applied sciences

Article

Traffic Light and Arrow Signal Recognition Based on a Unified Network

Tien-Wen Yeh [1], Huei-Yung Lin [1,2] and Chin-Chen Chang [3,*]

[1] Department of Electrical Engineering, National Chung Cheng University, Chiayi 621, Taiwan; tw950107@gmail.com (T.-W.Y.); lin@ee.ccu.edu.tw (H.-Y.L.)
[2] Advanced Institute of Manufacturing with High-Tech Innovations, National Chung Cheng University, Chiayi 621, Taiwan
[3] Department of Computer Science and Information Engineering, National United University, Miaoli 360, Taiwan
[*] Correspondence: ccchang@nuu.edu.tw

Abstract: We present a traffic light detection and recognition approach for traffic lights that utilizes convolutional neural networks. We also introduce a technique for identifying arrow signal lights in multiple urban traffic environments. For detection, we use map data and two different focal length cameras for traffic light detection at various distances. For recognition, we propose a new algorithm that combines object detection and classification to recognize the light state classes of traffic lights. Furthermore, we use a unified network by sharing features to decrease computation time. The results reveal that the proposed approach enables high-performance traffic light detection and recognition.

Keywords: autonomous vehicle; computer vision; traffic light recognition; convolutional neural networks

check for updates

Citation: Yeh, T.-W.; Lin, H.-Y.; Chang, C.-C. Traffic Light and Arrow Signal Recognition Based on a Unified Network. *Appl. Sci.* **2021**, *11*, 8066. https://doi.org/10.3390/app11178066

Academic Editor: Darko Babić

Received: 1 August 2021
Accepted: 30 August 2021
Published: 31 August 2021

1. Introduction

Advanced driver assistance systems currently installed in vehicles, such as autonomous driving vehicles, have achieved favorable results. These systems greatly reduce the necessity of driver control in locations with invariable landscapes, such as highways and enclosed parks. They are also capable of planning routes and navigating vehicles to destinations while helping them to avoid obstacles. However, for self-driving technology to be more widely applicable, additional landscapes (e.g., city streets) must be incorporated into these systems. Specifically, self-driving technology must be able to detect traffic conditions and determine whether to continue driving or stop according to traffic signals. Traffic light detection techniques are typically categorized into two classes. One class involves the detection of traffic lights and communicating with neighboring vehicles through vehicle-to-infrastructure (V2I) communications [1]. The second class involves the detection of traffic light positions and states by using vehicles' sensors. The first class is usually more expensive to implement than the second class.

Traffic light recognition methods that use vehicle onboard sensors have been extensively studied [2]. In several methods, advanced image processing techniques are primarily applied to the sequence images captured by in-vehicle cameras. Learning-based methods have become increasingly popular because of their excellent classification performance [3–5]. However, detection accuracy remains unsatisfactory due to the presence of multiple disturbance factors in outdoor environments, such as incomplete light shapes, dark light states, and partial occlusion. These issues are troublesome to overcome with computer vision and image processing techniques. However, the research on convolutional neural networks (CNN) for traffic light detection [6,7] has contributed to the development of learning-based methods with effective feature extraction for classification.

Conventional image-based traffic light detection methods can misjudge traffic situations when background features that resemble traffic lights are detected. Traffic light detection requires high accuracy because it affects subsequent vehicle control decisions. Accordingly, map-based detection methods have been proposed as a supplement to image-based detection methods, where the aim is to reduce misjudgment and enhance accuracy.

In this paper, we propose a traffic light recognition approach for traffic lights using deep neural networks. Our approach focuses on the detection of arrow signal lights. For traffic light detection, we use map data to facilitate detection by restricting the region of interest (ROI). We use two cameras with different focal lengths to capture nearby and faraway scenes. For recognition, we propose a technique that combines object detection and classification. In addition, we propose a unified network by sharing features to decrease training and computation. The results demonstrate the effectiveness of the proposed method in detecting traffic lights.

The contributions of the proposed approach are as follows: (i) using map information and two various focal length cameras for traffic light detection at different distances, (ii) proposing a technique that combines object detection and classification to solve the issue of multiple light state classes, and (iii) integrating the network for detecting traffic lights to a unified network by sharing feature maps for efficiency.

2. Related Works

Conventional image-based techniques for detecting traffic lights mostly utilized computer vision algorithms [8]. Captured images were first transformed into multiple color spaces. Features were then extracted for detection. In machine learning-based approaches [9], image features such as the histogram of oriented gradients or Harr-like operators were used for support vector machine (SVM) or adaptive boosting (AdaBoost) classification techniques. Fregin et al. [10] presented a traffic light detection method that integrates the depth information obtained using a stereo camera. Müller et al. [11] presented a dual-camera system that uses multiple focal length settings to expand the extent for detecting traffic lights. They used a camera with long focal length for faraway traffic light detection and a camera with a wide-angle lens for nearby traffic light detection.

Several techniques that compute the positions of traffic lights using deep neural networks have been extensively studied. Weber et al. [6,12] proposed the DeepTLR and HDTLR techniques, which use CNN for traffic light detection and classification. Sermanet et al. [13] proposed an approach for object detection, recognition, and localization. They introduced a multi-scale method with a sliding window. It can be efficiently performed in a CNN. Recently, several networks for object detection have been utilized for detecting traffic lights. For example, Behrendt et al. [14] presented an approach that uses the You Only Look Once (YOLO) framework [15] to detect traffic lights. Traffic lights sometimes appeared as small elements in images, and a common solution to this problem was to decrease the stride of a neural network for feature preservation. Müller and Dietmayer [16] used the single-shot multi-box detector method [17] and focused on small traffic light detection. Bach et al. [18] presented a unified traffic light recognition system that can perform state classification (circle, straight, left, right) by using a faster region-based CNN (Faster R-CNN) structure [19].

