

# RBNet: A Deep Neural Network for Unified Road and Road Boundary Detection

Zhe Chen<sup>1(✉)</sup> and Zijing Chen<sup>2</sup>

<sup>1</sup> School of Information Technology, UBTECH Sydney Artificial Intelligence Centre,  
The University of Sydney, Darlington, NSW 2006, Australia  
zche4307@uni.sydney.edu.au

<sup>2</sup> Faculty of Engineering and Information Technology,  
Centre for Artificial Intelligence, University of Technology Sydney,  
Ultimo, NSW 2007, Australia  
z.j.chen219@gmail.com

**Abstract.** Accurately detecting road and its boundary on the images is an essential task for vision-based autonomous driving systems. However, prevailing methods either only detect road or add an extra processing stage to detect road boundary. In this work, we introduce a deep neural network, called Road and road Boundary detection Network (RBNet), that can detect both road and road boundary in a single process. In specific, we first investigate the contextual relationship between the road structure and its boundary arrangement and then model them with a Bayesian network. By implementing the Bayesian model, the RBNet can learn to simultaneously estimate the probabilities of a pixel on the image belonging to the road and road boundary. Comprehensive evaluations are carried out based on the well-known road benchmark, which can demonstrate the compelling performance of the proposed method.

**Keywords:** Deep learning · Deep convolutional neural network · Road detection · Boundary detection

## 1 Introduction

Detecting road and its boundaries is the basis for the autonomous vehicles to navigate routes and avoid obstacles. Despite that various sensors are mounted on the vehicle to help the system perceive the environment, visual sensors, such as video cameras, can provide informative cues at a lower cost. Using the monocular colour image captured by the cameras, the goal of road and road boundary detection is to identify whether the pixels belong to the road areas and road boundaries.

To detect road areas, segmentation techniques like [1] are usually introduced to tackle the problem. Using either traditional classifiers [2, 3] or deep neural networks [4, 5], the category of each pixel can be estimated. However, these methods are not aware of the existence of the road boundaries. As for the boundaries, they are not simply the edge of the detected road areas. In general, the road

boundaries can be defined as the marks that split the road and non-road areas, such as the curb stones between vehicle and pedestrian paths and the white lines between the road and parking areas. Detecting road boundaries is challenging because they have various forms. Some studies [6, 7] attempt to identify the road boundary using a trained classifier while others [8, 9] simply refine the contours without knowing their locations. Few studies attempt to consider both road detection and road boundary detection in a unified framework.

In this paper, we summarise the road and road boundary using a Bayesian network and propose the RBNet to tackle the corresponding detection task. A critical observation of our study is that although the road boundaries have diversified appearance, they do contain abundant structural information which is helpful to define the road areas. As a result, we conclude that there exists a contextual connection between the road areas and the road boundaries, which can be formulated as a graphical probabilistic model. Following the concluded model, RBNet is introduced to detect both road and road boundaries in a single step. The training procedure of the RBNet can be formulated as a multi-task learning problem and we share the visual features across different tasks. After properly training the RBNet, we evaluate the effectiveness of the proposed network on the widely-used KITTI road benchmark [10] and report the performance on its official website<sup>1</sup>. Favourable performance can be achieved by the proposed method against other competitive algorithms. Statistical results also verify the existence of the contextual relationship between road boundaries and road areas in the road scene.

## 2 Related Work

The vision-based road and road boundary detection methods can be divided into two groups: model-based and learning-based. Model-based methods tend to build a shape model or appearance model to describe the road. For shape model, the boundaries of the road are commonly represented by curves like Bezier Splines [11] and Cubic Splines [12]. Then random sample consensus (RANSAC) is usually used to find the fittest parameters for the curves. For appearance model, as an example, [13] describes the road as a linear combination of different color planes and the color distribution of each pixel is used to decide whether it belongs to the road. These model-based methods are accurate when similar road pattern appears, but they would be vulnerable to the complicated road scenes. Different from model-based approaches, learning-based methods mainly adopt a trained classifier to distinguish the road areas from non-road areas [3].

