.-H. Chan, H.-T. Chen, W.-C. Teng, C.-W. Liu, and J.-H. Chuang, ‘‘MI3: Multi-intensity infrared

illumination video database,'' in Proc. Vis. Commun. Image Process. (VCIP), Dec. 2015, pp. 1–4.

[3] P. J. Lu, J.-H. Chuang, and H.-H. Lin, ''Intelligent nighttime video surveillance using multi-intensity infrared illuminator,'' in Proc. World Congr. Eng. Comput. Sci., vol. 1, 2011, pp. 19–21.

[4] Y.-T. Chen, J.-H. Chuang, W.-C. Teng, H.-H. Lin, and H.-T. Chen, ''Robust license plate detection in nighttime scenes using multiple intensity IR-illuminator,'' in Proc. IEEE Int. Symp. Ind. Electron., May 2012, pp. 893–898.

[5] J.-H. Chuang, W.-J. Tsai, C.-H. Chan, W.-C. Teng, and I.-C. Lu, ''A novel video summarization method for multi-intensity illuminated infrared videos,'' in Proc. IEEE Int. Conf. Multimedia Expo. (ICME), Jul. 2013, pp. 1–6.

[6] S. Ren, K. He, R. Girshick, and J. Sun, ''Faster R-CNN: Towards realtime object detection with region proposal networks,'' IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[7] K. He, G. Gkioxari, P. Dollár, and R. Girshick, ''Mask R-CNN,'' in Proc. IEEE Int. Conf. Comput. Vis., Oct. 2017, pp. 2961–2969.

[8] 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.

[9] J. Redmon and A. Farhadi, ''YOLOv3: An incremental improvement,'' 2018, arXiv:1804.02767.

[10] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, ''YOLOv4: Optimal speed and accuracy of object detection,'' 2020, arXiv:2004.10934.

[11] 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. Cham, Switzerland: Springer, 2016, pp. 21–37.

[12] K. Guo, S. Wu, and Y. Xu, ''Face recognition using both visible light image and near-infrared image and a deep network,'' CAAI Trans. Intell. Technol., vol. 2, no. 1, pp. 39–47, 2017.

[13] S. Cho, N. Baek, M. Kim, J. Koo, J. Kim, and K. Park, ''Face detection in nighttime images using visible-light camera sensors with two-step faster region-based convolutional neural network,'' Sensors, vol. 18, no. 9, p. 2995, Sep. 2018.

[14] H. Jiang and E. Learned-Miller, ''Face detection with the faster R-CNN,'' in Proc. 12th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG), May 2017, pp. 650–657.

[15] H. Nam and B. Han, ''Learning multi-domain convolutional neural networks for visual tracking,'' in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 4293–4302.