Journal of Engineering Research xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Engineering Research

journal homepage: www.journals.elsevier.com/journal-of-engineering-research

Learning deep feature fusion for traffic light detection

Ehtesham Hassan[a,*], Yasser Khalil[b], Imtiaz Ahmad[c]

[a] Kuwait College of Science and Technology, Kuwait
[b] Huawei Research, Canada
[c] Kuwait University, Kuwait

ARTICLE INFO

Keywords:
Object detection
Traffic light detection
YOLO
Handcrafted features
ICF

ABSTRACT

Traffic light detection in real-world conditions is challenging because of the positioning of lights, variety in shapes and scales, and similarity with other objects. The paper presents a deep learning-based traffic light detection system by learning the fusion of handcrafted features. The handcrafted features for object detection focus on specific attributes such as shape, color, or texture. The objective of this work is to incorporate handcrafted features into the network learning process such that the resulting detector parameters are robust to input variations, sensor noise, and atmospheric noise. The proposed detection framework is based on the latest You only look once (YOLO) architecture, trained with the fusion of different information channels in the Integral Channel Features (ICF). The approach demonstrates a qualitative approach for identifying the optimal layer for additional feature injection in the network, and the selection of ICF channels to be applied for fusion. The validation of the proposed detector on the Bosch small traffic light dataset achieved the best mAP score of 55.70 % on the testing set. Further, a qualitative comparison of the proposed detector's performance with that of other recent methods is presented, along with an analysis using auxiliary experiments.

Introduction

A robust traffic light detector plays an important role in many applications, such as intelligent transport systems, autonomous vehicles, and driver assistance systems. The problem has attracted significant attention across research communities [1], nevertheless, the existing state-of-the-art methods do not guarantee perfect detection of all types of traffic lights without false candidates in all possible conditions. Jensen et al. [2] present a rich survey on the different shapes and types of traffic lights across the world, and the associated difficulties in traffic light detection, including environment related issues.

Traffic light detection poses difficulties due to shape variations, occlusion, and viewpoint problems. In such cases, the color and shape features can provide complementary cues for detection. The present work proposes a robust traffic light detector using a deep neural network model by applying the latest version of You only look once (YOLO) [3] which has shown outstanding results in terms of speed and accuracy compared to other detectors. To address the YOLO's problems in small object detection, the presented method enhances the algorithm by fusing handcrafted features, specifically Integral channel features (ICF) [4], with the YOLO's deep learned features. The ICF is selected as it captures the combined information of color channels, gradient magnitude, and gradient histogram. The fusion of deep features and handcrafted features is learned in the network training process by carefully feeding the complementary features. The evaluation of the proposed traffic light detector is shown on the Bosch small traffic lights dataset [5] due to its high precision in labeling traffic light objects. To the best of the available information, the learning-based fusion of deep features with handcrafted features has not been attempted for traffic light detection. The contributions of this work are summarized as follows:

– The paper presents a traffic light detection system using the latest YOLO object detector.
– The system applies a learning-based fusion of handcrafted features with deep features in the YOLO architecture with a novel method to feed in additional features by following a qualitative approach.
– The system is evaluated on the Bosch traffic light dataset with detailed comparative analysis. The testing results demonstrate an increased mAP of 55.70 %.

* Corresponding author.
E-mail address: e.hassan@kcst.edu.kw (E. Hassan).

https://doi.org/10.1016/j.jer.2023.100066
Received 23 October 2022; Received in revised form 23 March 2023; Accepted 30 March 2023

The existing work on traffic light detection with different features is discussed in the next section. An introduction to the ICF, the YOLO, and modifications are discussed in the preliminary section. The subsequent sections present the proposed learning-based feature fusion for traffic light detection, and the evaluation and analysis of the proposed traffic light detector's performance. The final section concludes and presents the perspective of the work.

