2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)
Windsor Oceanico Hotel, Rio de Janeiro, Brazil, November 1-4, 2016

Lane Marking Detection based on Convolution Neural Network from Point Clouds

Bei He[1], Rui Ai[1], Yang Yan[1] and Xianpeng Lang[1]

***Abstract*— Lane marking detection, as a key technique for the Highly Automated Driving (HAD) Map, has drawn much attention recently. Traditional methods mainly focused on processing color images, where the performance was affected by illumination variation, over-exposure and occlusion severely. Confronted with above problems, this paper proposes a lane detection algorithm based on Convolution Neural Network (CNN) from point clouds. Our contributions are twofold. On one hand, a CNN framework via gradual up-sampling is introduced, where robust and accurate detection results are achieved. Before applying the CNN model, we also design pre-processing steps, including point clouds registration, the road surface segmentation and orthogonal projection. On the other hand, we propose to analyze the layout of lanes by utilizing the global information and domain knowledge. Hence, false detections caused by ground arrows and texts could be eliminated. Visual and quantitative experiments demonstrate the effectiveness of our algorithm.**

I. Introduction

Highly Automated Driving Map (HAD Map) aims to provide accurate and robust road information, including road surface markings, roadside barriers and various traffic signs [1]. It plays a vital role in autonomous driving assistance systems, which draws much attention of researchers and engineers recently. As one of the most key elements, lane markings refer to club-shaped structures with high reflectivity values on the road surface which can be utilized to manage and control traffic activities.

Traditional methods usually detected lane markings from color images or videos [2], [3], [4]. In general, there exist two main reasons for detecting from color images [5]. First, digital cameras are much cheaper than range scanners. Second, there are many mature methods, which can detect targets from color images or videos well. However, various challenging problems, including illumination variation, over-exposure and occlusion by cars/pedestrians, affect the performance of detection methods severely [6]. Consequently, many methods tried to extract lane markings from point clouds by directly processing them represented by reflectivity values [7], or transforming to 2D geo-referenced reflectivity images [8], [9]. Recently data acquisition vehicles for the digital map are usually equipped with a laser scanner, several cameras and a GNSS/IMU system. The collecting system gathers color images, point clouds, position, heading and attitude data at the same time. Compared with the digital camera, the laser scanner can collect more accurate and rich information of the 3D scene [7]. Hence, we can register point clouds and easily obtain a top-view image of digital 3D roads, where lane markings own different reflectivity values compared with others. Therefore, we try to detect lane markings from point clouds with the reflectivity values.

The Convolution Neural Network (CNN) outperforms traditional methods in many computer vision tasks [10], [11], [12], [13], [14], [15]. Known as its high accuracy and robustness, CNN is usually designed as an end-to-end framework. Researchers don't have to explicitly design complex features or process models for different tasks, since features can be learned automatically. In addition, CNN can process images fastly with a Graphics Processing Unit (GPU).

Aimed at problems of existing methods, this paper presents a lane marking detection algorithm based on CNN. The contributions are twofold. On one hand, we propose a novel CNN framework to detect lane markings from point clouds. After point clouds registration, the road surface segmentation and orthogonal projection, the reflectivity image is generated, where the CNN model can be applied. The gradual up-sampling strategy of the CNN model is designed for robust and precise detection. On the other hand, considering the CNN model cannot utilize the global information and domain knowledge, we analyze the layout of the road surface. Accurate detection results are achieved by utilizing the length, distance and angular difference simultaneously, where false alarms of ground arrows and texts can be eliminated.

This paper is organized as follows. In Section 2, we review related works about lane marking detection. In Section 3, we introduce our method, including pre-processing point clouds to obtain the reflectivity image, designing the CNN model for detection and analyzing the topology of the road surface. We discuss the experimental results in Section 4. Finally, we conclude our work in Section 5.

II. Relate Works

Researches of lane marking detection can be classified into two categories. One extracts lane markings from color images or videos [2], [4], [1]. The other tries the detection from point clouds, such as [8], [9], [7].