Recent methods for traffic light recognition can be divided into two classes. Methods in the first class detect a traffic light, cut the region of the traffic light, and deliver the traffic light information to a classifier for light state recognition [14]. Methods in the second class simultaneously detect a traffic light's position and recognize its light state [16,18]. When the location of an object is forecasted with a high confidence and a bounding box, an extra branch is utilized for the light state prediction. Other than the recognition of basic circular lights, a traffic light recognition system must typically recognize multiple types of arrow lights used in multiple countries. However, few studies have explored this issue [12,18]. A two-stage technique is generally used, first classifying light colors and then classifying arrow types.

One disadvantage of image-based methods for detecting traffic lights is the false positives due to the presence of similar background features. To reduce incorrect detection, a simple solution is to limit the ROI in an image when searching a traffic light. The location of the traffic light can also provide more information that improves the accuracy of detection. In this case, the aim is to ameliorate image-based methods. The traffic light detection method based on maps operates on the basis that traffic lights are placed at steady positions under normal circumstances. Global positioning system (GPS) and light detection and ranging (LiDAR) are commonly used to establish high-definition (HD) maps and annotate traffic light positions on a vehicle's route [20,21]. When a vehicle is moving, map and localization information are utilized to compute the location where a traffic light will appear in a slight region. Furthermore, this information can be used to verify the presence of the next traffic light.

3. Approach

Figure 1 presents the flowchart illustrating the proposed approach. For the input image, we used map information to crop the image and obtain an approximate traffic light position. Subsequently, we introduced a traffic light detection and recognition approach for traffic lights that is based on deep neural networks. Finally, we output the traffic light positions and light signal types.

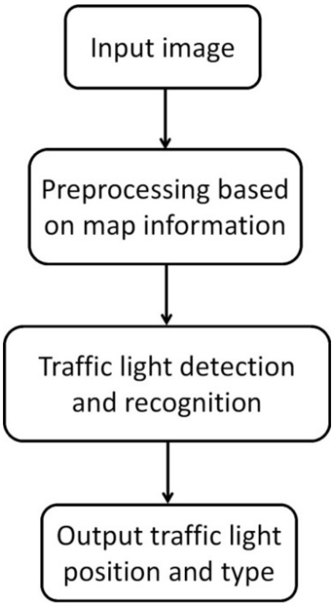
Input image
Preprocessing based on map information
Traffic light detection and recognition
Output traffic light position and type

Figure 1. Flowchart of the proposed approach.

3.1. Preprocessing Based on Map Information

Conventional image-based traffic light detection methods can misjudge traffic situations due to features in the background that appear similar to traffic lights. Traffic light detection requires a high level of accuracy because it affects subsequent decisions on vehicle control. Accordingly, we propose a map-based detection approach, not to replace image-based detection methods but to supplement them, with the aim of reducing misjudgment and thus enhancing accuracy.

In our approach, we integrated a HD map for detecting and recognizing traffic lights. We utilized a pre-constructed HD map with annotated traffic lights, including ID, position, and vertical and horizontal angle information. The position between traffic lights and a vehicle can be obtained using the HD map and the LiDAR data when the vehicle is moving. This information is used to crop an image to obtain an approximate position of a traffic light. Due to the nature of the registration of images and LiDAR data, the traffic lights cannot be correctly identified. Then, the segmented ROI is delivered to neural networks for precisely detecting locations and recognizing light states. Figure 2 shows the results of traffic light detection that uses a combination of image and LiDAR data.

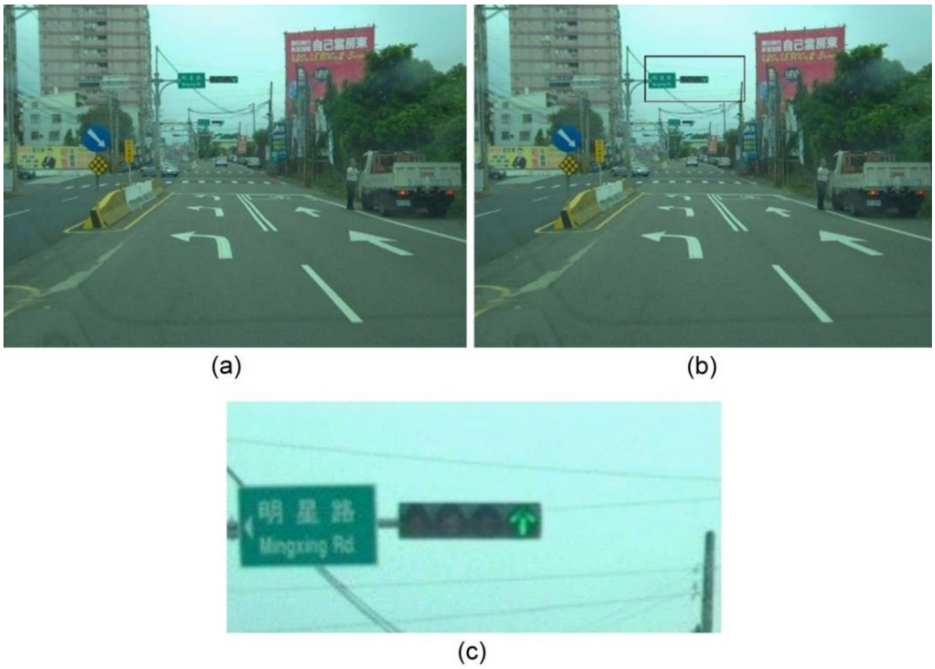

Figure 2. Results of traffic light detection using a combination of image and LiDAR data. (**a**) Input image, (**b**) processed result, and (**c**) the segmented image.

3.2. Traffic Light Detection and Recognition

Computation cost is a concern in traffic light detection. YOLOv3 [22] balances between accuracy and processing speed. Therefore, we integrated YOLOv3 into our approach for traffic light detection. The network framework of YOLOv3 can be divided into three parts. Figure 3 shows the network structure of YOLOv3. First, Darknet-53 is used to extract feature maps from the input images. Next, the feature pyramid network (FPN) integrates low-level and high-level features to produce feature maps of three scales. Finally, the prediction layer predicts objects of varying sizes in the feature maps. Figure 4 shows the input and output of the detection network.