With the recent development of deep learning techniques, convolutional neural networks (CNNs) have achieved record-breaking performance in image segmentation-related tasks. With deep CNN, road areas can be effectively segmented from the images [4, 14]. The major obstacle in improving the performance of CNN-based segmentation methods is that high-level semantic features are too coarse to define the contours or the boundaries of an instance. To relieve

<sup>1</sup> [http://www.cvlibs.net/datasets/kitti/eval\\_road.php](http://www.cvlibs.net/datasets/kitti/eval_road.php).

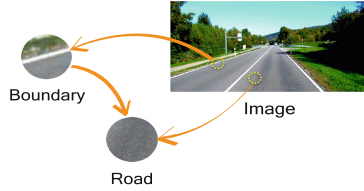
the boundary issue, fully convolutional neural network (FCN) [15] fuses results from low-level feature maps. A similar work [4] facilitates precise localization by following the architecture of “U-net” [16], which consists of a contracting network to capture contexts and a symmetric expanding network to enrich details. Such architectures are helpful for obtaining spatial details but are still weak on boundaries. In another group of studies, the boundaries of the segments are refined using conditional random fields (CRFs) in a post-processing step. CRF-based methods integrate score maps generated by CNN with pairwise features [17], whose inference can be efficiently carried out by high-dimensional filtering [8]. CRF-based methods are advantageous for refining contours but they are not aware of the existence of the boundaries. More closely related studies tend to detect the object boundaries to refine the segmentation. For example, LRR [18] distinguishes boundaries with masking operations and uses Laplacian reconstruction to improve accuracy. [6] first detects obstacle boundaries with CNN and then obtains the road areas by using a graph-cut algorithm. Although these studies are active in identifying boundaries, the contextual relationship between the detected boundaries and the segmentation results is not sufficiently studied. With the consideration of this context, the label noise problem [24] can also be avoided to some extents.

### 3 Road and Road Boundary Detection Network

In this section, we first summarise the road and road boundary detection tasks using the unified Bayesian network model which tends to formulate the relationship among road, road boundaries and the input image in the same probabilistic graph. Following the structure of Bayesian network, we then introduce a deep neural network, called RBNet, to simultaneously detect road and road boundaries.

#### 3.1 The Bayesian Network Model

Tackling the road and road boundary detection separately would be time-consuming and requires carefully designed algorithms to fuse the results for better performance. To relieve this issue, we attempt to simultaneously detect the road areas and road boundaries by formulating them in the unified model. In specific, we find that the identification of road areas is not only influenced by the local appearance but also affected by the road boundaries. For example, if a pixel on the image is enclosed by the road boundaries, it can be considered as the road areas as well, regardless of its visual appearance. However, the road boundaries can not be directly defined based on the edges of the road areas because these areas may not always relate to the road boundaries. As an example, the edges of the road areas may be around the image border instead of actual road boundaries. Therefore, it is more adequate to define the road boundaries only based on the visual appearance. Accordingly, we summarise the relationships among road areas, road boundaries and input image as a Bayesian network, whose detailed structure is illustrated in Fig. 1.



**Fig. 1.** The Bayesian network for road and road boundary detection. In this model, road area, road boundary, and the image are three nodes in the graph. The directed arrows represent the dependencies among the nodes.

Formally, we consider the road, road boundary and input image as three nodes in the Bayesian network. Meanwhile, we refer the road detection and the road boundary detection as the pixel-wise classification. Suppose  $r_{x,y}$  denotes the labeling of the pixel at  $(x, y)$  with respect to road areas on the input image  $I$ . Then segmenting road areas aims at acquiring an assignment  $R = \{r_{x,y}\}$  which allocates 1 to in-road pixels and 0 to the rest. Similarly, suppose  $b_{x,y}$  denotes the labeling of a pixel at  $(x, y)$  which takes 1 if the pixel belongs to the road boundaries and 0 otherwise. We refer  $B = \{b_{x,y}\}$  as the classification results for road boundaries on the whole image. According to the graph illustrated in Fig. 1, the joint probability of  $R$ ,  $B$ , and  $I$  is given as:

$$P(R, B, I) = P(R|B, I)P(B|I)P(I) \tag{1}$$

Therefore, the road detection and the road boundary detection can be solved by estimating by estimating  $P(B)$  and  $P(R)$ . According to the Eq. 1, we have:

$$P(B) = \sum_{R \in \{-,+\}, I \in \{-,+\}} P(R, B, I) \tag{2}$$

$$P(R) = \sum_{B \in \{-,+\}, I \in \{-,+\}} P(R, B, I) \tag{3}$$

Let  $I^+$  represent that there exists road in the image and  $I^-$  represent the opposite. Suppose we are working on the urban road images, we can assume that there is always a road in the camera’s view and thus consider the probability of  $P(I^+)$  as always being 1. As a result, the  $P(B)$  can be computed by:

$$P(B) = P(B|I^+) \tag{4}$$

As for the road detection task, it requires the computation of marginal probability  $P(R)$ . Based on Eq. 3, we have:

$$P(R) = P(R|B^-, I^+)P(B^-|I^+) + P(R|B^+, I^+)P(B^+|I^+) \tag{5}$$

where  $B^+$  is denoted as the collection of pixels which are identified as road boundaries and  $B^-$  is the counterpart. Accordingly,  $B = \{B^+, B^-\}$ .

### 3.2 The Deep Neural Network for Road and Road Boundary Detection

As mentioned above, the road and road boundary detection can be delivered by inferring  $P(B)$  and  $P(R)$  based on Eqs. 4 and 5 respectively. However, empirically estimating the probabilities would be unreliable and arduous because the dependencies in the Bayesian network could be ambiguous due to environmental noises and complex road scenarios. To obtain a faithful estimation for the probabilities, we employ the deep neural network, RBNet, to learn the dependencies statistically.

In order to properly implement RBNet, we first decompose the task of inferring graphical model of Bayesian network into several independent sub-tasks and then introduce the corresponding task-specific sub-networks to solve them. As a result, the training procedure can be formulated as a multi-task learning problem. In the following sections, we denote  $l$  as the loss function for each pixel.

**Road Detection.** The road detection task can be achieved by inferring  $P(R)$  based on Eq. 5. The inference requires the sum of  $P(R|B^+, I^+)P(B^+|I^+)$  and  $P(R|B^-, I^+)P(B^-|I^+)$ . Computing  $P(R|B^-, I^+)$  and  $P(R|B^+, I^+)$  is simple and straight forward. On one hand, the  $P(R|B^-, I^+)$  stands for the probability of each pixel in the image that belongs to the road area when no boundary is detected, which can be interpreted as the common segmentation task. On the other hand, the estimation of  $P(R|B^+, I^+)$  can be regarded as the prediction of road areas only based on the road boundary detection results. However, for example, computing  $P(R|B^+, I^+)P(B^+|I^+)$  should not be regarded as element-wise multiplication between the  $P(R|B^+, I^+)$  and  $P(B^+|I^+)$ , because a road boundary pixel may have effects on the road pixels in a larger image region. To properly infer  $P(R)$ , we rewrite the Eq. 5 in the following form:

$$P(R) = P(R|B^-, I^+) + (P(R|B^+, I^+) - P(R|B^-, I^+)) P(B^+|I^+) \quad (6)$$

given the fact that  $P(b_{x,y}^+|I^+) + P(b_{x,y}^-|I^+) = 1$  for  $b_{x,y}^+ \in B^+$  and  $b_{x,y}^- \in B^-$ . Consequently, inferring  $P(R)$  can be achieved by computing the addition of two terms on the right side of Eq. 6.

For computing the term  $P(R|B^-, I^+)$ , it can be obtained via the segmentation results directly. We introduce a subnetwork in RBNet to compute the term, whose overall training loss can be defined as:

$$L_1(\theta_1) = \sum_{x,y} l(r_{x,y}^*, r_{x,y}(\theta_1)) \quad (7)$$

where  $r_{x,y}^*$  represents the ground-truth at the location  $(x, y)$  for road detection and  $r_{x,y}(\theta)$  denotes the output of the network parameterized by  $\theta$  at the same location.