Related works

The early traffic light detectors were focused on the use of handcrafted features as inputs to classifiers [6,7]. Zhou et al. [8] proposed a traffic light detection and recognition algorithm that uses features extracted from both color and HOG. Similarly, the HOG features were applied in traffic light recognition task in [9,10]. Ji et al. [11] assumed that no traffic light existed below a predefined threshold line. Salarian et al. [12] applied Hue-Saturation-Value (HSV) color-based modeling for detecting the traffic lights. There are also works on the fusion of handcrafted features for traffic light detection combining brightness, color, and blob cues [13]. Ozcelik et al. [14] used a modified AdaBoost classifier as the learning-based detector for traffic light detection using aggregated channel features. Jensen et al. [15] applied a YOLO network architecture-based detector to the public LISA dataset. Behrendt et al. [5] proposed a system for detection and tracking of traffic lights in real-time using the first version of YOLO detection [16] with the ability to detect traffic lights with a small width. Pon et al. [17] proposed a hierarchical architecture based on a modified Faster R-CNN that detects both traffic lights and sign labels. Müller and Dietmayer [18] used the modified Single shot multi box detector for traffic light detection. The fusion of handcrafted features in deep learning networks has also been attempted in many other applications. Nguyen et al. [19] combined handcrafted features with deep features to build a face recognition system. Nie et al. [20] applied fusion in deep neural network for mitosis detection and recognition. Similarly, Georgescu et al. [21] learned the combination of deep features with Bag-of-visual-words for face expression recognition.

The approach presented here demonstrates the latest YOLO detector with a fusion of handcrafted features for traffic light detection tasks. Handcrafted features are designed to encode domain knowledge. Deep features are robust enough to deal with varying scenarios. Therefore, the objective is to learn the optimal fusion of ICF with deep features, which is expected to improve the detector's performance. All evaluation experiments are conducted on the public Bosch traffic light dataset, as it has the best labeling accuracy and high-resolution images compared to other traffic light datasets.

Preliminaries

A brief discussion on the ICF and YOLOv3's architecture is as follows.

Integral channel features (ICF)

ICF are a series of image channels computed from the input image using linear and non-linear transformations. A channel refers to a representation of the input image. The evaluated channels were color channels (RGB, Gray, HSV, and LUV), gradient magnitude, and gradient histogram. The proposed work reevaluates the combination of channels because of the varying challenges in the present application scenario. The window size parameter for HOG [22] computation depends on the 2D dimensions of the deep features to be fused with. For example, if the 2D dimensions of a layer where features are to be fused are 13 × 13 and the input image is 416 × 416 in size, then the window size will be 32 × 32 to generate 13 × 13 HOGs. For each window, the HOG is computed for six bins. Other HOG parameters, including cell and block sizes used for normalization purposes, are also set equal to the window size. The block stride is set equal to the window size with no overlap.

YOLOv3 architecture

The YOLOv3 architecture is based on an input image of dimensions 416 × 416. YOLOv3 consists of 107 layers, 75 of which are convolutional. The YOLOv3 predicts bounding boxes at three different scales. The first detection scale is used for detecting large objects, and the second and third are for medium and small sized objects, respectively. Each detection scale predicts a 3D tensor encoding the bounding box coordinates, objectness score, and class predictions. Other than convolutional layers, YOLOv3 includes residual layers that combine feature maps from two layers using element-wise addition. In addition, the up-sample layer provides more semantic information. The architecture also uses a route layer to produce feature maps, either by bringing forward another layer's features or by computing them using features from two different layers.

Modifications in YOLOv3 architecture

YOLOv3 is fast and treats detection as a single regression problem, reducing the complexity of the design. Though is has significant speed improvement, it lags in terms of accuracy with small objects. The network is further modified following the suggestions in AlexeyAB [23] to improve detection accuracy on small objects. These modifications are mainly targeting the third detection layer, as it is dedicated to detecting small objects. First, the up-sampling factor of the 97th layer is updated from 2 to 4. Second, the routing of the 11th layer instead of 36th layer at the 98th layer. The modifications are shown in Fig. 1. The detection scale of the original YOLOv3 architecture detects objects at dimensions 52 × 52, with the input image size of 416 × 416. The 52 × 52 dimension is a result of an up-sampling factor of 2 from the second detection layer. After applying the modification at the 97th layer, increasing the up-sampling factor to 4 results in increasing the layer's output dimensions to 104 × 104. Furthermore, the 98th layer output is routed to the nearest layer having 104 × 104 dimensions, i.e., the 11th layer.