The method proposed by [2] was a representative work of lane marking detection. Firstly, [2] generated the top-view image from the original color one, and then filtered it with the difference of Gaussian filter. The filtered image was binarized, where the vertical and straight candidate lines were extracted by simplified hough transformation. Finally, a post-processing was adopted to refine splines in top-view and original images simultaneously. The disadvantage was

[1]The authors are from Baidu L3 Business Unit, China. {hebei01, airui01,yanyang01,langxianpeng}@baidu.com

Fig. 1. The schematic of our algorithm.

that it easily preserved wrong lines, such as ground arrows on the road surface and guard bars beside the road boundary. [6] discussed recent progresses of road and lane marking detection in past several years. More methods utilized the LiDAR and IMU with the global positioning information as important complements for the image-to-world correspondence. However, extracting lane markings still relied on traditional computer vision methods to process color images. [1], [4] applied the CNN model to improve the robustness of detection results. Due to the restricted view of the camera, complete and accurate shapes of objects on the road surface cannot be described well, such as solid lane markings, ground arrows and diversion lines. Additionally, those methods were also affected by illumination variation and occlusion on the road surface severely.

As an active remote sensing technology, laser scanning can capture 3D geo-spatial data with unprecedented details of the road. From point clouds, it was easy to extract accurate shapes of lane markings. [8] was a early work about traffic sign and lane marking detection based on the collected LiDAR data. It detected traffic signs from point clouds by clustering, and fit a plane using the RANdom Sample Consensus (RANSAC) algorithm to remove outliers. Furthermore, lane markings were segmented from the reflectivity image via utilizing hough transformation. The method in [9] contained 3 steps for lane marking detection. The first step was curb extraction, where original point clouds were partitioned based on the vehicle trajectory data and detected curbs. The second step generated the geo-referenced reflectivity image. The third step extracted lane markings by introducing a point-density-dependent segmentation method. Compared with them, [7] designed 2 steps to extract lane markings from point clouds. The first step was to segment lane markings directly from the road surface of point clouds based on the difference of reflectivity values. [7] combined a global threshold and multiple local thresholds for binarization. The second step was to cluster those lane markings into different categories with deep boltzmann machines and the Principal Component Analysis (PCA) methods. The test scenes in that paper were very ideal. First, their point clouds were denser than what we captured for the HAD Map production. Second, lane markings on their point clouds were very clear, without occlusion or wear. Thus, this method cannot deal with real-world scenes in the HAD Map production well.

In conclusion, existing methods mainly aimed to detect lane markings for autonomous driving and concentrated on the read-time performance. Since they were not designed for the HAD Map production, the accuracy and completeness of road markings were not guaranteed. Therefore those methods cannot be adopted directly. Consequently, this paper proposes a novel method for effective lane marking extraction based on the reflectivity image of point clouds.

III. Proposed Algorithm

Our algorithm comprises 3 steps, which is illustrated in Fig.1. First, the collected point clouds should be pre-processed. With the position, heading and attitude data recorded by GNSS/IMU, we register frames of point clouds together to obtain dense results. Via orthogonal projection, we extract the road surface and obtain the top-view reflectivity image. Second, we propose a CNN model to detect candidate lane markings from the above image. Third, we analyze the layout of lanes and classify them into different groups of lanes. False detections, such as ground arrows and texts on the road surface, are removed by the restriction of the length and spatial position. In the meanwhile, lanes' left and right components can also be distinguished.

A. Pre-process Point Clouds

Point clouds of a single frame are sparse, where global information cannot be utilized. As a result, we use the recorded GNSS/IMU data, including the position, heading and attitude, to register frames of point clouds firstly. From the dense point clouds, the road surface is segmented by fitting several local planes. The details of the segmentation method can be referred in [16]. We only preserve point clouds of road surfaces, while other objects, e.g. cars, barriers and trees are removed. Therefore, the occlusion and false detection results brought by those decrease naturally. Then we generate the top-view image via the orthogonal projection along the z-axis, and guarantee the y-axis is parallel to the driving direction. In that case, the lane markings are also parallel to the y-axis approximately. During imaging, we apply the reflectivity values in consideration of the fact that reflectivity difference between lane markings and others is obvious. Most traditional problems of detection methods from color images or videos, such as illumination variation, over-exposure and the occlusion by cars/pedestrians, can be resolved.