The second term on the right side of the Eq. 6 can be viewed as a residual of the  $P(R|B^-, I^+)$  when computing the  $P(R)$ . We name it as the contextual residual. The form of this contextual residual suggests that it can be computed based on the road boundary detection results,  $P(B^+|I^+)$ . Using the road labels as well, the training loss of the sub-network that predicts the contextual residual can be defined as:

$$L_2(\theta_2) = \sum_{x,y} l(r_{x,y}^*, r_{x,y}(\theta_2, B^+(\theta_3)) + r_{x,y}(\theta_1)) \quad (8)$$

where the  $B^+(\theta_3)$  represents the  $P(B^+|I^+)$  estimated in the sub-network for road boundary detection task parameterised by  $\theta_2$ . The symbol  $\theta_3$  denotes the parameters of the sub-network for computing contextual residual.

**Road Boundary Detection.** Based on Eq. 4,  $P(B)$  can be directly inferred from  $P(B|I^+)$ . Suppose the ground-truths of road boundaries is given by  $\{b_{x,y}^*\}$ , the loss to train the network:

$$L_3(\theta_3) = \sum_{x,y} l(b_{x,y}^*, b_{x,y}(\theta_3)) \tag{9}$$

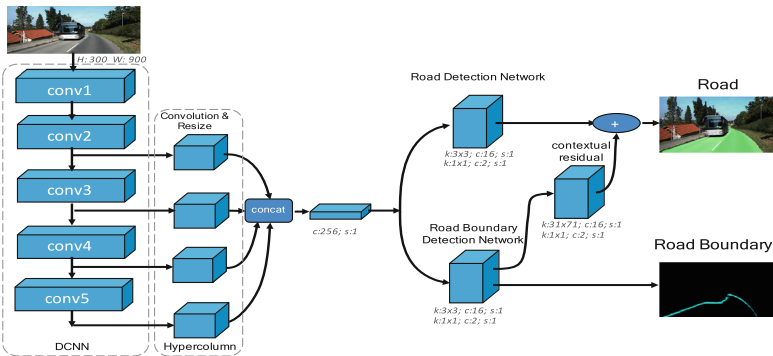
where  $b_{x,y}(\theta_3)$  represents the estimation of the road boundaries for a pixel located at  $(x, y)$  using the sub-network parameterised by  $\theta_3$ . We manually labeled the road boundaries as supervision information based on the ground-truths for road area detection task.

**Multi-task Learning.** Overall, we formulate the training procedure of the RBNet as a multi-task learning problem. Let  $\Theta$  denote all the parameters of RBNet and thus  $\Theta = \{\theta_1, \theta_2, \theta_3\}$ . The general training loss of RBNet is defined as:

$$Loss(\Theta) = \mu_1 L_1(\theta_1) + \mu_2 L_2(\theta_2) + \mu_3 L_3(\theta_3) \tag{10}$$

where  $\mu_i$  represents the loss weight for the corresponding task. As a result, training RBNet can be viewed as minimising the overall loss function with respect to the  $\Theta$ . Furthermore, in order to train the RBNet in an end-to-end manner, we make the CNN feature sharable for each task, which means that the subset of  $\Theta$  for computing the visual features are shared among  $\theta_1, \theta_2$  and  $\theta_3$ . Sharing features could also bring other advantages. For instance, both abstract semantics and fine spatial details could be maintained to ensure good performance.

**Implementation Details.** Figure 2 shows the detailed architecture of the proposed RBNet. As illustrated, the process of RBNet involves several steps. In the first step, we use a pre-trained deep convolutional neural network (DCNN) model to extract visual features, which usually have five convolution blocks. Afterward, we adopt hypercolumn-like architecture, whose details can be found at [19], to fuse and interpret features



**Fig. 2.** Detailed architecture of RBNet. The blue cubes represent convolution layers. The symbols  $k$ ,  $c$ , and  $s$  below the cubes respectively represent the kernel sizes, the channel numbers, and the strides for the corresponding convolution operations.

extracted from different depths. Following this are three task-specific networks, including road boundary detection network, semantic segmentation network, and the contextual residual network.