Learning based feature fusion for traffic light detection

The proposal for feature fusion starts with identifying a location and space for handcrafted features within the YOLOv3 architecture. Therefore, as a first step, one of the 75 convolutional layers in the network architecture is identified as a location for handcrafted feature fusion. A set of reserved filters in the identified convolutional layer is used to input the handcrafted features. The size of the reserved filters

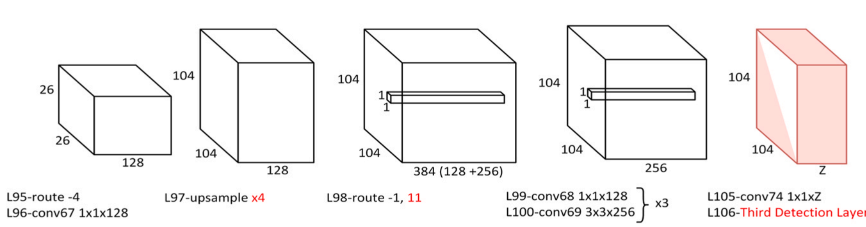

Fig. 1. Modifications applied to YOLOv3's architecture.

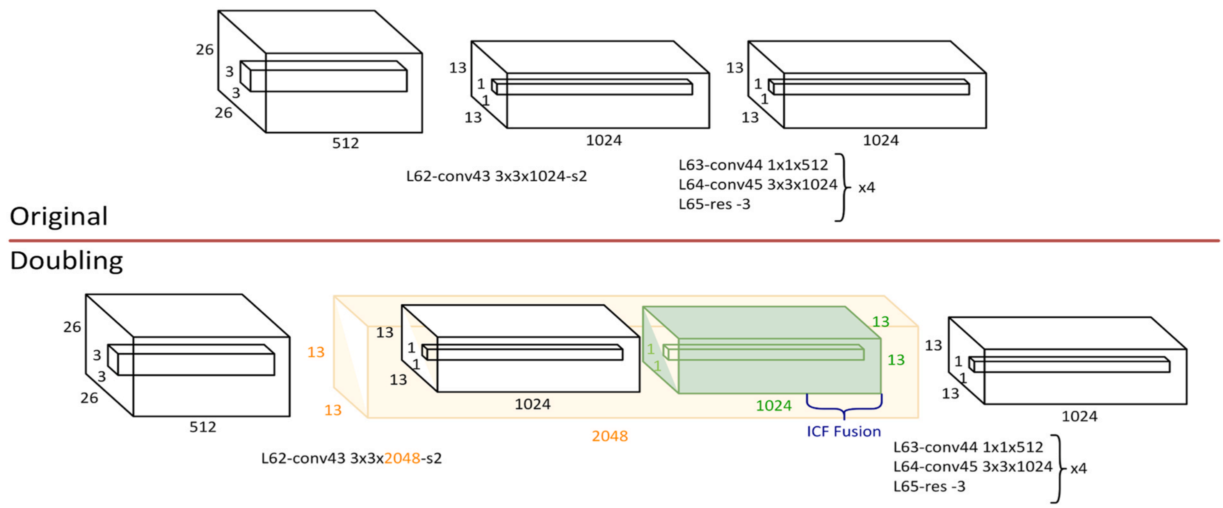

Fig. 2. Example of ICF fusion with deep features.

should be equivalent to the depth of the handcrafted features used for fusion. The actual fusion takes place during learning when the contents of the selected filters (deep features) are replaced with handcrafted features. For storing the handcrafted features, additional filters are stacked onto the original filters, which are created by doubling the depth of the original filters. The depth of the handcrafted features is usually less than the number of the stacked filters. The untouched stacked filters are by default filled with deep features, as the network trains itself. The filter's doubling procedure at a specific layer is shown in Fig. 2. The top side of the figure shows a portion of the original YOLOv3 network, and the bottom side presents the doubled number of filters at the same layer (62nd layer shown as an example). The light orange color represents the new depth and consists of the original number of filters concatenated with additional filters of the same depth shown in green. The figure also highlights the ICF insertion in the portion of the stacked filters. To fuse handcrafted features with deep features, their 2D dimensions must match the width and height dimensions of the layer. For example, if the input image size is 416 × 416 and fusion is to occur at the 62nd layer. In YOLOv3′s architecture, the filter dimensions at the 62nd layer are 13 × 13. Therefore, before the fusion, the selected ICF channels should be 13 × 13 to fit into the stacked filters. In the testing phase, the input image dimensions are fixed throughout the process; hence, the fused ICF dimensions at the 62nd layer are the same throughout. The random parameter in the training phase changes the size of the input image every 10 iterations. Consequently, the size of input changes by a factor of 32, starting from 320 × 320 and going up to 608 × 608. This requires changing ICF dimensions prior to fusion to accommodate all 10 different dimensions.

