International Core Journal of Engineering Volume 10 Issue 6, 2024
ISSN: 2414-1895 DOI: 10.6919/ICJE.202406_10(6).0009

Deep Learning based Forklift Obstacle Detection and Distance Estimation Using Single RGB Camera

Xinwei Liu[1,2], Yimin Song[1], Min Fu[2], and Guoyun Ye[2]

[1] Tianjin University, Tianjin, China

[2] Ningbo Ruyi Joint Stock Co. Ltd, Ninghai, Zhejiang, China

Abstract

Ensuring the precise detection of road obstacles is imperative for the safe implementation of autonomous driving, particularly for forklift. However, existing methods encounter difficulties when analyzing road scenes that feature intricate backgrounds. This is primarily due to the limitations of supervised approaches, as they fail to identify unknown objects that are not represented in the training dataset. To overcome this challenge, we present a novel road obstacle detection and distance estimation method that employs an autoencoder equipped with semantic segmentation, trained exclusively on data derived from different road scenes. Our method relies solely on a color image captured by a standard in-vehicle RGB camera as the input. By utilizing an autoencoder architecture composed of a semantic image generator as the encoder and a photographic image generator as the decoder, we generate a resynthesized image. Extensive experiments demonstrate that our proposed method achieves performance levels comparable to those of existing approaches, even in the absence of postprocessing. Furthermore, when postprocessing techniques are applied, our method outperforms state-of-the-art approaches on the Lost and Found dataset. Additionally, evaluations conducted on real-world and synthetic road obstacles, reveal that our method significantly surpasses the performance of a supervised approach that explicitly focuses on road obstacle detection and distance estimation. Thus, our deep-learning-based forklift obstacle detection and distance estimation method represents a practical solution that facilitates the advancement of industrial vehicle autonomous driving systems.

Keywords

Obstacle Detection; Distance Estimation; Forklift; Deep Learning; Autonomous Driving.

1. Introduction

In recent times, the development of advanced driving support systems has witnessed remarkable progress, with the ultimate aim of realizing fully autonomous driving. Human-machine interfaces play a pivotal role in these systems, providing drivers with the necessary support for safe, secure, and comfortable driving, especially in the field of forklift driving. By effectively communicating changes in the driving environment, such as traffic congestion, weather variations, and road obstacles, drivers are empowered to navigate their vehicles with increased awareness. This information, initially detected by preceding vehicles, is seamlessly transmitted to subsequent vehicles, establishing a cooperative network that enhances the overall driving experience.

The Ministry of Communications of the Transport in China released a report highlighting that in 2020, a staggering 480,000 road obstacles were identified, amounting to an alarming average of nearly 1,600 obstacles encountered each day. The presence of these obstacles is a major contributing factor

ISSN: 2414-1895 DOI: 10.6919/ICJE.202406_10(6).0009

to severe accidents. As a result, there is an urgent social imperative to develop an automated road obstacle detection system, as the current practice involves manual identification and elimination of these obstacles.

Intelligent forklifts, equipped with essential artificial intelligence capabilities, are autonomous vehicles capable of independently handling the storage and retrieval of goods without human involvement. In the warehousing industry, characterized by abundant and valuable products, the utilization of shelving and stacking methods for storage purposes proves to be highly effective. By replacing manual handling with intelligent forklifts, the efficiency of warehousing transportation can be significantly enhanced, resulting in standardized and intelligent management of warehouse operations [1]. Given its direct impact on the efficient functioning of intelligent forklifts in warehousing, obstacle detection plays a vital role in ensuring the safe operation of these intelligent vehicles. As a critical research area, obstacle detection has gained prominence, emerging as a pivotal technology for intelligent vehicles [2-4].

Multiple approaches have been proposed to recognize the driving environment in this context. However, these approaches are ill-suited for commercially available vehicles (e.g. forklift), as they rely on specialized sensors like stereo cameras, LI-DAR, and radar. Furthermore, the adoption of such sensors for autonomous driving is economically infeasible and imposes significant power requirements. Alternatively, a deep-learning-based approach emerges as a potential alternative to sensor-dependent methods. However, collecting a substantial volume of data necessary for supervised learning proves impractical due to the substantial variability in colors, shapes, sizes, and textures of road obstacles, as illustrated in Figure. 1. Consequently, training a classifier to reliably detect road obstacles for forklift, including unknown objects, becomes an almost insurmountable challenge.

In this paper, we propose a forklift obstacle detection and distance estimation method based on an autoencoder integrated with semantic segmentation. Notably, the method exclusively relies on a color image captured by a typical in-vehicle RGB camera as its input. Utilizing an autoencoder architecture consisting of a semantic image generator [2] as the encoder and a photographic image generator [3] as the decoder, the method generates a resynthesized image. By quantifying the perceptual loss [3] between the input and resynthesized images and incorporating the entropy of the semantic image, an anomaly map is produced. Subsequently, postprocessing techniques are employed on the anomaly map to accurately localize road obstacles within the image. Specifically, a widely-used technique for calculating visual saliency in images [4][5] is applied to sharpen the anomaly map.