Especially, we propose to normalize the intensity values with the mean and standard deviation of all point clouds. Then the values are re-scaled to $[0, 255]$ according to the minimal and maximal ones. Our algorithm generates the reflectivity image with the resolution 6144×3072 for each $6144 \times 3072cm^2$ patch. Hence, one pixel is equivalent to $1cm^2$ region in the real world, which meets the demands of our HAD Map production (Of course, the scale can be adjusted according to different application demands).

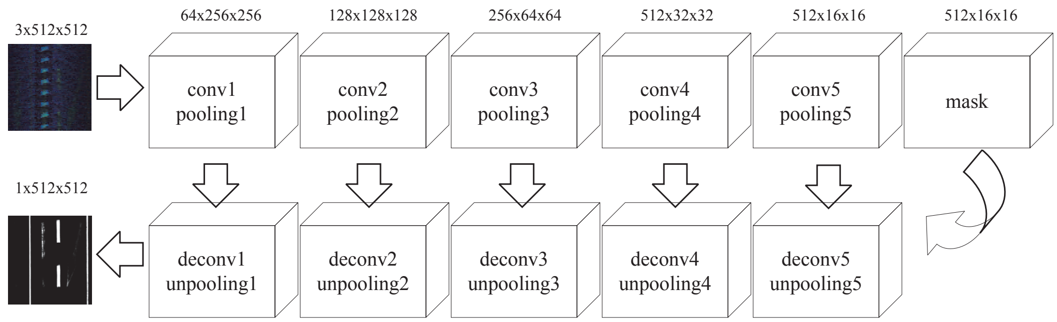


Fig. 2. The convolutional neural network structure for lane marking detection.

B. CNN Model for Lane Marking Detection

Recent methods [4], [1] introduced the CNN model into lane marking detection from color images or videos. In their framework, the output patch is much smaller than the input one. There exist 2 reasons. On one hand, the traditional CNN framework was designed for classification [10], [17]. The iteration cannot converge when the feature dimension is too high. On the other hand, up-sampling is not an easy task, where huge parameters need to be learned [13], [14]. Furthermore, segmentation based on CNN cannot be utilized directly, since club-shaped structures are much harder to be drawn out than blob-shaped ones. Consequently, [4] applied the bi-linear interpolation on the low-resolution mask, while [1] preferred to segment by sliding windows with dense overlaps.

Aimed to solve problems of low accuracy [4] or high computation cost [1] brought by above methods, this paper proposes a novel CNN framework. Inspired by [18], we utilize the original information by gradual up-sampling. As shown in Fig.2, each up-sampling layer is composed of de-convolution and un-pooling layers, which combines the original and down-sampling images. During training, we crop reflectivity images with the resolution 6144×3072 into 512×512 patches. Then the small patches are fed into our CNN framework, where the outputs are the same size. Our framework is indeed an end-to-end one, where the output patch can be utilized for detection directly. To be mentioned, the parameters should be learned layer-by-layer so that the iteration can converge safely.

During detection, we slide the window on the whole image with the resolution 6144×3072, where no overlap between windows is required. Hence, we only need to repeat 12×6 times for each image. Compared with [1], the 4×4 CNN outputs are replaced with 512×512 ones, where the computation cost is reduced significantly. For each detection window, we crop the patch and apply the CNN framework to solve it. Hence, each pixel is assigned with the corresponding probability ranging from $[0, 1]$. To avoid the dominate noise to affect the following analysis, we add a Gaussian filter to smooth outputs. Via binarization, component connectivity analysis and line fitting, we can obtain candidate lane markings easily. The lane markings are denoted by the three-order curves in this paper.