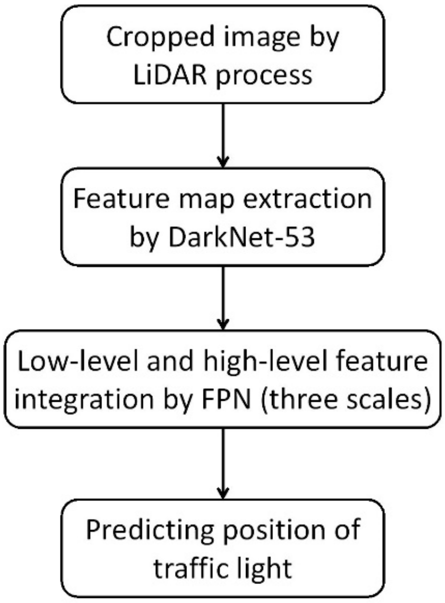
Cropped image by LiDAR process
Feature map extraction by DarkNet-53
Low-level and high-level feature integration by FPN (three scales)
Predicting position of traffic light

Figure 3. Network structure of YOLOv3.

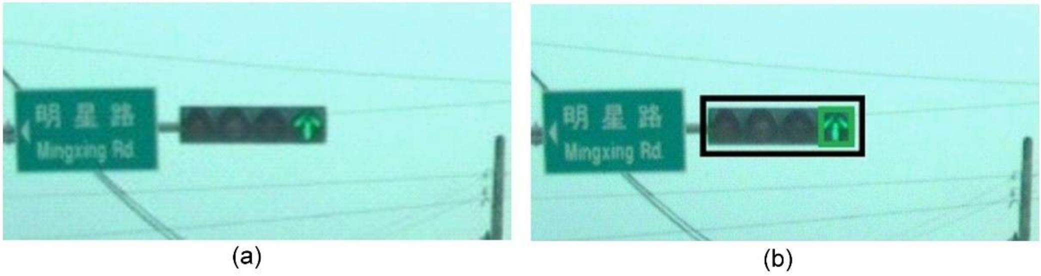

Figure 4. Detection network input and output. (**a**) Network input, and (**b**) network output.

For traffic light recognition, we introduced a new technique combining object detection and classification. We used YOLOv3-tiny [22] to detect and classify light states. Having a framework similar to that of YOLOv3, YOLOv3-tiny also comprises three parts, with the main difference being the feature-extraction network. Specifically, YOLOv3-tiny has fewer convolution layers and pooling layers, and its FPN produces feature maps of only two scales, in which the prediction layer predicts objects of different sizes. The smaller number of convolution layers contributes to its higher speed but reduces its accuracy.

In the proposed approach, there are four classes of light states: RedCircle, YellowCircle, GreenCircle, and Arrow. The light states of Arrow are then further classified into the LeftArrow, StraightArrow, and RightArrow classes using LeNet [23]. For example, when a light state of Red–Left–Right is established, YOLOv3-tiny detects one RedCircle and two Arrows. Moreover, the two Arrows will be recognized as LeftArrow and RightArrow classes by LeNet. Then, we can obtain the final traffic light state by merging the results obtained from the two networks.

3.3. *Unified Network*

Three networks were employed in our approach, namely YOLOv3 in the first stage, YOLOv3-tiny, and LeNet in the second stage. If the networks are trained separately, the resultant weights can only be used to optimize the results for each stage. If they can share feature maps in one unified network, better results could be achieved from end-to-end training. Moreover, the training speed and inference would also increase.

In a unified network, the three subnets share feature maps and, thus, the inputs and frameworks of the second and third subnets change. Shared feature maps replace images as the input. Feature extraction is removed from the subnets, and only their prediction function is retained. First, the third-layer feature maps of YOLOv3 that are generated through the FPN are extracted, and these maps are then cropped to retain the areas with traffic lights that are indicated by the YOLOv3 detection results. These areas are then converted to a fixed size through interpolation and adopted as the input feature maps for YOLOv3-tiny. After the inputs are subjected to the convolutional layer and FPN, YOLOv3-tiny then predicts the positions of traffic lights in feature maps of varying scales. At this point, the unified network has predicted the positions of traffic lights and their respective signals. The remaining task for the network is to judge and predict whether a traffic light has arrow lights. If a traffic light has arrow lights, the network must predict the arrow light type. Similar to the previous procedure, the second-layer feature maps produced by YOLOv3-tiny through the FPN are extracted and cropped to retain the areas of arrow lights as indicated by the detection results of YOLOv3-tiny. These areas are then converted to a fixed size and used as the input feature maps for LeNet. After the maps undergo convolution, global average pooling, and softmax layer processing, the network is able to predict arrow light types. Therefore, the unified network can recognize the position of traffic lights, the position of light signals on these traffic lights, and light signal types on the input images. Figure 5 illustrates the unified network architecture.