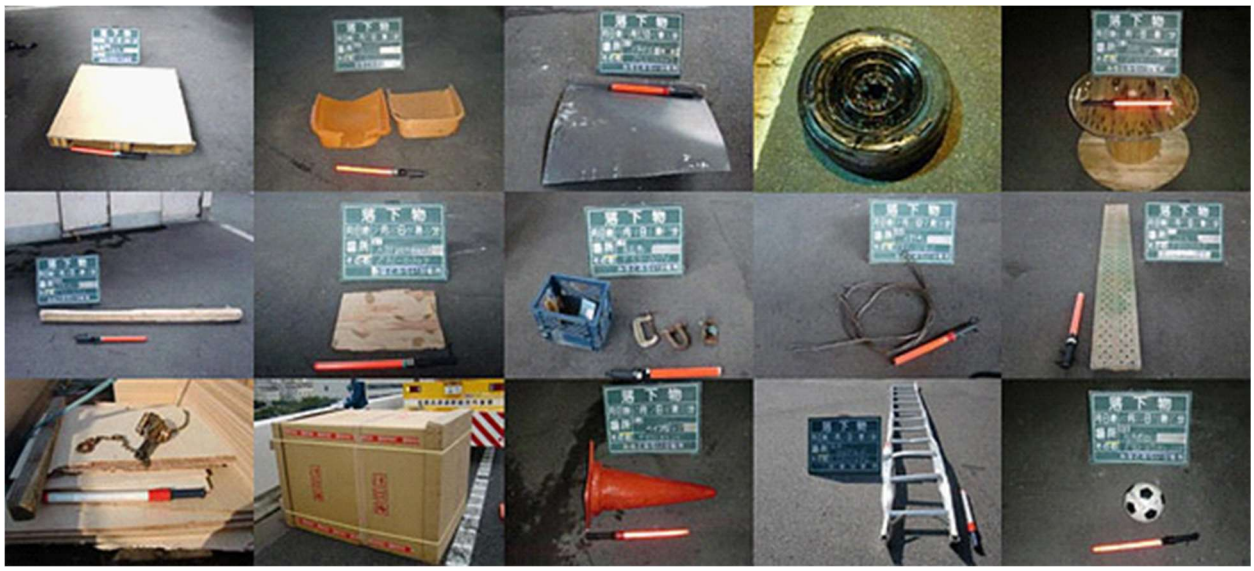

Figure 1. Examples of road obstacles (e.g., burst tire debris, road cones, plywood, square lumber, and scrap, a soccer ball).

ISSN: 2414-1895 DOI: 10.6919/ICJE.202406_10(6).0009

2. Related Work

In the early investigations focused on forklift obstacle detection in in door environments, the application of stereo vision techniques was prominent. Notably, Hancock [8] utilized laser reflectance and stereo vision to successfully detect small road obstacles situated at significant distances. Similarly, William et al. [9] employed a multibaseline stereo technique to identify small road obstacles, approximately 14 cm in height, even at distances exceeding 100 m. Recent studies also continued to adopt stereo cameras or the structure-from-motion technique for road obstacle detection. Subaru Eyesight [10], for instance, stands as a well-known stereo-vision-based system capable of robustly detecting large road obstacles. Additionally, the commercially available Mobileye system [11] demonstrates the ability to reliably detect large obstacles at close range using only a monocular camera. Furthermore, Tokudome et al. [12] introduced a novel real-time environment recognition system for autonomous driving that leverages a LIDAR sensor.

Diverging from special-sensor-based strategies, machine-learning-based approaches operate by harnessing a standard in-vehicle camera to extract raw images, subsequently leveraging advanced machine learning techniques such as autoencoders [14][15], uncertainty-based methodologies [16][17], and generative adversarial networks (GANs) [18][19] to transform these images into informative features. Autoencoder-based methods [14][15] involve comparing small input patches with the output of a shallow autoencoder trained exclusively on road textures. This enables the differentiation of road patches from other types. However, these approaches also encompass normal objects like vehicles, traffic signs, and buildings, leading to a notable number of false positives. Uncertainty-based techniques [16][17], relying on the Bayesian SegNet framework, incorporate an uncertainty threshold to identify potentially mislabeled regions, including unknown objects. Nonetheless, these methods also generate a considerable number of false positives in irrelevant regions, such as boundary regions associated with semantic labels like roads, vehicles, buildings, sky, and nature. GAN-based approaches [18] employ an adversarial autoencoder to process an image and measure the feature loss between the output and input images. While these methods can classify entire images, they lack the capability to specifically detect anomalies within them. Moreover, in GAN-based approaches [19], an algorithm searches for the latent vector that produces an image closely resembling the input, based on a GAN trained to represent an original distribution. However, this computationally intensive process does not effectively identify anomalies.

Despite its advantages, the supervised training of this approach tends to exhibit limitations when faced with images containing intricate backgrounds. The primary drawback lies in the network's exclusive focus on learning mislabeled semantic maps for typical objects, neglecting the crucial aspect of learning mislabeled semantic maps for unknown objects. Additionally, the training process for the discrepancy network is notably complex, as achieving end-to-end training proves to be a challenging endeavor.