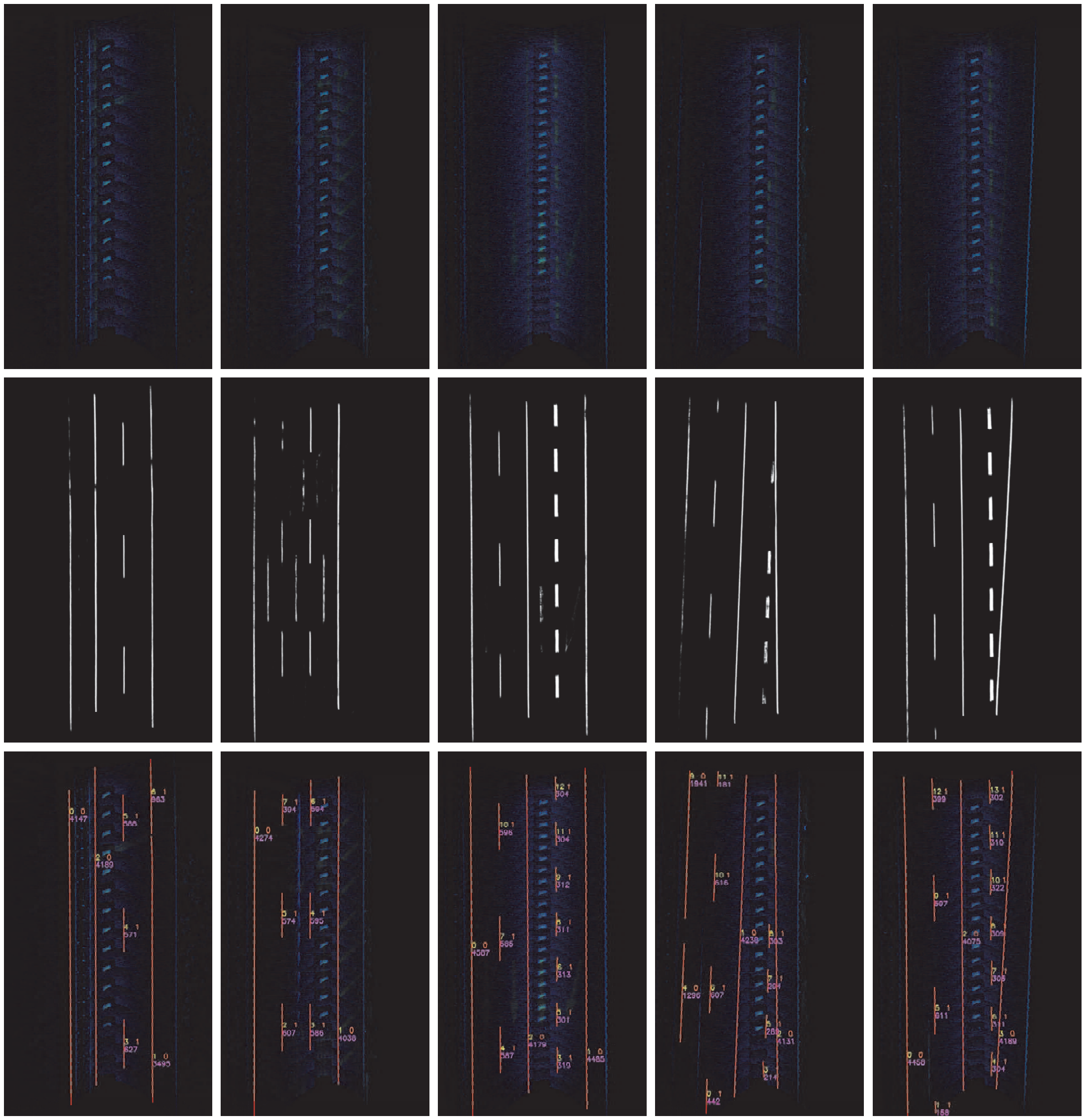
C. Analyze the Layout of Lanes

The above CNN framework extracts short line segments, which may belong to broken lane markings or worn solid ones. Firstly, we connect the nearest line segments to form integral lane markings along y-axis. Hence, broken lane markings can be re-constructed, when the worn solid lane markings are restored in the meanwhile.

The CNN framework also detects road markings, including ground arrows and texts, which should be removed. To decide whether a candidate lane marking is correct, the global context, called by the topology of lanes, is utilized. Consequently, we investigate the layout of lanes in the 3D space. We remove significantly false candidate lane markings based on 2 rules. First, if the length of a candidate lane marking is less than $0.5m$, it should be eliminated. That is because too short lane markings cannot be fitting accurately. Second, since the y-axis is parallel to the driving direction, lane markings with large angles between them and the driving direction are removed. The threshold is fixed as $\pi/4$ in our experiments.

In order to remove residual wrong lane markings, this paper proposes to derive higher-level information, which is described as follows.