System evaluation

The evaluation of the proposed traffic light detector is done on the Bosch traffic light dataset which consists of 5000 training and 8334 testing images of 1280 × 720 pixels. The traffic light objects vary approximately from 1 to 85 pixels. The annotations consist of four class labels: “Red”, “Yellow”, “Green”, with “Off” representing the non-operating lights.

Methodology

The mean average precision (mAP) is used for the detector's performance evaluation which is a preferred metric to analyze object detection models [16,17,24]. The metric is computed as the mean of the average precision for all classes. A precision-recall (PR) curve for each class is computed by requiring the detections to have an intersection over union (IoU) greater than the specified threshold. A detection is a true positive if the IoU matches the threshold; otherwise, it is a false positive. After computing the PR curve, the AP per class is the area under the curve. Behrendt et al. [5] use PR curves for measuring the detector's performance on the Bosch dataset with IoU values greater than 0.3 and 0.5. Similarly, Pon et al. [17] use the mAP metric for the detector's evaluation. The proposed detector is also analyzed using the PR curves for detections with IoU ≥ 0.5.

Experiments and results

The initial evaluation of the original YOLOv3 on the Bosch dataset achieved a maximum mAP of 33.60 %. Behrendt et al. [5] provided only PR curves without mAP scores (reproduced in Fig. 3). The blue line in the figure represents the evaluation of all ground truth traffic lights. The yellow and red lines represent the evaluation for only ground truth traffic lights with a width greater than 5px and 10px, respectively. The present evaluation focuses on the solid blue line with respect to all ground truth objections without any exception, where the IoU ≥ 0.5 is used for labeling an object. The approximate mAP for the solid blue line in the figure is computed as 0.72∗0.5 = 0.36 (36 %). Fig. 4 illustrates the approximate area under the solid blue line, where the black region denotes the actual area under the curve, except for the two small regions that were left out. The purple regions are additional areas included to ease the calculation and compensate for the two small, uncounted regions. For the same PR curve, Pon et al. [17] have approximated mAP as 40 %. Evaluation of the Bosch dataset on the original YOLOv3 architecture is expected to give a lower mAP compared to Behrendt et al. [5], as the YOLOv3 is trained on input images of sizes ranging from 320 × 320 to 608 × 608; rather than on the original image dimensions of 1280 × 720. On the other hand, Behrendt et al. [5] trained the YOLOv1 network on several crops of the input images, giving it the advantage of detecting almost all small objects.

Improvement on YOLOv3 architecture

The YOLOv3′s last detection layer is responsible for detecting small objects and contributes most to detecting the traffic lights, compared to the other two detection scales, which are responsible for detecting large and medium sized objects. The modifications discussed earlier on YOLOv3′s network improved the detection when dealing with small objects before proceeding to the feature fusion step, resulting in a mAP score of 41 %. Hence, achieving an improvement of 1 % over the generous approximation computed by Pon et al. [17].