3. The Proposed Method

3.1 Obstacle Detection

While our basic idea aligns with that presented in [20], our implementation diverges significantly from existing methods, offering a more practical solution. Our implementation comprises three distinct components: an autoencoder, an anomaly calculator, and a postprocessor, as illustrated in Figure 2. Initially, the autoencoder takes a color image (a) captured by a standard in-vehicle camera as input. Using a semantic segmentation technique [2], the autoencoder generates a semantic map (b) and employs a photographic image synthesis technique [3] to create a resynthesized image (c). Subsequently, the anomaly calculator combines the perceptual loss (d) and entropy (e) associated with the semantic map to generate an anomaly map (f). Finally, the postprocessor sharpens the anomaly map for each local region (g), resulting in an obstacle score map (h). The subsequent sections provide a comprehensive description of these steps.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202406_10(6).0009

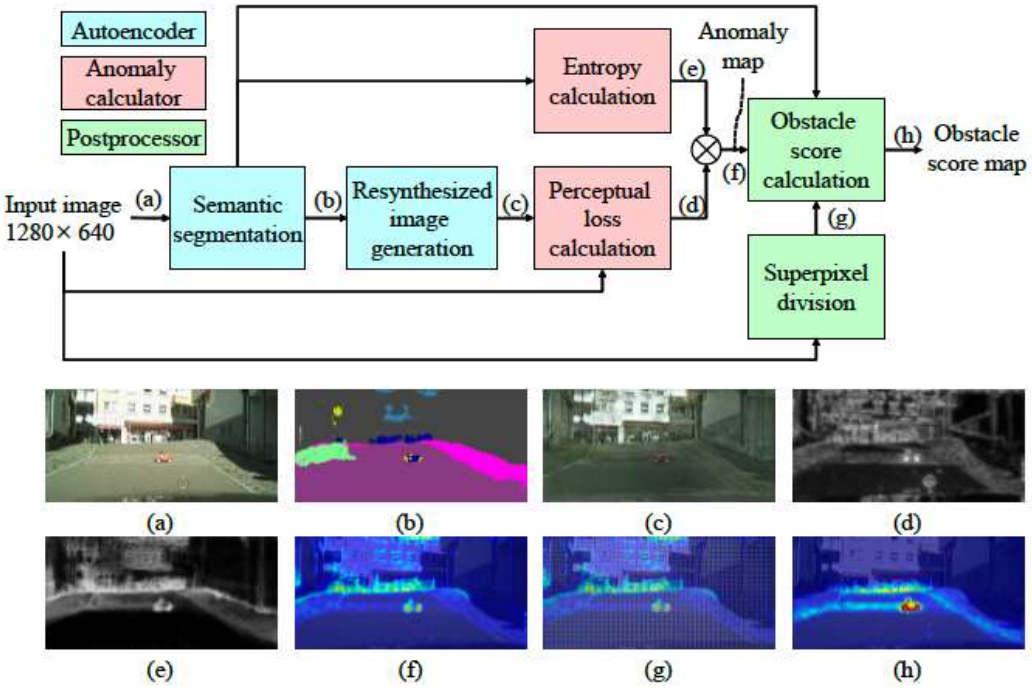

Figure 2. Schematic overview of our road obstacle detection system. (a) Input image, (b) semantic map, (c) resynthesized image, (d) perceptual loss, (e) entropy map, (f) anomaly map, (g) superpixels, and (h) obstacle score map.

3.1.1 Autoencoder

Within our system, the autoencoder is comprised of two essential modules: semantic segmentation and resynthesized image generation, as depicted in Figure 2. These modules play a crucial role in generating a semantic map and a resynthesized image, which are subsequently forwarded to the anomaly calculator and postprocessor. To achieve semantic segmentation, we employ ICNet [2], a representative technique capable of segmenting the input image into 20 distinct semantic labels, such as road, car, traffic light, and traffic sign. Our input images are sourced from the publicly available Cityscapes dataset, widely used for evaluating and training vision algorithms [23]. To accommodate GPU memory constraints, we downscale the Cityscapes training dataset to a resolution of 1,280 × 640 pixels. After fixing the semantic segmentation model parameters, we concatenate the semantic segmentation module with the resynthesized image generation module. For resynthesized image generation, we utilize an advanced technique called cascaded refinement network [3], which processes the semantic map to generate an image (referred to as the resynthesized image) identical to the input image from the Cityscapes dataset [23]. Among the three types of components, only the autoencoder requires learning the model parameters.

Our algorithm addresses the challenge of improving the quality of resynthesized images by introducing a straightforward solution. Rather than connecting the decoder directly to the output of the last layer (i.e., the softmax layer), we establish a connection with an intermediate layer (i.e., the convolution layer immediately prior to the softmax layer). This simple adjustment proves highly effective without requiring additional functions like instance segmentation and instance-level feature embedding, which are essential in Pix2PixHD [22]. Unlike Resynth, which trains the encoder and decoder separately due to memory limitations, our algorithm facilitates seamless end-to-end learning and enables rapid inference by concatenating lightweight deep neural networks.

3.1.2 Obstacle Detector

The obstacle detector plays a vital role in the system and includes two key modules: entropy calculation and perceptual loss calculation, as illustrated in Figure 2. Its primary objective is to generate an anomaly map, which assigns an anomaly score to each pixel in the input image, and subsequently transmits this map to the postprocessor. When tackling the challenge of estimating