Specifically, considering the powerful expression power and efficiency, we adopt the use of ResNet50 [20] network as the pre-trained DCNN model. The features from *conv2*, *conv3*, *conv4* and *conv5* blocks are connected to the hypercolumn, followed by two fully convolution layers. Subsequently, three task-specific sub-networks are employed to fulfill the goal of multi-task learning described by Eq. 10. In the boundary detection network and road detection network, we use convolution layers with small kernels to tackle the corresponding tasks. To capture structural contexts in a broader region, we use larger kernels for the convolution layer in predicting the contextual residual. In our implementation, the loss function  $l$  is defined as multinomial logistic loss function.

## 4 Experiment

### 4.1 Setup

In this section, we comprehensively evaluate the effectiveness of the RBNet and also compare it with other competing algorithms. Since the detection of road boundary benefits road detection based on the summarised Bayesian network, we major evaluate the performance on road detection benchmark. To best unfold and assess the performance of the proposed approach, we conduct the evaluation on the KITTI road benchmark [10], where the results of evaluated methods can be made publicly accessible on the official website<sup>2</sup>. KITTI road detection benchmark divides the images into three sets, which are urban marked (UM), urban multiple marked lanes (UMM) and urban unmarked (UU). There are in total 289 images for training and 290 images for testing.

**Evaluation Metrics.** We follow the evaluation metrics as discussed in [21] in KITTI road benchmark. The metrics include maximum F1-measure (MaxF), average precision (AP), precision rate (PRE), recall rate (REC), false positive rate (FPR), and false negative rate (FNR). The four latter measures are computed at the working point of MaxF. According to KITTI’s evaluation system, all the results are transformed into birds-eye view space for evaluation.

**Training.** While training RBNet, we randomly flip and crop the training images and add small disturbance to RGB channels of input data. The input images are resized into a uniform size of  $300 \times 900$ . While training the RBNet, the loss weights  $\mu_1$ ,  $\mu_2$ ,  $\mu_3$  in Eq. 10 are set as 1, 1, and 0.1 respectively. We use 100k training epochs and the learning rate is decayed from 0.01 using “poly” policy. The hardware used for all the computation is a cluster node (8 cores @ 3.50 GHz, 32,GB RAM) accelerated with a GPU card (NVIDIA Tesla K20c 5 GB). The overall processing time of RBNet is 0.18 second per frame.

<sup>2</sup> [http://www.cvlibs.net/datasets/kitti/eval\\_road.php](http://www.cvlibs.net/datasets/kitti/eval_road.php).

**Boundary Refinement.** To demonstrate the effects of detecting road boundary, we use the detected boundaries to refine the detected road areas by eliminating potential false positives. Specifically, we first find the left and right boundaries of the road based on the boundary detection results and then refine the confidence score of each pixel according to their relative locations to the identified boundaries. Outside pixels could be viewed as false results. If better road detection results are obtained, the contextual relationship between road and road boundaries can be proved.

## 4.2 Results

In this section, we thoroughly compare the performance of the RBNet for general road detection with other state-of-the-art methods on the KITTI road benchmark. The compared algorithms include Up\_Conv [4], DDN [5], FTP [12], FCN\_LC [22], SPRAY [23], and StixelNet [6]. MaxF and AP are mainly used for comparison.

Table 1 shows the results of the evaluation on different categories of tasks in the KITTI benchmark. The MaxF and AP scores of UM road, UMM road, and UU road in KITTI benchmark are presented in the table. The effectiveness of the proposed method has been demonstrated since our method achieves the highest scores on both MaxF for each category.

**Table 1.** Performance for the per-category result. “UM”, “UMM” and “UU” represent the detection task for urban marked road, urban multiple marked lane, and urban unmarked road respectively. The “Lane” represents the ego-lane detection task based on the “UM”. Bold fonts refer to the best performance.