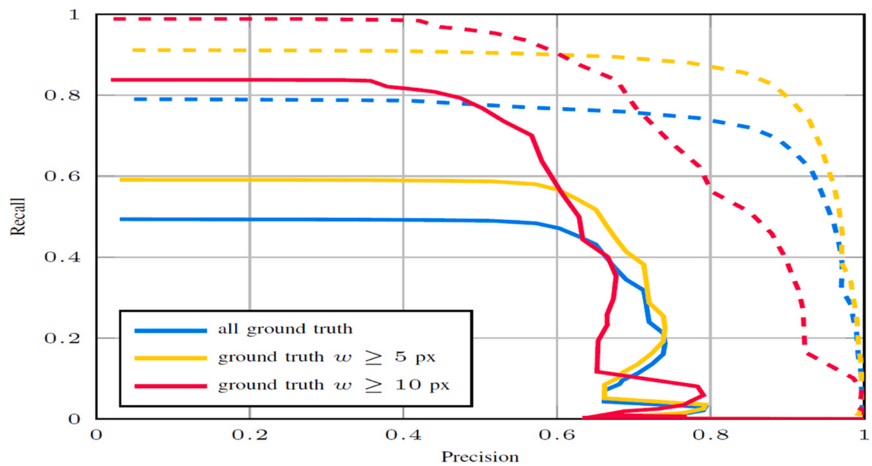

Fig. 3. PR curves from Behrendt et al. [5]. Solid and dashed lines represent IoU ≥ 0.5 and IoU ≥ 0.3 respectively.

Identifying the layer for feature fusion

For learning based feature fusion, the proposed work has two prerequisites:

– Identification of the layer and provide the space for additional features (ICF).
– Best combination of ICF channel features to be fused.

The fusion step starts at the 10th layer, fusing all ICF channels except the RGB. After determining the best layer location that gives the highest mAP score, the experiments determine the best ICF combination. Table 1 and Fig. 5 illustrate the mAP results and PR.

curves for the fusion of 14 ICF channels (gray, HSV, LUV, gradient, and HOG) into the YOLOv3 at different layers. The fusion at the 7th layer achieved the highest mAP of 35.1 %.

Determining the best combination of ICF channels for fusion

Next, the best combination of ICF channels is identified to further improve the detection rate on the Bosch dataset. Table 2 and Fig. 6 present the maximum mAP scores and PR curves for fusing different combinations of ICF channels at the 7th layer.

The results establish that HSV and LUV color channels improve the detector's overall performance. The last two experiments test gradient histograms contribution to traffic light detection. In summary, the best layer position to fuse ICF channels is at the 7th layer, and the best ICF group for fusion is Group C consisting of HSV and LUV color channels, Gradient

Table 1
mAP for fusion of ICF, except RGB, with modified YOLOv3 deep features.

Layer	Number of Filters	Doubled Number of Filters	mAP (%)
5	128	256	27.3
6	64	128	28.7
7	128	256	35.1
9	64	128	32.5
10	128	256	25.6

histograms result in the highest mAP score (47.0 %), which is 7 % higher than the generous mAP approximation of Behrendt et al. [5] computed by Pon et al. [17]. This proves the effectiveness of ICF fusion in improving the detector's performance over merely relying on deep features.

Discussion on detection results

Sample results of the proposed traffic light detector correctly detecting traffic light objects are shown in Figs. 7 and 8. It is observed that all missed detections are even difficult for the human eye to detect. Fig. 9 shows false and missed detections using the modified version of YOLOv3 (top row); however, using the proposed detector they were correctly detected proving the effectiveness of ICF fusion (bottom row). It is noticed that without ICF fusion, the detector confused "Red" and "Yellow". The probable reason why "Red" is harder to detect is due to plenty of background noise coming from car brake lights. Additionally, from a distance it becomes harder to differentiate between "Red" and "Yellow" colors.

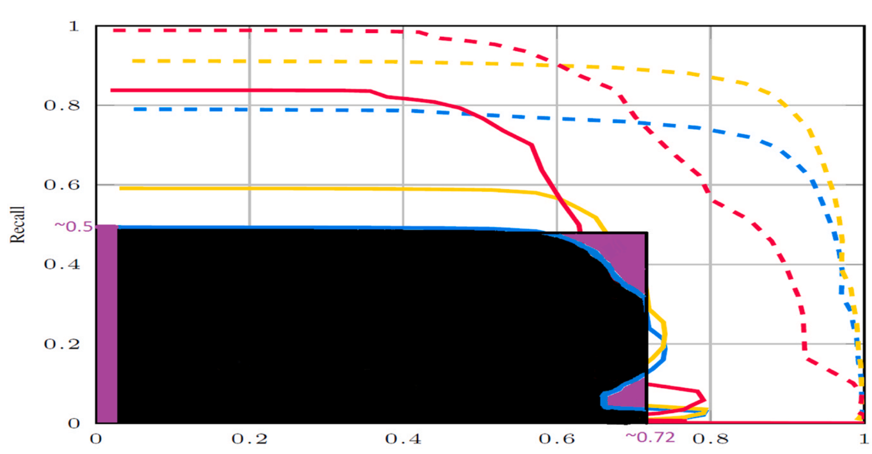

Fig. 4. Computation of area under the solid blue line shown in Fig. 3.