1) We link short line segments to form lane markings, which means each lane marking may consist of one or more line segments sharing the same direction. Then, the length, the confidence and the principle direction are calculated for each lane marking. We sort the candidate lane markings from right to left according to the corresponding average x coordinates.
2) Among all, 2 lane markings are selected to compose one lane. For each lane marking, we search its left-neighboring one and compute the distance between them as the width of the constructed lane. The average lane width is $3.7m$ in our experiments. If the distance between two lane markings is less than $3.7m - 1.5m$ or larger than $3.7m + 1.5m$, we would ignore them. Otherwise, if the confidence values c_f of those lane markings are both larger than 0.5, we let them compose a lane.
3) Then we explore whether a lane marking is in the middle of the new constructed lane. If a lane marking is in the middle of a lane, and the distance from either



(a) (b) (c) (d) (e)

Fig. 3. Detection results of our algorithm for straight lane markings. (a)-(e) correspond to scenes of the speed-limiting band, ground arrows, ground texts, diversity lines and the fork. For each scene, the reflectivity image, the CNN output and the final result are listed from top to bottom, respectively. The index, the length (pixel/cm) and the type of a lane marking are plotted by yellow, orange and red colors.

side of the lane is larger than $0.5m$, we regard this as a wrong lane marking. We delete that along with all children line segments. However, if above distances are both less than $0.3m$, that lane marking can compose the one of a double lane marking, which means it should not be deleted. We go back to 2) until no new lane could be found.

The confidence value of a candidate lane marking is defined as:

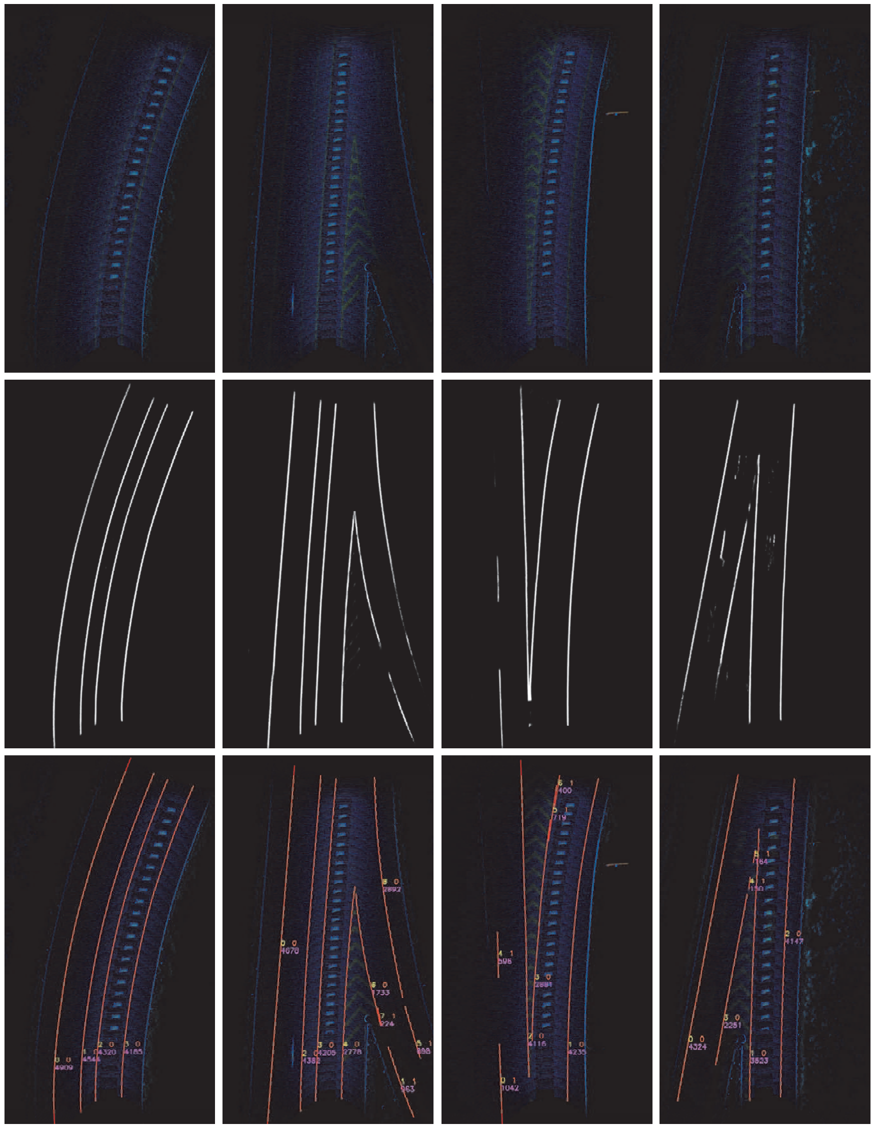
$$c_f = \sum_{i=1}^{n} \frac{f_c(i)}{n} \times l/H, \tag{1}$$