Methods	UM		UMM		UU		Lane	
	MaxF	AP	MaxF	AP	MaxF	AP	MaxF	AP
StixelNet [6]	85.33 %	72.14 %	92.98 %	92.89 %	86.06 %	72.05 %	-	-
SPRAY [23]	88.14%	91.24%	89.69%	93.84%	82.71%	87.19%	83.42%	86.84%
FCN_LC [22]	89.36 %	78.80 %	93.26 %	87.15 %	86.27%	75.37 %	-	-
FTP [12]	91.20%	90.60 %	91.85 %	87.98 %	89.62%	88.93%	-	-
DDN [5]	93.65 %	88.55 %	94.17 %	92.70 %	91.76 %	86.26 %	-	-
Up_Conv [4]	92.20 %	88.85 %	95.52 %	92.86 %	92.65 %	<b>89.20 %</b>	89.88%	<b>87.52%</b>
<b>RBNet</b>	<b>94.77 %</b>	<b>91.23 %</b>	<b>96.06 %</b>	<b>93.49%</b>	<b>93.21 %</b>	89.18%	<b>90.54%</b>	82.03%

**Table 2.** Overall performance for KITTI’s benchmark based on all the “UM”, “UMM” and “UU” test sets. Best scores are presented in bold.

Methods	MaxF	AP	PRE	REC	FPR	FNR
StixelNet [6]	89.12 %	81.23 %	85.80 %	92.71 %	8.45 %	7.29%
SPRAY [23]	87.09 %	91.12 %	87.10 %	87.08 %	7.10 %	12.92 %
FCN_LC [22]	90.79 %	85.83 %	90.87 %	90.72 %	5.02 %	9.28 %
FTP [12]	91.61 %	90.96 %	91.04 %	92.20 %	5.00 %	7.80 %
DDN [5]	93.43 %	89.67 %	<b>95.09 %</b>	91.82 %	<b>2.61%</b>	8.18 %
Up_Conv [4]	93.83 %	90.47 %	94.00 %	93.67 %	3.29 %	6.33 %
<b>RBNet</b>	<b>94.97 %</b>	<b>91.49 %</b>	94.94 %	<b>95.01 %</b>	2.79 %	<b>4.99 %</b>



By combining the results of all the “UM”, “UMM” and “UU” road, overall performance of the evaluated algorithms are illustrated in the Table 2. In this measurement, the proposed RBNet has also outperformed other algorithms in many criteria, including MaxF, AP and so on, which proves both the correctness of the summarised Bayesian network and the robustness of the RBNet in general road detection. Some qualitative results on KITTI benchmark are illustrated in the Fig. 3.



**Fig. 3.** Qualitative results on KITTI road detection benchmark. The results are from: (a) UM; (b) UMM; (c) UU; and (d) Lane. The detected road boundaries and the segmented road areas are shown in yellow color and green color respectively. (Color figure online)

## 5 Conclusion

In this work, we formulate the road detection and road boundary detection problem into a unified Bayesian network model based on the contextual relationship between road boundaries and road areas in an image. We then propose the RBNet to estimate the probabilities of the Bayesian network. The RBNet can detect road boundaries and road areas in a single processing step. The empirical study on KITTI benchmark proves the effectiveness and validity of RBNet. For the future research, we will accelerate the processing speed to meet real-time demand.

## References

1. Zhang, W., Yang, J., Jia, W., Kasabov, N., Jia, Z., Zhou, L.: Unsupervised segmentation using cluster ensembles. In: Loo, C.K., Yap, K.S., Wong, K.W., Beng Jin, A.T., Huang, K. (eds.) *ICONIP 2014*. LNCS, vol. 8836, pp. 76–84. Springer, Cham (2014). doi:[10.1007/978-3-319-12643-2\\_10](https://doi.org/10.1007/978-3-319-12643-2_10)
2. Xiao, L., Dai, B., Liu, D., Zhao, D., Wu, T.: Monocular road detection using structured random forest. *Int. J. Adv. Robot. Syst.* **13**(3), 101 (2016)