ISSN: 2414-1895 DOI: 10.6919/ICJE.202406_10(6).0009

semantic labels for an unknown object, we make certain assumptions. Firstly, we acknowledge that the semantic map inherently harbors ambiguity surrounding the unknown object. Secondly, the resynthesized image displays substantial differences in appearance compared to the input image due to this inherent ambiguity. To quantify the level of ambiguity, we compute the entropy for the semantic map. Furthermore, we evaluate the dissimilarities in appearance using the perceptual loss [3]. The anomaly score is defined as the product of these calculated measures. Specifically, we define the entropy for the semantic map as follows:

$$W = T_{bi}\left(-\sum_{s} r^s \log(r^s)\right)$$

Here, r^s is the probability of the s-th semantic label estimated using the semantic segmentation technique [2] and T_{bi} is a bilinear-interpolation-based up-converter that upconverts the resynthesized image to the same resolution as the input image. Further, we define the perceptual loss between the input and resynthesized images as follows:

$$\Psi = \sum_{k=1}^{5} T_{bi}(\Psi^k)$$
$$\Psi^k = \|\Upsilon^k(E) - \Upsilon^k(F)\|_1$$

Here, E and F are the input and resynthesized images, respectively. In addition, W is the output from the k-th hidden layer of VGG19 [24]. Specifically, W(k = 1, . . . , 5) are given by the outputs from conv1 2, conv2 2, conv3 2, conv4 2, and conv5 2, as shown in Figure 3. Thus, we obtain the output from each hidden layer using VGG19 on the input and resynthesized images. Additionally, we define the L1 norm between the output from the k-th hidden layer for the input image and the output from the k-th hidden layer for the resynthesized image as perceptual loss Ψ^k. Further, we calculate the total perceptual loss Ψ by adding perceptual loss Ψ^k (l = 1, . . . , 5) after adjusting its resolution with upconverter T_{bi}. Finally, we generate anomaly map $\forall$ by taking the element-wise product of perceptual loss Ψ and entropy Φ as follows:

$$\forall = \Psi \odot \Phi$$

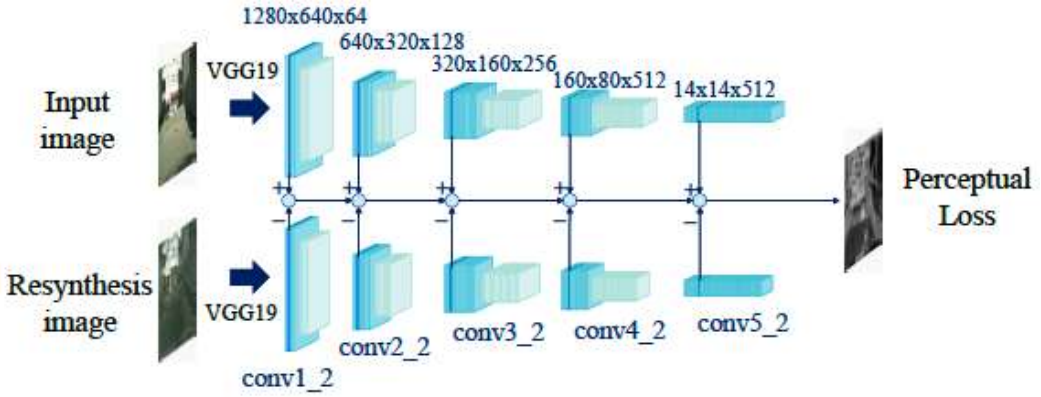

Figure 3. Overview of the module responsible for calculating the perceptual loss.

3.2 Obstacle Distance Estimation

The perspective transformation of a monocular vision system can be modeled using the pinhole camera principle. The image acquisition process involves projecting the three-dimensional

ISSN: 2414-1895 DOI: 10.6919/ICJE.202406_10(6).0009

coordinates of the real-world road surface onto the two-dimensional image plane of the camera. In contrast, distance measurement is the inverse of the image acquisition process. Therefore, to estimate the distance of a target, the coordinates of the target on the camera's image plane must first be obtained. Then, utilizing parameters such as the camera's horizontal and vertical field of view, pitch angle, mounting height, and focus, the image plane coordinates are converted into three-dimensional road surface coordinates. The distance between the obstacle target and the camera is subsequently calculated based on these coordinates. It is important to note that the camera's vertical and horizontal field of view, as well as its focus, are fixed parameters, while the pitch angle and mounting height of the camera require manual determination.

3.2.1 Monocular Vision Imaging Model

The three-dimensional road coordinate system is illustrated in Figure 4, where the road surface is denoted by plane ABK. The camera positioned at point O captures the road area represented by ABCD. G represents the intersection of the camera's optical axis OG with the road surface, while J corresponds to the camera's vertical projection on the road surface. The camera's mounting height is indicated by OJ, and the mapping axis of the road's X-axis on the image plane is defined as the x-axis. Angle EOF signifies the camera's horizontal field of view, with a magnitude of 2β. Point G serves as the origin of the coordinate system, with the positive Y-axis indicating the direction of vehicle advancement. Assuming point P (XP, YP) as the obstacle target, points N and M refer to the projections of point P on the Y-axis and X-axis, respectively.

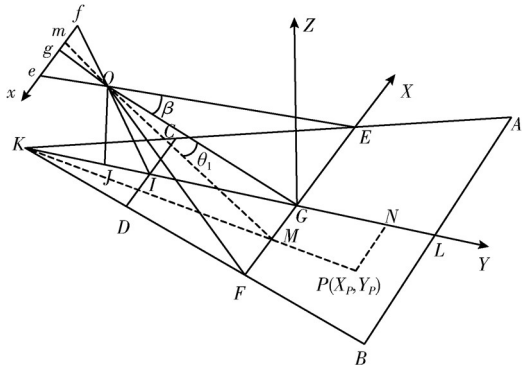

Figure 4. 3D road coordinate system.