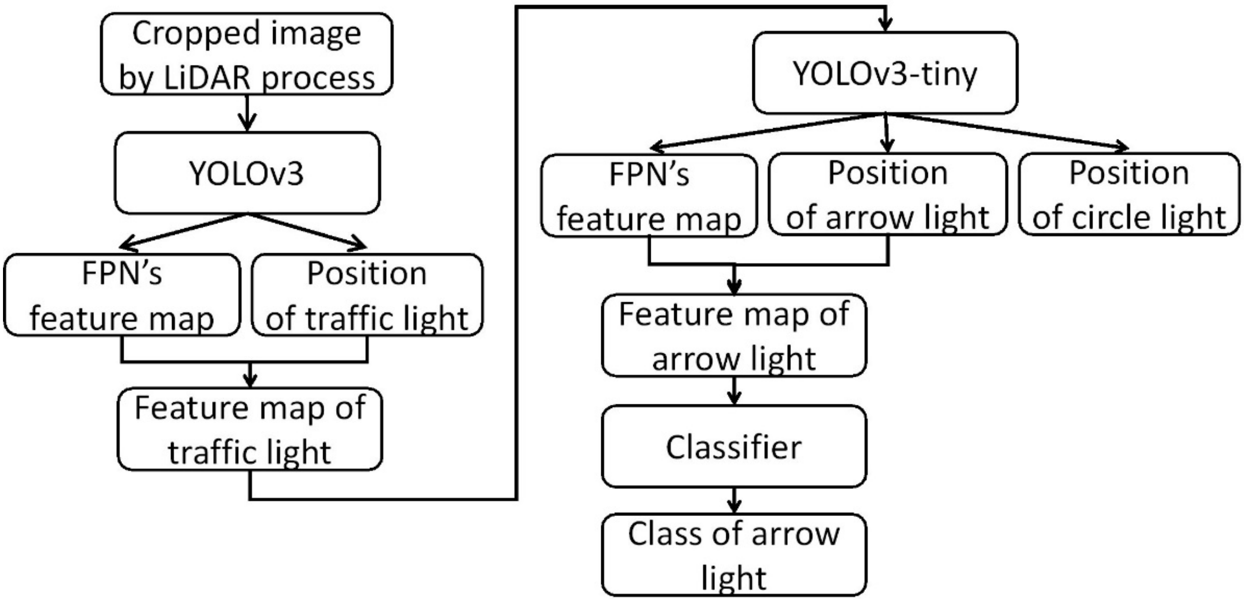

Figure 5. Flowchart of the unified network.

The loss function is calculated by summing the losses from YOLOv3′s traffic light prediction, YOLOv3-tiny's light signal detection, and LeNet's arrow light recognition. By summing the losses of the three subnets, we modified the unified network to account for the overall loss of all performed tasks. The loss functions of YOLOv3-tiny and YOLOv3 remain unchanged, but cross entropy loss is applied for LeNet. Moreover, our detection network is based around segmented images. Hence, our training images are segmented for simulating LiDAR processing (see Figure 6 as an illustration). Each image of traffic lights is segmented thrice. Then, the traffic light positions are randomly placed in the segmented image. The dataset includes six primary classes: Green, Yellow, Red, Straight, StraightRight, and Close. We performed data augmentation by rotating the arrow light images to generate more training data. Figure 7 shows an example of data augmentation.

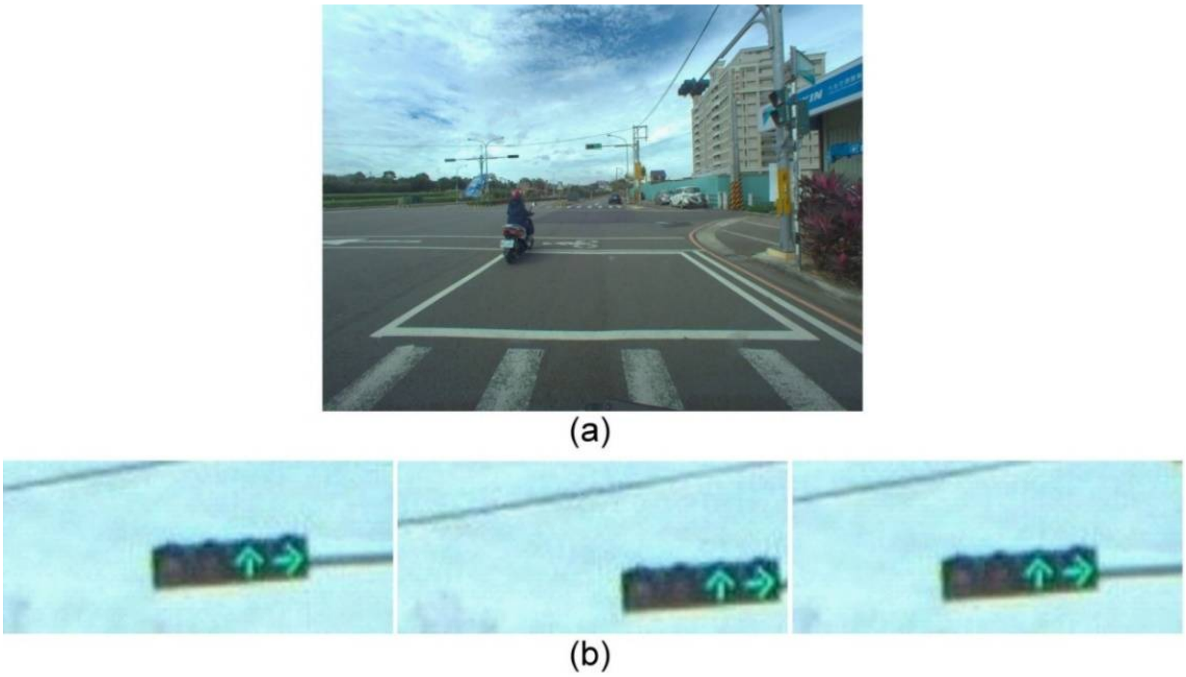

Figure 6. Training images. (**a**) Input image, and (**b**) images cropped using LiDAR data.

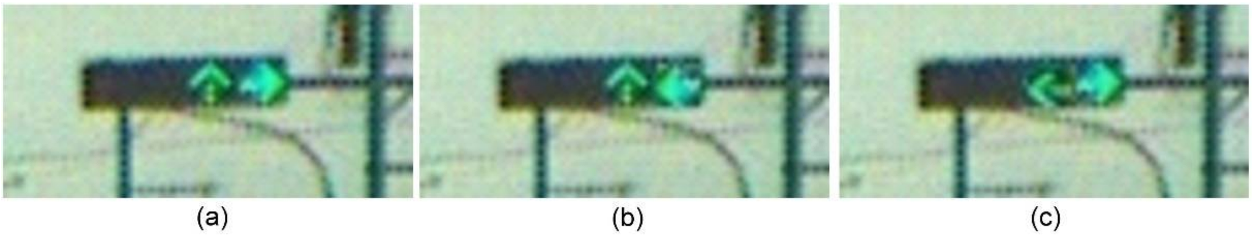

Figure 7. Augmented data. (**a**) Training image, (**b**) augmented image 1, and (**c**) augmented image 2.