where $f_c(i)$ is the output probability of the CNN model for the pixel i within the lane marking region. n, l and H refer to the total number of pixels belong to the lane marking, the corresponding length and the height of the reflectivity image respectively.

IV. Experiments and Discussions

We implemented our algorithm in C++, ran it on a $K40$ GPU and generated 12729 reflectivity images of the highway as the dataset. The true lane markings are annotated manually. We select 2729 from them for training, while the rests are applied for testing. Visual and quantitative results are listed in order. Lane marking detection results of our algorithm in various scenes are shown in Fig.3 and Fig.4. Those 2 figures refer to straight and curve lane markings respectively. In the following, the recall and precision ratios of Yu's [7] and ours are compared.

Only the road markings on the road surface may affect the performance of our lane marking detection after the pre-process module. In Fig.3, scenes of the speed-limiting band,



(a) (b) (c) (d)

Fig. 4. Detection results of our algorithm for curve lane markings. The shapes of curves are various. For each scene, the reflectivity image, the CNN output and the final result are listed from top to bottom, respectively. The index, the length (pixel/cm) and the type of a lane marking are plotted by yellow, orange and red colors.

ground texts, ground arrows, diversity lines and the fork are illustrated from (a) to (e). Due to the robustness of the CNN framework, lane markings, along with several road markings, are highlighted from the road surface. As verified in the second rows, pixels belonged to the true lane markings are strengthened, even when the reflectivity values are weak, e.g., in the (d). In the outputs of the CNN framework, curbs, blank road surfaces and residual artifacts of cars are all eliminated. However, ground texts and arrows are still preserved, as shown in (c) and (d). Our algorithm utilizes the layout of lanes and removes those false alarms via length and spatial position filters. In the meanwhile, the lengths and types of lane markings are calculated. The yellow, red and orange colors refer to the index, type and length (pixel/cm) of a lane marking.

Our detection also reports good performance on various curve lane markings in Fig.4. The CNN framework can extract lane markings from road surfaces robustly and accurately. The reflectivity values of lane markings in (b) and (c) are weak. However, in the CNN outputs, those pixels gain high probabilities, which helps the following layout analysis. Although ground arrows and speed limit bands are preserved, we apply the global context to re-construct lanes in the 3D space. Hence, true lane markings can be detected correctly.

Quantitative comparison of detection results between Yu's [7] and ours is shown in Table.I. The recall and precision ratios are calculated according to the pixels instead of the number of the lines since the length of lane marking is infinite. Pixels in the lane markings are regarded as the right ones if the absolute distances between them and the ground truth are less than the threshold. To meet the demand of our HAD Map production, the threshold is fixed as $10cm$. From

TABLE I
QUANTITATIVE COMPARISONS OF YU'S [7] AND OURS.

Recall		Precision	
Yu's [7]	Ours	Yu's [7]	Ours
85.31%	**93.80%**	81.47%	**95.49%**

the table, we can figure out that our algorithm outperforms Yu's [7] both on the recall and precision ratios owing to our novel CNN framework and the analysis of the 3D layout. In experiments, the time cost of our algorithm to process one reflectivity image is $28.2s$ averagely which is enough for our off-line map production. To sum up, visual and quantitative results demonstrate our algorithm's high accuracy and robustness.

V. CONCLUSION

This paper designs a lane detection algorithm, which introduces the CNN framework on points clouds. First, we pre-process the original points clouds, via point cloud registration, the road surface segmentation and orthogonal projection. Then the generated reflectivity images can be fed into the CNN framework. Second, we add the gradual up-sampling method into the traditional CNN model, where robust and accurate detection results are calculated. Due to the same size of input and output patches, the time cost of our algorithm is low actually. Third, considering false alarms caused by ground arrows and texts, the layout of lanes is analyzed. Hence, those wrong lane markings are removed owing to the global information and domain knowledge. Experimental results verify our algorithm's high accuracy and robustness visually and quantitatively.