The image captured by the camera is depicted in Figure 5, where the rectangular region abcd represents the imaging plane of the road surface ABCD. The point p(xp, yp) in the image plane corresponds to the point P(XP, YP) in the road coordinate system. Furthermore, the points e, f, g, i, l, m, and n in the image plane correspond to the points E, F, G, I, L, M, and N in the road coordinate system, respectively. H and W denote the height and width of the image plane, respectively.

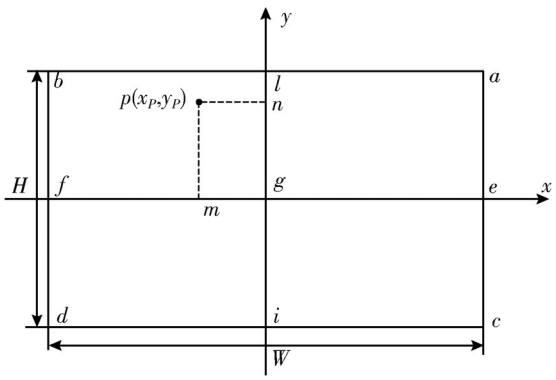

Figure 5. 2D image plane coordinate system.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202406_10(6).0009

The imaging model of the camera in the positive lateral direction of the Y-axis is depicted in Figure 6. The projections of the target object on the Y-axis and the y-axis are indicated by the points N and n, respectively, highlighting the relationships GN = YP and gn = yp. The angle ∠GON, formed between the optical axis OG and the projection point N of the target on the Y-axis, is defined as θ0. Moreover, the camera's vertical field of view, denoted by ∠IOL, corresponds to an angle of 2α. The camera's pitch angle, represented by ∠JOG, is equal to γ. Additionally, the y-axis represents the mapping axis of the road's Y-axis onto the y-axis of the image plane.

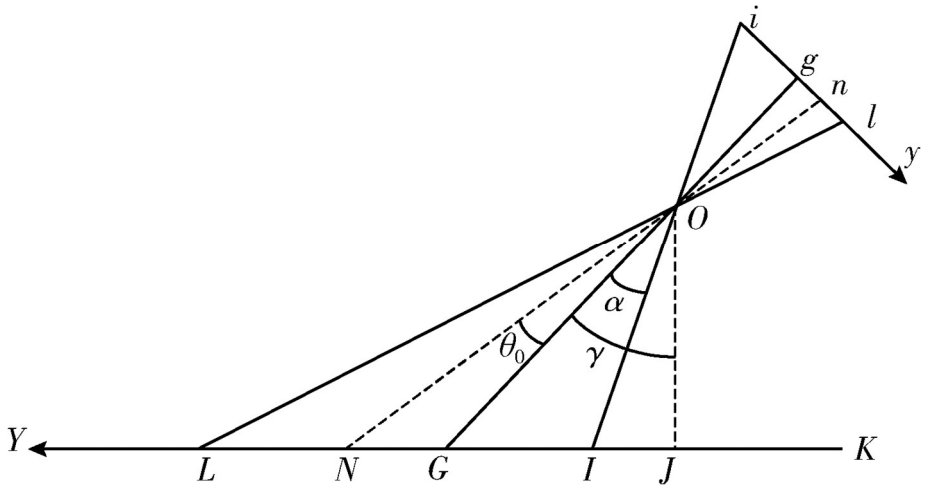

Figure 6. Y-axis direction front and side camera imaging model.

3.2.2 Cubic Curve Fitting of Measurement Data

Within the image plane coordinate system illustrated in Figure 5, it is observed that as the detection target moves further away from the camera, the distance information represented by the pixels in the image becomes more significant. However, this also causes a degradation in the clarity of information features, leading to notable deviations in the predicted positional information of distant targets by the detection algorithm. To enhance the accuracy of ranging, this study utilizes a third-degree Bezier curve equation to fit the actual measured distance of obstacles with the algorithmically corrected predicted distance. The process involves data interpolation to construct equations that effectively reduce ranging errors.

4. Experimental Results

To assess the efficacy of our approach in detecting road obstacles and estimating the distance, we conducted evaluations using a most commonly used dataset. Given our emphasis on identifying unknown anomaly objects, we purposely abstained from utilizing any prior knowledge related to road obstacles during the training phase.

4.1 Obstacle Detection Results

In order to conduct a comprehensive evaluation of our road obstacle detection method, we employed the publicly available dataset "Lost and Found" [6]. This dataset offers precise human-marked labels as ground truth instead of using conventional bounding boxes to delineate road obstacle regions. Following a well-established methodology [20], we quantified the accuracy of the detected road obstacle regions by comparing the resulting anomaly maps to the ground-truth anomaly annotations using ROC curves and the AUROC metric. Our evaluation included a comparative analysis against representative existing methods (Resynth [20], restricted Boltzmann machine [15], and uncertainty-based method [17]) as baselines. Furthermore, we examined the performance of our approach with both obstacle score maps (with postprocessing) and anomaly maps alone (without postprocessing).