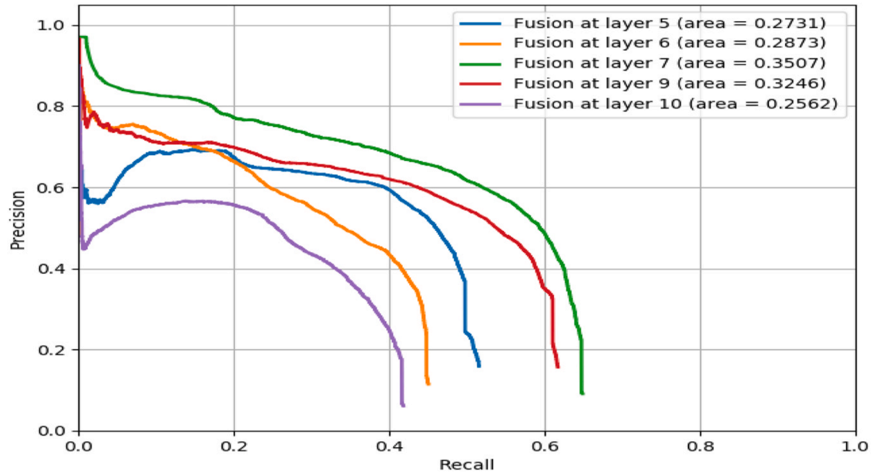

Fig. 5. PR curves for ICF (Gray, HSV, LUV, Grad., and HOG) fusion at different layers.

Table 2
mAP for fusion of different ICF combinations at the 7th layer of modified YOLOv3.

Group	Fused ICF	ICF channels	mAP (%)
A	RGB + Gray + HSV + LUV + Grad. + HOG	17	45.5
B	Gray + HSV + LUV + Grad. + HOG	14	44.8
C	HSV + LUV + Grad. + HOG	13	47.0
D	Gray + LUV + Grad. + HOG	11	42.2
E	LUV + Grad. + HOG	10	43.1
F	Gray + HSV + Grad. + HOG	11	44.0
G	Gray + HSV + LUV	7	45.8
H	Gray + HSV + LUV + HOG	13	45.3

Auxiliary experiments for improving detection rates

It is observed that the fusion of max-normalized handcrafted features results in a 1 % mAP increase over the unnormalized handcrafted features, which is followed in all experiments. The evaluation process also experimented with the positioning of fusion before and after the activation function on the convolutional layer, where ICF fusion before the activation step gave a better mAP score. In addition, the process evaluated the ICF fusion before the batch normalization. Training becomes harder when ICF are fused before batch normalization. As the ICF channels are fused, feature maps now contain both handcrafted and deep features. Next, when batch normalization is applied to the hybrid feature maps, features end up within the standard normal distribution. This transformation changes the contextual information of the fused ICF channels, hence making network training difficult. Behrendt et al. [5] proposed training the detector network on crops of the original input images. The objective is to process the entire image and allow the network to discover by itself the relevant objects, rather than physically alter the input size. This makes it tougher for the detector to process its input images and accurately detect objects.

Benchmarking with recent methods

Pon et al. [17] trained a modified Faster R-CNN on the Bosch dataset based on the original image size. However, the proposed detector is trained on input images resized from 1280 × 720 to a dimension ranging from 320 × 320 to 608 × 608. Furthermore, all extracted ICF features for fusion are computed on even smaller dimensions to match the input layer's dimensions. The best performance of the proposed detector is a mAP of 47.0 %, whereas Pon et al. [17] report a mAP of 54 %. Nevertheless, to prove the robustness of the proposed detector, the dimension of input images is increased from 416 × 416 to 704 × 704 only at the testing phase, without retraining the network. The training uses images on spatial scales ranging from 320 × 320 to 608 × 608. The experiment achieved the best mAP of 55.70 % due to the combination of ICF features fused at the 7th layer. In addition, 704 × 704 is