In the future, we will extend our algorithm to detect road markings simultaneously, such as ground arrows, texts and diversion lines. Several road markings, e.g., diversion lines, are usually large and with irregular shapes on the road surface. It is hard to extract complete and accurate contours from color images or videos, since diversion lines are always occluded by other cars/pedestrians. Hence, the proposed algorithm is more suitable. In the meanwhile, those road markings can restrict the regions of lane markings, where the corresponding accuracy and robustness of lane marking detection can be improved further.

REFERENCES

[1] Brody Huval, Tao Wang, Sameep Tandon, Jeff Kiske, Will Song, Joel Pazhayampallil, Mykhaylo Andriluka, Royce Cheng-Yue, Fernando Mujica, Adam Coates, et al., "An empirical evaluation of deep learning on highway driving," *arXiv preprint arXiv:1504.01716*, 2015.

[2] Mohamed Aly, "Real time detection of lane markers in urban streets," in *IEEE Intelligent Vehicles Symposium*. IEEE, 2008, pp. 7–12.

[3] Bei He, Rui Ai, Yang Yan, and Xianpeng Lang, "Accurate and robust lane detection based on dual-view convolutional neutral network," in *2016 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2016, pp. 1041–1046.

[4] Jihun Kim and Minho Lee, "Robust lane detection based on convolutional neural network and random sample consensus," in *Neural Information Processing*. Springer, 2014, pp. 454–461.

[5] Sibel Yenikaya, Gökhan Yenikaya, and Ekrem Düven, "Keeping the vehicle on the road: A survey on on-road lane detection systems," *ACM Computing Surveys (CSUR)*, vol. 46, no. 1, pp. 2:1–2:43, 2013.

[6] Aharon Bar Hillel, Ronen Lerner, Dan Levi, and Guy Raz, "Recent progress in road and lane detection: a survey," *Machine Vision and Applications*, vol. 25, no. 3, pp. 727–745, 2014.

[7] Yongtao Yu, J. Li, Haiyan Guan, Fukai Jia, and Cheng Wang, "Learning hierarchical features for automated extraction of road markings from 3-d mobile lidar point clouds," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 2, pp. 709–726, Feb 2015.

[8] Xin Chen, Brad Kohlmeyer, Matei Stroila, Narayana Alwar, Ruisheng Wang, and Jeff Bach, "Next generation map making: geo-referenced ground-level lidar point clouds for automatic retro-reflective road feature extraction," in *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, 2009, pp. 488–491.

[9] Haiyan Guan, Jonathan Li, Yongtao Yu, Cheng Wang, Michael Chapman, and Bisheng Yang, "Using mobile laser scanning data for automated extraction of road markings," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 87, pp. 93–107, 2014.

[10] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[11] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jagannath Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2014, pp. 580–587.

[12] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross B Girshick, Sergio Guadarrama, and Trevor Darrell, "Caffe: Convolutional architecture for fast feature embedding.," in *ACM Multimedia*, 2014, vol. 2, p. 4.

[13] Jonathan Long, Evan Shelhamer, and Trevor Darrell, "Fully convolutional networks for semantic segmentation," *arXiv preprint arXiv:1411.4038*, 2014.

[14] Wei Liu, Andrew Rabinovich, and Alexander C Berg, "Parsenet: Looking wider to see better," *arXiv preprint arXiv:1506.04579*, 2015.

[15] Chenyi Chen, Ari Seff, Alain Kornhauser, and Jianxiong Xiao, "Deepdriving: Learning affordance for direct perception in autonomous driving," *arXiv preprint arXiv:1505.00256*, 2015.

[16] Bertrand Douillard, J Underwood, Vsevolod Vlaskine, A Quadros, and S Singh, "A pipeline for the segmentation and classification of 3d point clouds," in *Experimental Robotics*. Springer, 2014, pp. 585–600.

[17] Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[18] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *arXiv preprint arXiv:1511.00561*, 2015.