ISSN: 2414-1895 DOI: 10.6919/ICJE.202406_10(6).0009

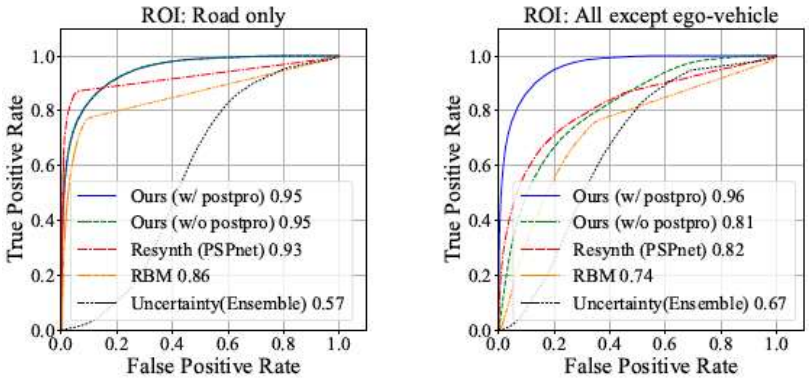

Figure 7. ROC curves and AUROC scores for the Lost and Found dataset.

Figure 7 illustrates the ROC curves and corresponding AUROC scores obtained through the evaluation of the methods. The left set of curves represents the evaluation confined to the road area, defined by the ground-truth annotations. Likewise, the right set of curves showcases the evaluation conducted across the entire images, excluding the regions occupied by the ego-vehicle. Comparatively, our approach without postprocessing exhibits performance on par with Resynth [20] and outperforms the other methods. Moreover, by incorporating postprocessing, our approach attains the highest AUROC scores among all the methods. Notably, the employment of postprocessing leads to a substantial improvement in road obstacle detection performance, as demonstrated in Figure 7.

In Figure 8, we present an example that demonstrates the generated maps for an image featuring a road obstacle, primarily situated in the middle of the road (a). However, the semantic segmentation module inaccurately assigns labels to the road obstacle region (b). Subsequently, the resynthesized image generation module generates an image exhibiting significant dissimilarities in appearance compared to the road obstacle region in the original input image (c). Consequently, the anomaly calculator produces relatively high anomaly scores surrounding the road obstacle (d). Finally, the postprocessor accentuates the road obstacle by amplifying the anomaly score exclusively within the road area while suppressing the anomaly score in other regions (e). The presence of a yellow car parked sideways introduces misclassification errors in the semantic segmentation process. These misclassifications not only compromise the quality of the resynthesized images but also contribute to false positives in road obstacle detection.

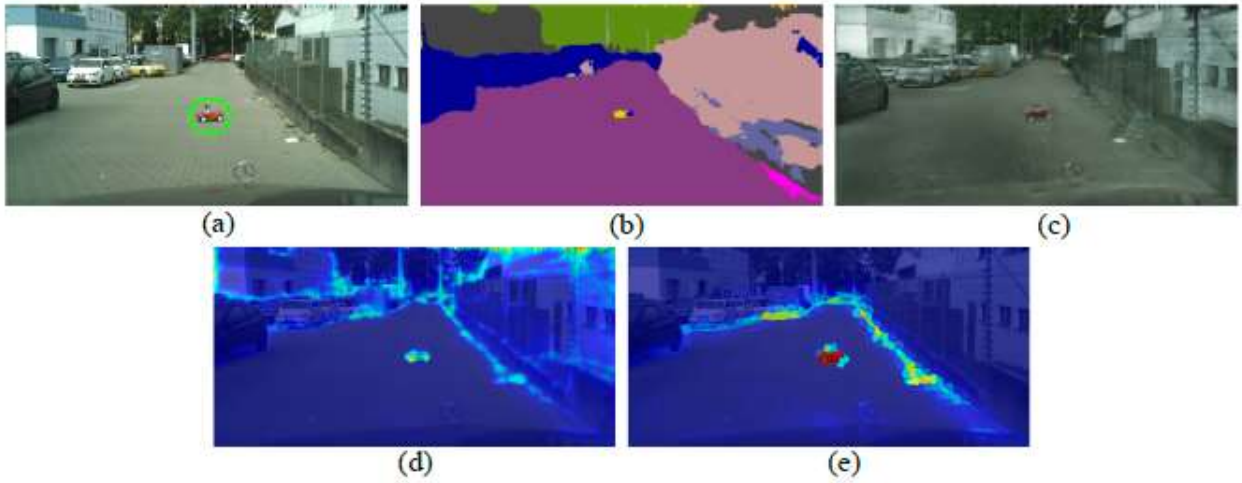

Figure 8. Example of maps generated for a synthetic image with a road obstacle. (a) Input image, (b) semantic map, (c) resynthesized image, (d) anomaly map, and (e) obstacle score map.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202406_10(6).0009