4. Results

We conducted several experiments on images from our dataset and the LISA dataset [2]. The proposed approach was run on a computer with 3.60 GHz Core™ i7-7700 CPU and 8 GB of memory. The used graphics card was an NVDIA GeForce GTX 1070Ti.

4.1. Dataset

Scenarios where our approach was to be applied pose two technical challenges. First, currently available public datasets contained vertically arranged traffic lights, which differed from the horizontally arranged traffic lights. Second, although arrow signals appeared frequently on Taiwan's roads, most studies only used classifiers to recognize circle lights. The establishment of a self-collecting dataset can solve these problems. Furthermore, the lack of arrow light images perplexed the training process of the network.

Several commonly used datasets can be obtained for detecting and recognizing traffic lights. However, they each used different formats and were therefore unsuitable for network training involving traffic lights in Taiwan. Therefore, we worked with the Industrial Technology Research Institute (ITRI). We collected a dataset for evaluation and training, and the dataset comprises data on two routes. The first was the route from Hsinchu High-Speed Railway Station to the ITRI campus, and the second was the route from Chiayi High-Speed Railway Station to National Chung Cheng University. The first and second routes, respectively, spanned 16 and 39 km and took 40 and 50 min to record. Two cameras having different focal lengths (3.5 and 12 mm) were placed below the rearview mirror of a vehicle for acquiring images. The image sequences were captured at 36 fps. The resolution of the captured image was a size of 2048 × 1536. Additionally, LiDAR data were recorded using a Velodyne Ultra Puck VLP-32C and used to segment approximate regions of traffic lights. The first and second routes were recorded thrice and once, respectively. We sampled five frames per second for processing and labeling the positions of traffic lights and the classes of light states. The labeled images contained 26,868 images and 29,963 traffic lights. Only the traffic lights with obvious light states were labeled and 14 classes of light state combinations were established.

In the LISA dataset, traffic lights were arranged vertically (see Figure 8 as an illustration). The traffic lights in Taiwan were arranged horizontally, as depicted in Figure 9. Additionally, the light states were distinct. In the LISA dataset, only a single light can be shown at a time in the traffic lights, whereas traffic scenarios in Taiwan can involve multiple combinations of traffic lights with multiple arrow light types. Our dataset contained mostly data on circular lights and several arrow lights. With respect to ROI size of traffic lights, the LISA dataset mainly contained traffic light regions in the range of 15 to 30 pixels. In our dataset, the images captured with a 3.5 mm lens camera consisted of traffic light regions in the ranges of 10 to 20 pixels, and the images obtained with a 12 mm lens camera consisted of traffic light regions in the ranges of 15 to 50 pixels.

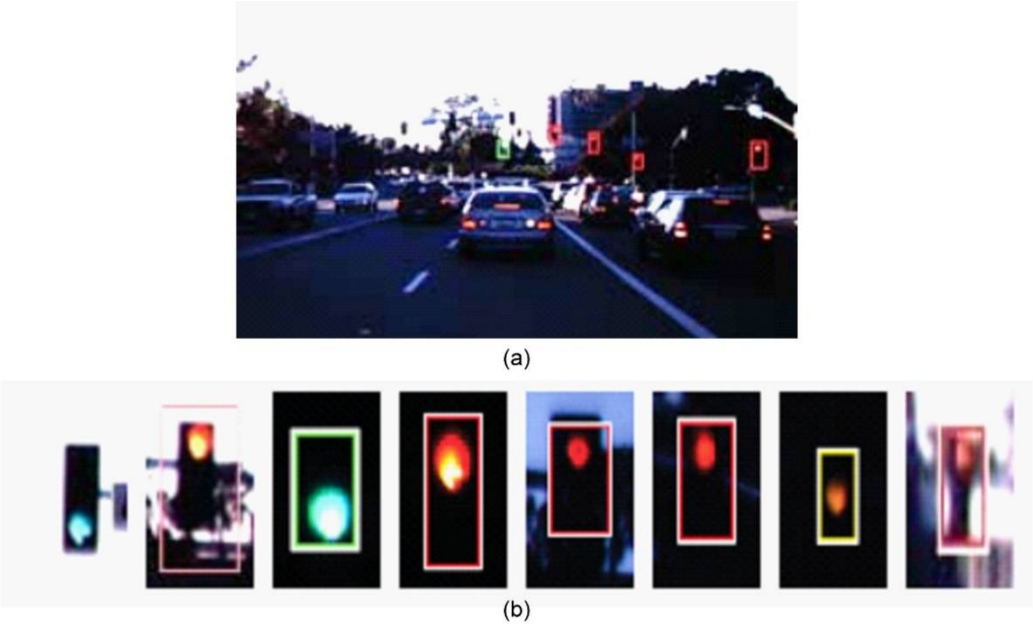

Figure 8. Examples in the LISA dataset. (**a**) Traffic scene, and (**b**) traffic lights.