3. Zhou, S., Gong, J., Xiong, G., Chen, H., Iagnemma, K.: Road detection using support vector machine based on online learning and evaluation. In: *Intelligent Vehicles Symposium (IV)*, pp. 256–261 (2010)
4. Oliveira, G.L., Burgard, W., Brox, T.: Efficient deep methods for monocular road segmentation. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE Press, New York (2016)
5. Mohan, R.: Deep Deconvolutional Networks for Scene Parsing. arXiv preprint [arXiv:1411.4101](https://arxiv.org/abs/1411.4101) (2014)
6. Levi, D., Garnett, N., Fetaya, E., Herzlyia, I.: StixelNet: a deep convolutional network for obstacle detection and road segmentation. In: *British Machine Vision Conference*, pp. 109:1–109:12. BMVC Press (2015)
7. Yuan, Y., Jiang, Z., Wang, Q.: Video-based road detection via online structural learning. *Neurocomputing* **168**, 336–347 (2015)
8. Krähenbühl, P., Koltun, V.: Efficient inference in fully connected crfs with gaussian edge potentials. In: *Advances in Neural Information Processing Systems*, pp. 109–117 (2011)
9. Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Trans. Pattern Anal. Mach. Intell.* **PP**, 1 (2017)
10. Geiger, A., Lenz, P., Urtasun, R.: Are we ready for autonomous driving? The kitti vision benchmark suite. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3354–3361. IEEE Press, New York (2012)
11. Aly, M.: Real time detection of lane markers in urban streets. In: *IEEE Intelligent Vehicles Symposium*, pp. 7–12. IEEE Press, New York (2008)
12. Laddha, A., Kocamaz, M.K., Navarro-Serment, L.E., Hebert, M.: Map-supervised road detection. In: *IEEE Intelligent Vehicles Symposium*, pp. 118–123. IEEE Press, New York (2016)
13. Alvarez, J.M., Salzmann, M., Barnes, N.: Learning appearance models for road detection. In: *IEEE Intelligent Vehicles Symposium*, pp. 423–429. IEEE Press, New York (2013)
14. Brust, C.A., Sickert, S., Simon, M., Rodner, E., Denzler, J.: Convolutional Patch Networks with Spatial Prior for Road Detection and Urban Scene Understanding. arXiv preprint [arXiv:1502.06344](https://arxiv.org/abs/1502.06344) (2015)
15. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431–3440. IEEE Press, New York (2015)
16. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) *MICCAI 2015*. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). doi:[10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)
17. Mostajabi, M., Yadollahpour, P., Shakhnarovich, G.: Feedforward semantic segmentation with zoom-out features. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3376–3385. IEEE Press, New York (2015)
18. Ghiasi, G., Fowlkes, C.C.: Laplacian pyramid reconstruction and refinement for semantic segmentation. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) *ECCV 2016*. LNCS, vol. 9907, pp. 519–534. Springer, Cham (2016). doi:[10.1007/978-3-319-46487-9\\_32](https://doi.org/10.1007/978-3-319-46487-9_32)
19. Hariharan, B., Arbeláez, P., Girshick, R., Malik, J.: Hypercolumns for object segmentation and fine-grained localization. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 447–456. IEEE Press, New York (2015)

20. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778. IEEE Press, New York (2016)
21. Fritsch, J., Kuhn, T., Geiger, A.: A new performance measure and evaluation benchmark for road detection algorithms. In: 16th International IEEE Conference on Intelligent Transportation Systems-(ITSC), pp. 1693–1700. IEEE Press, New York (2013)
22. Mendes, C.C.T., Frémont, V., Wolf, D.F.: Exploiting fully convolutional neural networks for fast road detection. In: IEEE International Conference on Robotics and Automation, pp. 3174–3179. IEEE Press, New York (2016)
23. Kühnl, T., Kummert, F., Fritsch, J.: Spatial ray features for real-time ego-lane extraction. In: 15th International IEEE Conference on Intelligent Transportation Systems, pp. 288–293. IEEE Press, New York (2012)
24. Liu, T., Tao, D.: Classification with noisy labels by importance reweighting. IEEE Trans. Pattern Anal. Mach. Intell. **38**, 447–461 (2012). IEEE Press