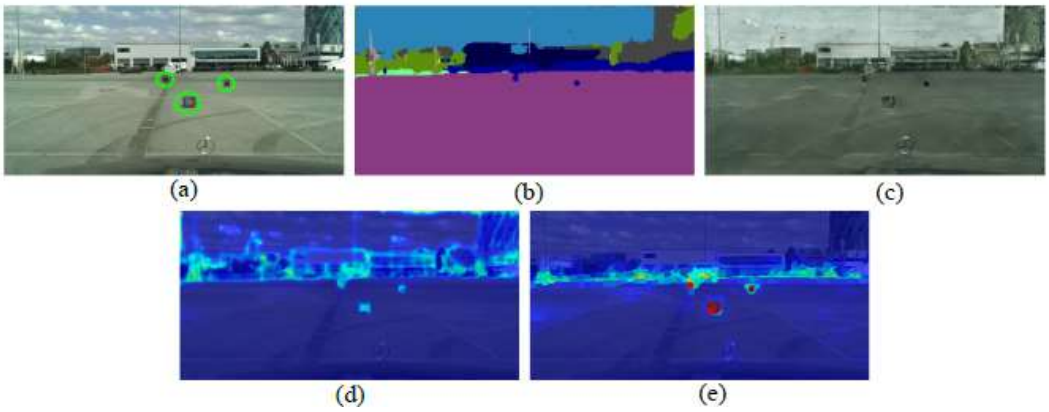

Figure 9. Example of maps generated for a synthetic image with multiple road obstacles. (a) Input image, (b) semantic map, (c) resynthesized image, (d) anomaly map, and (e) obstacle score map.

The maps depicted in Figure 9 showcase an example where an image contains multiple obstacles (a). Even though the obstacles are relatively small, the anomaly score accurately reflects their presence (e). In line with the previous observations, misclassifications occur in the proximity of the obstacle regions (b), thereby impacting the quality of resynthesized images (c).

4.2 Obstacle Distance Estimation Results

In this study, images were captured using a wide-angle camera with a resolution of 4032×3024 pixels, a focal length of 4.216ms, and an aperture size of f/1.8. The camera provided a horizontal field of view of 65° and a vertical field of view of 59.6°. The camera was mounted at a height of 1.2m with a pitch angle of 87.9°. The implemented program was based on Python 3.6, utilizing the YOLOv4 object detection algorithm implemented in the TensorFlow deep learning framework (version 1.11). The parallel computing library used was CUDA Toolkit 9+cuDNN 7.1. Detected obstacles were represented by rectangular bounding boxes indicating their positions, and the object recognition and distance estimation results were displayed within these boxes.

To evaluate the performance of the proposed method, obstacle cars were strategically placed on the road at distances ranging from 5m to 80m from the camera. Distance measurements were conducted at 5m intervals, and additional measurements were taken by randomly placing vehicles on the road surface. A total of 60 sets of actual distance measurements and initial algorithm-based distance measurements were recorded, resulting in a dataset of 120 distance values. Table 1 presents 14 representative sets of distance data in the first and second columns. The third column of Table 1 shows the distance measurements obtained by applying the Canny edge detection operator and correcting the position of the detected objects' lower edge. The results demonstrate that as the distance of the detected obstacles increases, the measurement error caused by inaccuracies in the detection bounding box position also increases. However, the proposed method for adjusting the detection bounding box position effectively reduces the measurement error. The average error decreased from 6.21m before the adjustment to 1.504m after the adjustment.

The experimental results indicate that within the range of actual distances from 5m to 50m, the average distance error is 0.5356m, with a maximum error of 1.0952m. However, beyond a distance of 50m, the measurement error increases due to the reduced pixel area occupied by the detected targets in the images. Specifically, at an actual distance of 70m, the maximum distance measurement error reaches 2.362m. Comparing the distance measurement results of this study with the data presented in Table 1 of reference [4], Table 4 of reference [6], and Table 1 of reference [15], a significant reduction in distance measurement errors can be observed for the same distances. This improvement can be attributed to two factors. Firstly, this study incorporates target position correction to mitigate the positional deviation caused by the object detection algorithm. In contrast, the differential detection method used in reference [15] fails to accurately determine the actual target position and does not

ISSN: 2414-1895 DOI: 10.6919/ICJE.202406_10(6).0009

include position correction. Secondly, the distance measurement algorithm in this study is optimized, and the measurement data is subjected to curve fitting, resulting in distance measurements that closely align with the actual distances. Conversely, the distance measurement methods employed in reference [4] and reference [6] are susceptible to horizontal vehicle offsets, which negatively impact their accuracy.

Table 1. Distance measurement results.

Actual distance (m)	Initial distance (m)	Distance after correction (m)	Distance after curve fitting (m)	Final distance error (m)	Final error (%)
5	5.3091	5.1770	5	0	0
10	9.0496	9.4581	9.2509	0.7491	7.491
15	14.2977	14.9761	14.7970	0.2030	1.353
20	21.2773	20.6372	20.5371	0.5371	2.685
25	25.9248	25.5887	25.5730	0.5730	2.292
30	28.5890	29.3857	29.4290	0.5710	1.903
35	41.6549	35.9956	36.0952	1.0592	3.129
40	48.0998	40.7217	40.7996	0.7996	1.999
45	50.0178	44.7131	44.7160	0.2840	0.631
50	48.5291	50.7808	50.5448	0.5448	0.109
55	61.6844	53.3838	52.9917	2.0083	3.651
60	78.9836	62.3644	61.1455	1.1455	1.909
70	93.6298	75.7761	72.3620	2.3620	3.374
80	90.7885	85.9256	80	0	0

5. Conclusion

This study introduces an innovative approach to road obstacle detection and distance estimation using an unsupervised autoencoder with semantic segmentation for forklift. Unlike conventional methods, our proposed approach does not rely on any prior knowledge of road obstacles, making it highly versatile. The method leverages a single-color image captured by a standard in-vehicle RGB camera as its input. By employing an autoencoder architecture consisting of a semantic image generator as the encoder and a photographic image generator as the decoder, a resynthesized image is generated. The method then calculates the perceptual loss between the input and resynthesized images and incorporates the entropy of the semantic image to generate an anomaly map. Finally, the approach employs visual-saliency-based postprocessing techniques to accurately localize road obstacles within the image.