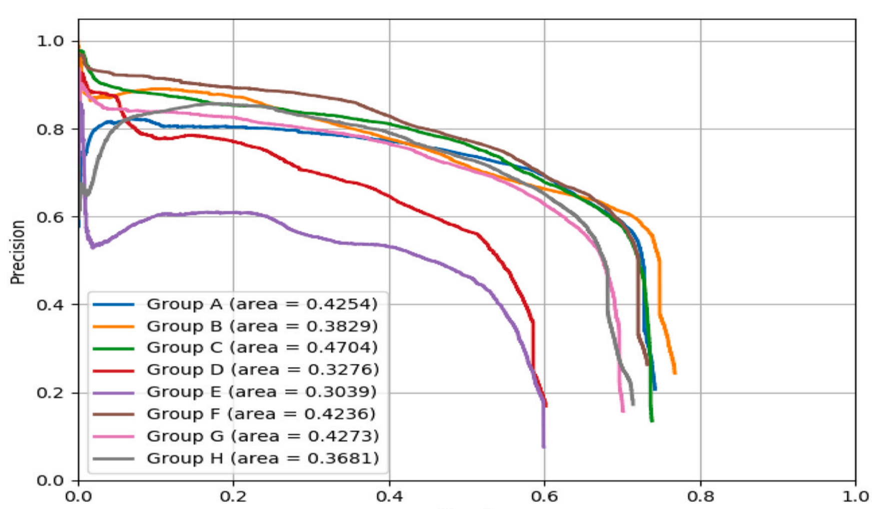

Fig. 6. PR curves for different ICF groups fused at the 7th layer of modified YOLOv3.

Fig. 7. Examples of true detections from the proposed traffic light detector.

Fig. 8. Examples of false and miss detections from the proposed traffic light detector.

Fig. 9. Detections with modified YOLOv3 (left), detections after feature fusion (right).

still smaller than the original input size in the Bosch dataset. The performance would further improve by passing the test images at the original scale. Again, retraining on higher input resolutions is also expected to improve the detector's performance.

Conclusion and future work

The paper presented a novel traffic light detector that learns the fusion of deep features computed by YOLOv3 with ICF for robust and accurate object detection invariant of occlusion, orientation, and shadow problems. The resulting deep features generate robust representations for object detection, addressing the contextual variations and confusions. The evaluation of the proposed detector on the Bosch traffic light dataset achieved the best mAP of 55.7 % with modified YOLOv3 and the ICF constituted with HSV and LUV color channels and gradient magnitude and orientation histograms. A novel empirical strategy is proposed to identify optimal layer for feature fusion with evaluation on a combination of channels from ICF for traffic light detection. The experiments show that performance improves by retraining the network on the original size of the training images. The results also establish that simple resizing of testing images in the proposed detector improves the detection rate from 47.00 % to 55.70 %. Further exploration in this work can focus on the analysis of handcrafted features on wavelets and Fourier transforms in the proposed fusion framework.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] M. Diaz, P. Cerri, G. Pirlo, M.A. Ferrer, D. Impedovo, 2015. A survey on traffic light detection. In: New Trends in Image Analysis and Processing – ICIAP Workshops, pp. 201–208.

[2] M. Jensen, M. Philipsen, A. Møgelmose, T. Moeslund, M. Trivedi, Vision for looking at traffic lights: Issues, survey, and perspectives, IEEE Trans. ITS 17 (7) (2016) 1800–1815.

[3] J. Redmon, A. Farhadi, Yolov3: An incremental improvement. arXiv, 2018. 〈https://doi.org/10.48550/arXiv.1804.02767〉.

[4] P. Dollár, Z. Tu, P. Perona, S. Belongie, Integral channel features, Br. Mach. Vis. Conf. (2009) 1–11, https://doi.org/10.5244/C.23.91

[5] K. Behrendt, L. Novak, R. Botros, A deep learning approach to traffic lights: Detection, tracking, and classification. In: IEEE ICRA, 2017, pp. 1370–1377. 〈https://doi.org/10.1109/ICRA.2017.7989163〉.

[6] R. Charette, F. Nashashibi, Real time visual traffic lights recognition based on spot light detection and adaptive traffic lights templates. In: IEEE IV Symposium, 2009, pp. 358–363.