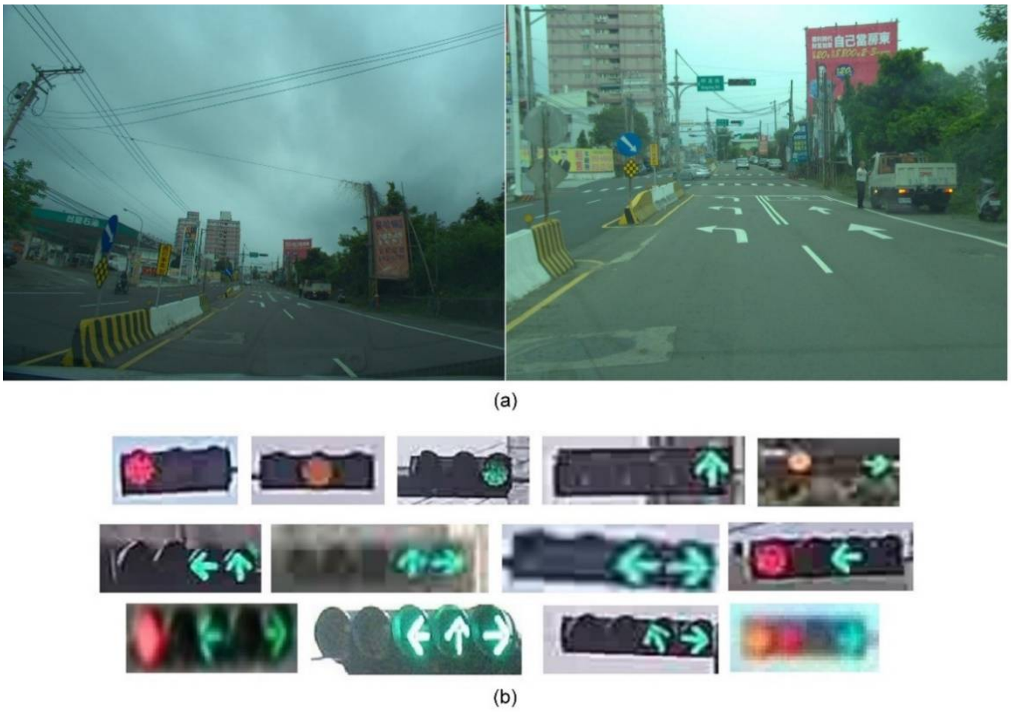

Figure 9. Examples in our dataset. (**a**) The traffic light images obtained with 3.5/12 mm lens cameras, and (**b**) traffic lights.

4.2. Evaluation

The mean average precision (mAP) score was used to evaluate the object detection results. The daytime images in the LISA dataset were utilized for training and testing to compare our approach with previous methods, as indicated in Table 1. The intersection over union (IoU) was set as 0.5. From the results, conventional detectors did not perform well because there were complicated scenes. In Table 1, the results reported in row 1 (color detector), row 2 (spot detector), and row 3 (aggregate channel features detector, ACF detector) were extracted from Jensen et al. [2], and those reported in row 4 (Faster R-CNN), row 5 (spot light detection), row 6 (modified ACF detector), and row 7 (multi-detector) were extracted from Li et al. [24]. The low accuracy of Faster R-CNN can primarily be attributed to the small regions of traffic lights, which were hard to detect after performing convolution layer-by-layer [24]. The proposed approach (results shown in the last row, Table 1) outperformed the other methods for both circular lights and arrow lights.

Table 1. LISA daytime dataset test results (mAP).

Method	Stop	StopLeft	Go	GoLeft	Warning	WarningLeft	All
Color detector	-	-	-	-	-	-	0.04
Spot detector	-	-	-	-	-	-	0.0004
ACF detector	-	-	-	-	-	-	0.36
Faster R-CNN	0.14	0.01	0.19	0.001	-	-	0.09
SLD	0.08	-	0.10	-	-	-	0.09
Modified ACF detector	0.63	0.13	0.40	0.37	-	-	0.38
Multi-detector	0.72	0.28	0.52	0.40	-	-	0.48
Our approach	0.70	0.40	0.88	0.71	0.52	0.24	0.66

Table 2 presents the accuracy and computation speed results of multiple network structures when they were applied to our dataset. Four network structures (including one with data augmentation) were compared. The first network structure used YOLOv3 to detect the traffic lights and AlexNet [25] to classify light states. The second network structure used the combined YOLOv3 + YOLOv3-tiny + LeNet approach. However, in this case, the three networks operated as independent networks. The unified network structure integrated these three subnets. These networks were trained and tested with the same dataset. Table 2 shows that the proposed methods achieved higher mAPs than that of the YOLOv3 + AlexNet network structure but required more computation costs. Relative to the first two networks, the one using LeNet for arrow light classification had a higher mAP but required more computation costs. As seen from Table 2, the two versions of the proposed unified network (the second-rightmost and rightmost columns in Table 2) had higher mAPs than the YOLOv3 + YOLOv3-tiny + LeNet network structure did. Furthermore, their computation speed was quicker because of sharing feature maps.

Table 2. Results obtained using our dataset.

Method	YOLOv3 + AlexNet	YOLOv3 + YOLOv3-tiny + LeNet	Unified Network	Unified Network
Data augmentation	-	-	-	✓
mAP	0.36	0.55	0.57	0.67
Speed (ms)	31	52	40	40

Images captured with varying distances contained traffic lights with varying sizes of ROIs. This appeared to affect the results of traffic light detection and recognition. In the proposed approach, 3.5 and 12 mm lens cameras were used for capturing images. Table 3 shows the mAP for varying ROI sizes of traffic lights (height measured in pixels), the size of traffic lights at varying distances, and the mAP for varying distances. The traffic light images captured by the 12 mm lens camera were bigger than those taken by the 3.5 mm lens camera. The images with bigger traffic lights provided better results for detection. The detection results for the images captured by the 12 mm lens camera were better than those for the images captured by the 3.5 mm lens camera for different distances. Nevertheless, we used the 3.5 mm lens camera for near scenes. At the distance of 0 to 15 m, traffic light images could only be taken by the 3.5 mm lens camera because of the cameras' field of view (see Figure 10 as an illustration).

Table 3. Relationship between mAP, traffic light size (height in pixels), and traffic light distance.