In addition, we propose an innovative method for distance measurement to obstacles in front of a moving vehicle using a monocular vision system. We address the limitations of conventional obstacle detection algorithms, which suffer from poor generalization and can only detect individual obstacles, by employing the YOLOv4 algorithm, a deep learning-based object detection method. To improve the accuracy of object localization, we enhance the Canny operator for edge detection, enabling us to precisely identify the actual position of the target. Furthermore, we adjust the target position by considering the vertical coordinate difference of the bottom edge of the detection box. By leveraging the camera projection model and employing geometric derivation, we establish a ranging model and utilize a third-order Bezier curve to fit the ranging data.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202406_10(6).0009

Through extensive experimentation, we have substantiated that the proposed method achieves comparable performance to existing techniques, even without the inclusion of optimization. Remarkably, when optimization is integrated, our method surpasses state-of-the-art approaches on a prominent publicly available dataset. This unsupervised deep-learning-based road obstacle detection and distance estimation method stands as a practical solution that propels the progress of autonomous driving systems for forklift.

References

[1] Ministry of Land, Infrastructure, Transport and Tourism: Number of fallen objects handled by expressway companies in 2018.

[2] Zhao, H., Qi, X., Shen, X., Shi, J., Jia, J.: ICNet for real-time semantic segmentation on high-resolution images. In: The European Conference on Computer Vision (ECCV). (2018).

[3] Chen, Q., Koltun, V.: Photographic image synthesis with cascaded refinement networks. 2017 IEEE International Conference on Computer Vision (ICCV) (2017) 1520–1529.

[4] Goferman, S., Zelnik-Manor, L., Tal, A.: Context-aware saliency detection. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. (2010) 2376–2383.

[5] Masao, Y.: Salient region detection by enhancing diversity of multiple priors. IPSJ Transactions on Mathematical Modeling and Its Applications 9 (2016) 13–22.

[6] Pinggera, P., Ramos, S., Gehrig, S., Franke, U., Rother, C., Mester, R.: Lost and found: detecting small road hazards for self-driving vehicles. In: 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). (2016) 1099–1106.

[7] Shutoko; Metropolitan Expressway Company Limited: Current state of road obstacles.

[8] Hancock, J.: High-speed obstacle detection for automated highway applications. Technical report, CMU Technical Report (1997).

[9] Williamson, T., Thorpe, C.: Detection of small obstacles at long range using multibaseline stereo. In: IEEE International Conference on Intelligent Vehicles. (1998).

[10] SUBARU: EyeSight. (http://www.subaru.com/engineering/eyesight.html).

[11] Yoffie, D.B.: Mobileye: The Future of Driverless Cars. HBS CASE COLLECTION. Harvard Business School Case (2015).

[12] Tokudome, N., Ayukawa, S., Ninomiya, S., Enokida, S., Nishida, T.: Development of real-time environment recognition system using lidar for autonomous driving. (2017) 1–4.

[13] Velodyne: HDL-64E. (http://velodynelidar.com/lidar/).

[14] Munawar, A., Vinayavekhin, P., Magistris, G.D.: Limiting the reconstruction capability of generative neural network using negative learning. 2017 IEEE 27th International Workshop on Machine Learning for Signal Processing (MLSP) (2017) 1–6.

[15] Creusot, C., Munawar, A.: Real-time small obstacle detection on highways using compressive rbm road reconstruction. 2015 IEEE Intelligent Vehicles Symposium (IV) (2015) 162–167.

[16] Alex Kendall, V.B., Cipolla, R.: Bayesian segnet: Model uncertainty in deep convolutional encoder-decoder architectures for scene understanding. In Tae-Kyun Kim, Stefanos Zafeiriou, G.B., Mikolajczyk, K., eds.: Proceedings of the British Machine Vision Conference (BMVC), BMVA Press (2017) 57.1–57.12.

[17] Lakshminarayanan, B., Pritzel, A., Blundell, C.: Simple and scalable predictive uncertainty estimation using deep ensembles. In Guyon, I., Luxburg, U.V., Bengio, S.,Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R., eds.: Advances in Neural Information Processing Systems 30. Curran Associates, Inc. (2017) 6402–6413.

[18] Akcay, S., Atapour-Abarghouei, A., Breckon, T.P.: Ganomaly: Semi-supervised anomaly detection via adversarial training (2018).

[19] Schlegl, T., Seeb¨ock, P., Waldstein, S., Schmidt-Erfurth, U., Langs, G.: Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. (2017) 146–157.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202406_10(6).0009

[20] Lis, K., Nakka, K., Fua, P., Salzmann, M.: Detecting the unexpected via image resynthesis. In: The IEEE International Conference on Computer Vision (ICCV). (2019).

[21] Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J.: Pyramid scene parsing network. In: CVPR. (2017).

[22] Wang, T.C., Liu, M.Y., Zhu, J.Y., Tao, A., Kautz, J., Catanzaro, B.: Highresolution image synthesis and semantic manipulation with conditional GANs. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2018).

[23] Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., Schiele, B.: The cityscapes dataset for semantic urban scene understanding. In: Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2016).

[24] Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: International Conference on Learning Representations. (2015).

[25] Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., S¨usstrunk, S.: Slic superpixels (2010).