[7] V. Haltakov, J. Mayr, C. Unger, S. Ilic, 2015. Semantic segmentation based traffic light detection at day and at night. In: Proceedings of the Pattern Recognition: 37th German Conference, Germany, Proceedings, 37, pp. 446–457. 〈https://doi.org/10.1007/978–3-319–24947-6_37〉.
[8] X. Zhou, J. Yuan, H. Liu, Real-time traffic light recognition based on c-hog features, Comput. Inform. 36 (4) (2017) 793–814, https://doi.org/10.4149/cai_2017_4_793
[9] A. Møgelmose, D. Liu, M.M. Trivedi, Detection of us traffic signs, IEEE Trans. Intell. Transp. Syst. 16 (6) (2015) 3116–3125, https://doi.org/10.1109/TITS.2015.2433019
[10] G. Mu, Z. Xinyu, L. Deyi, Z. Tianlei, A. Lifeng, Traffic light detection and recognition for autonomous vehicles, J. China Univ. Posts Telecommun. 22 (1) (2015) 50–56, https://doi.org/10.1016/S1005-8885(15)60624-0
[11] Y. Ji, M. Yang, Z. Lu, C. Wang, Integrating visual selective attention model with hog features for traffic light detection and recognition, IEEE IV Symp. (IV) (2015) 280–285.
[12] M. Salarian, A. Manavella, R. Ansari, A vision based system for traffic lights recognition. In: SAI Intelligent Systems Conference, 2015, pp. 747–753. IEEE. 〈https://doi.org/10.1109/IntelliSys.2015.7361224〉.
[13] S. Aziz, E.A. Mohamed, F. Youssef, Traffic sign recognition based on multi-feature fusion and elm classifier, Procedia Comput. Sci. 127 (2018) 146–153.
[14] Z. Ozcelik, C. Tastimur, M. Karakose, E. Akin, A vision based traffic light detection and recognition approach for intelligent vehicles," International Conference on Computer Science and Engineering (UBMK), Antalya, Turkey, 2017, pp. 424–429. 〈https://doi.org/10.1109/UBMK.2017.8093430〉.
[15] M. Jensen, K. Nasrollahi, T.B. Moeslund, Evaluating state-of-the-art object detector on challenging traffic light data, IEEE CVPR Workshops (2017) 882–888, https://doi.org/10.1109/CVPRW.2017.122
[16] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: unified, real-time object detection, Proc. IEEE CVPR (2016) 779–788, https://doi.org/10.1109/CVPR.2016.91
[17] A.D. Pon, O. Andrienko, A. Harakeh, S.L. Waslander, A. A Hierarchical Deep Architecture and Mini-batch Selection Method for Joint Traffic Sign and Light Detection. 15th Conference on Computer and Robot Vision (CRV), Toronto, Canada, 2018, pp. 102–109. 〈https://doi.org/10.1109/CRV.2018.00024〉.
[18] J. Müller, K. Dietmayer, Detecting traffic lights by single shot detection, arXiv (2018), https://doi.org/10.48550/arXiv.1805.02523
[19] D.T. Nguyen, T.D. Pham, N.R. Baek, K.R. Park, Combining deep and handcrafted image features for presentation attack detection in face recognition systems using visible-light camera sensors, Sensors 18 (2018) 699, https://doi.org/10.3390/s18030699
[20] W. Nie, Y. Yan, T. Hao, C. Liu, Y. Su, Mitosis event recognition and detection based on evolution of feature in time domain, Mach. Vis. Appl. 29 (8) (2018) 1249–1256.
[21] M.I. Georgescu, R.T. Ionescu, M. Popescu, Local learning with deep and handcrafted features for facial expression recognition, arXiv (2018), https://doi.org/10.48550/arXiv.1804.10892
[22] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection. In: IEEE Computer Society CVPR, 2005. CVPR 2005. vol. 1, 2005, pp. 886–893. IEEE. 〈https://doi.org/10.1109/CVPR.2005.177〉.
[23] AlexeyAB, Windows and linux version of darknet yolo v3 & v2 neural networks for object detection (6), 2017. 〈https://github.com/AlexeyAB/darknet〉. Published in (2017).
[24] F. Doğan, I. Turkoğlu, Comparison of deep learning models in terms of multiple object detection on satellite images, J. Eng. Res. 10 (3A) (2022) 89–108, https://doi.org/10.36909/jer.12843