Traffic Light Size	**0–5**	**5–10**	**10–15**	**15–20**	**20–25**	**25–30**	**30–35**	**35–40**	**40–45**	**45–50**	**50–55**	**55–60**	**60–65**	**65–70**
mAP	0.33	0.48	0.82	0.80	0.76	0.75	0.69	0.74	0.77	0.63	0.69	0.69	0.55	0.62
Distance	0–15		15–30		30–45		45–60		60–75		75–90		90–100	
Traffic light size (3.5 mm)	62		20		15		9		7		6		5	
Traffic light size (12 mm)	-		43		39		29		21		15		15	
Distance	0–15		15–30		30–45		45–60		60–75		75–90		90–100	
mAP (3.5 mm)	0.38		0.62		0.57		0.47		0.46		0.37		0.32	
mAP (12 mm)	-		0.79		0.68		0.69		0.70		0.63		0.51	

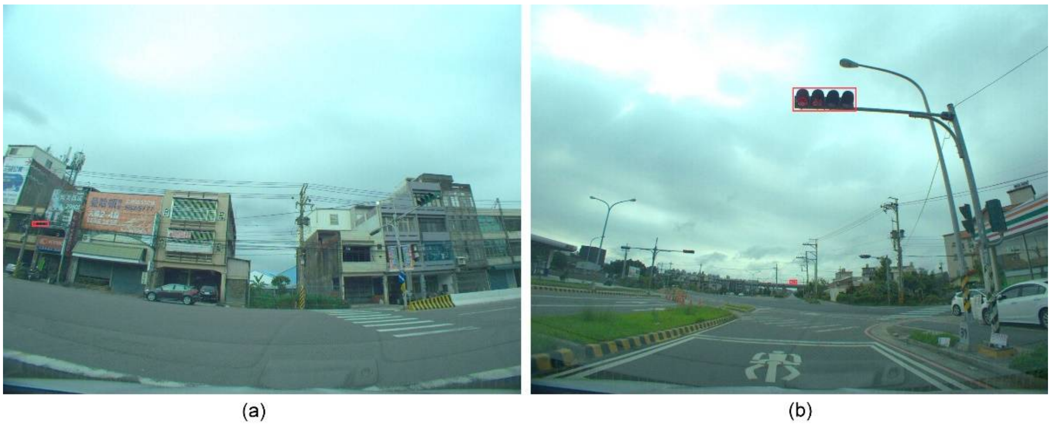

Figure 10. Traffic lights at the distance of 0 to 15 m could only be taken by the 3.5 mm lens camera. (**a**) A traffic light is at the border of the image, and (**b**) two traffic lights are faraway.

The proposed approach contained three phases, namely traffic light detection, initial light state classification, and arrow type recognition. Table 4 shows the mAPs for each network phase. The mAPs of initial light state classification and arrow type recognition were computed based on the results of traffic light detection computed by the previous subnet. Table 4 depicts that larger errors primarily occurred in the second subnet. Additionally, the mAP of the Green class was the lowest because the arrow light and green light appeared similar from faraway. Table 5 displays the mAPs for all classes. The classes with more training images had higher mAPs. However, data augmentation did not improve the accuracy for the classes with insufficient samples.

Table 4. mAP of each network phase.

	Detection	**State**				**Type**		
Class	Traffic Light	Red	Yellow	Green	Arrow	Left	Straight	Right
mAP	0.97	0.93	0.90	0.64	0.91	0.87	0.98	0.97

Table 5. mAP for each class.

Class	**Close**	**Red**	**Yellow**	**Green**	**Left**	**Straight**	**Right**
mAP	0.43	0.78	0.79	0.76	No data	0.55	No data
Class	Red Left	Red Right	Straight Left	Straight Right	Left Right	Red Left Right	Straight Left Right
mAP	0.55	0.45	0.64	0.87	0.84	No data	0.69

Finally, several detection results of the unified network are shown in Figure 11. The results show that our approach can obtain the desired performance.

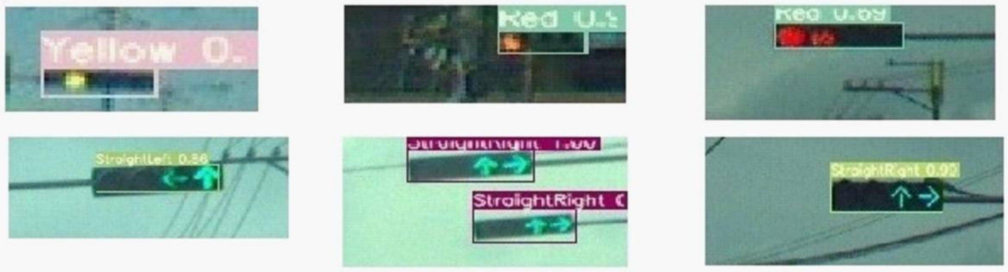

Figure 11. Several detection results of the unified network.

5. Conclusions

We have presented a traffic light detection and recognition approach for traffic lights that is based on a convolutional neural network. For traffic light detection, two cameras were used with different focal lengths to capture nearby and faraway scenes. The map information was utilized to facilitate traffic light detection by restricting the ROI. For traffic light recognition, we proposed a technique that combines object detection and classification to solve the issue of multiple light state classes in many urban traffic scenes. Additionally, a unified network was proposed by sharing features to reduce training and computation costs. The experiments performed using the LISA dataset and our dataset have demonstrated that the proposed approach performed better than the previous methods.

The proposed approach, despite its ability to detect and recognize traffic lights, did not achieve 100% accuracy, indicating room for improvement. For traffic light detection and recognition, a low tolerance for error is required because the safety of passengers and others is at stake. The proposed approach may be improved by future research through improvement of the detection method and the dataset used.

Author Contributions: Methodology, T.-W.Y. and H.-Y.L.; Supervision, H.-Y.L. and C.-C.C.; Writing—original draft, T.-W.Y.; Writing—review and editing, H.-Y.L. and C.-C.C. All authors have read and agreed to the published version of the manuscript.

Funding: The support of this work in part by the Ministry of Science and Technology of Taiwan under Grant MOST 106-2221-E-194-004 and the Advanced Institute of Manufacturing with High-tech Innovations (AIM-HI) from The Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan is gratefully acknowledged.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Abboud, K.; Omar, H.A.; Zhuang, W. Interworking of dsrc and cellular network technologies for v2x communications: A survey. *IEEE Trans. Veh. Technol.* **2016**, *65*, 9457–9470. [CrossRef]
2. Jensen, M.B.; Philipsen, M.P.; Møgelmose, A.; Moeslund, T.B.; Trivedi, M.M. Vision for looking at traffic lights: Issues, survey, and perspectives. *IEEE Trans. Intell. Transp. Syst.* **2016**, *17*, 1800–1815. [CrossRef]
3. Caesar, H.; Bankiti, V.; Lang, A.H.; Vora, S.; Liong, V.E.; Xu, Q.; Krishnan, A.; Pan, Y.; Baldan, G.; Beijbom, O. nuscenes: A multimodal dataset for autonomous driving. *arXiv* **2019**, arXiv:1903.11027.
4. Ramanishka, V.; Chen, Y.T.; Misu, T.; Saenko, K. Toward driving scene understanding: A dataset for learning driver behavior and causal reasoning. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7699–7707.
5. Waymo. Waymo Open Dataset: An Autonomous Driving Dataset. 2019. Available online: https://waymo.com/open/ (accessed on 7 October 2019).
6. Weber, M.; Wolf, P.; Zöllner, J.M. Deeptlr: A single deep convolutional network for detection and classification of traffic lights. In Proceedings of the 2016 IEEE Intelligent Vehicles Symposium (IV), Gothenburg, Sweden, 19–22 June 2016; pp. 342–348.

7. Yeh, T.W.; Lin, H.Y. Detection and recognition of arrow traffic signals using a two-stage neural network structure. In Proceedings of the the 6th International Conference on Vehicle Technology and Intelligent Transport Systems (VEHITS 2020), Prague, Czech, 2–4 May 2020; pp. 322–330.
8. Fregin, A.; Müller, J.M.; Dietmayer, K.C.J. Feature detectors for traffic light recognition. In Proceedings of the 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), Yokohama, Japan, 16–19 October 2017; pp. 339–346.
9. Kim, H.K.; Park, J.H.; Jung, H.Y. Effective traffic lights recognition method for real time driving assistance system in the daytime. *Int. J. Electr. Comput. Eng.* **2011**, *5*, 1429–1432.
10. Fregin, A.; Müller, J.M.; Dietmayer, K.C.J. Three ways of using stereo vision for traffic light recognition. In Proceedings of the 2017 IEEE Intelligent Vehicles Symposium (IV), Los Angeles, CA, USA, 11–14 June 2017; pp. 430–436.
11. Müller, J.M.; Fregin, A.; Dietmayer, K.C.J. Multi-camera system for traffic light detection: About camera setup and mapping of detections. In Proceedings of the 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), Yokohama, Japan, 16–19 October 2017; pp. 165–172.
12. Weber, M.; Huber, M.; Zöllner, J.M. Hdtlr: A cnn based hierarchical detector for traffic lights. In Proceedings of the 2018 21st International Conference on Intelligent Transportation Systems (ITSC), Maui, HI, USA, 4–7 November 2018; pp. 255–260.
13. Sermanet, P.; Eigen, D.; Zhang, X.; Mathieu, M.; Fergus, R.; LeCun, Y. Overfeat: Integrated recognition, localization and detection using convolutional networks. In Proceedings of the 2nd International Conference on Learning Representations (ICLR 2014), Banff, AB, Canada, 14–16 April 2014.
14. Behrendt, K.; Novak, L.; Botros, R. A deep learning approach to traffic lights: Detection, tracking, and classification. In Proceedings of the 2017 IEEE International Conference on Robotics and Automation (ICRA), Singapore, 29 May–3 June 2017; pp. 1370–1377.
15. Redmon, J.; Divvala, S.K.; Girshick, R.B.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
16. Müller, J.M.; Dietmayer, K.C.J. Detecting traffic lights by single shot detection. In Proceedings of the 2018 21st International Conference on Intelligent Transportation Systems (ITSC), Maui, HI, USA, 4–7 November 2018; pp. 266–273.
17. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.E.; Fu, C.Y.; Berg, A.C. Ssd: Single shot multibox detector. In Proceedings of the 14th European Conference on Computer Vision (ECCV2016), Amsterdam, The Netherlands, 11–14 October 2016; pp. 21–37.
18. Bach, M.; Stumper, D.; Dietmayer, K.C.J. Deep convolutional traffic light recognition for automated driving. In Proceedings of the 2018 21st International Conference on Intelligent Transportation Systems (ITSC), Maui, HI, USA, 4–7 November 2018; pp. 851–858.
19. Ren, S.; He, K.; Girshick, R.B.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *39*, 1137–1149. [CrossRef] [PubMed]
20. Fairfield, N.; Urmson, C. Traffic light mapping and detection. In Proceedings of the 2011 IEEE International Conference on Robotics and Automation, Shanghai, China, 9–13 May 2011; pp. 5421–5426.
21. Hirabayashi, M.; Sujiwo, A.; Monrroy, A.; Kato, S.; Edahiro, M. Traffic light recognition using high-definition map features. *Robot. Auton. Syst.* **2019**, *111*, 62–72. [CrossRef]
22. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
23. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [CrossRef]
24. Li, X.; Ma, H.; Wang, X.; Zhang, X. Traffic light recognition for complex scene with fusion detections. *IEEE Trans. Intell. Transp. Syst.* **2018**, *19*, 199–208. [CrossRef]
25. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Commun. ACM* **2012**, *60*, 84–90. [CrossRef]