Improving Road Detection Results Based on Ensemble Learning and Key Samples Focusing

Siqi Fan, Fenghua Zhu, Hui Zhang, Yisheng Lv, Xiao Wang, Gang Xiong, Fei-Yue Wang

Abstract— **Road detection is fundamental for many applications, especially vision-based autonomous driving systems. To improve the accuracy of the detection results, most of previous research focus on designing feature encoders and classifiers. In this paper, a road detection method is proposed based on ensemble learning and key samples focusing. A road detection network is designed, which integrates classification results based on different feature combinations by weighted voting. The outputs of the network are further processed by morphological transformation. To focus on key samples, a novel loss function is proposed. The loss function can attach importance to hard samples and pay different attention to missed detection and false detection. The method is evaluated on KITTI dataset, and its effectiveness is verified.**

I. INTRODUCTION

Road detection is the basis of trajectory planning for the intelligent vehicles [1] [2]. Various sensors have been installed on intelligent vehicles to perceive the environment. However, visual sensors are still the most popular sensors as they can provide rich information at a low cost. Therefore, road detection using computer vision technologies is usually a fundamental task in developing autonomous driving systems. Road detection can be regarded as a segmentation problem, in which the goal is to classify pixels in an image based on monocular visual information. The methods for road detection can be mainly divided into two categories, which implement the segmentation by traditional models and deep neural networks, respectively. In traditional models, the pixels are classified by traditional classifiers [3] [4] using a variety of features extracted manually, while deep neural networks implement end-to-end segmentation of the road regions [5] [6].

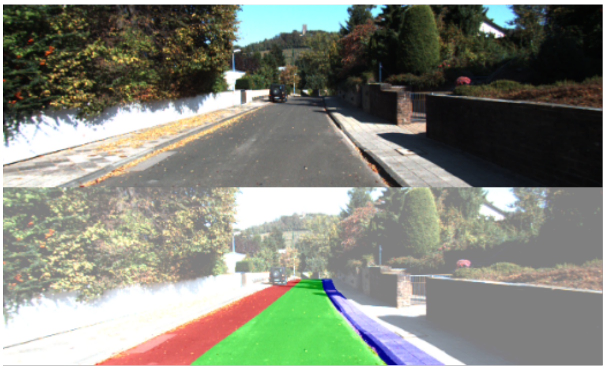

Fig. 1. Schematic diagram of road detection results: The correct detection is colored in green, while the red and blue area are corresponding to the missed and false detection, respectively.

Because of the complicated road scenes, the hypothesis space that need to be searched is huge, however, the number of data available for training is limited. Therefore, there always be some error detections. They can be mainly divided into two types, as shown in Fig. 1: missed detection (False Negative, red area) and false detection (False Positive, blue area). The goal of the optimization is to reduce missed and false detection, thereby making the road detection results better. Most of the research focus on extracting better features and designing complicated classifiers, but that usually causes overfitting.

The ensemble learning method integrates multiple weak classifiers to form a strong one. There are two common types of ensembles: bootstrap aggregating (Bagging) and boosting. Bagging trains each model using a subset of the training set. Similar to that, attribute bagging trains models using subsets in feature space. Different from bagging, boosting trains each new model to emphasize the training samples that the previous models misclassified. Methods based on ensemble learning can improve classification results using weak classifiers. As the existing approaches can basically achieve road detection, the idea of ensemble learning can be used to obtain better road detection results. The road area corresponds to a variety of features. This paper integrates the classification results obtained under different feature combinations, which can reduce the risk and achieve error cancellation. At the same time, hypothetical space is more open compared to a single classifier. The method is more

* This work was supported in part by the National Key R & Development Program of China under Grant 2018YFB1700202, NSFC U1811463, U1909204, 61773381, 61533019, Basic and applied basic research fund of Guangdong Province 2019B1515120030. (Corresponding author: Yisheng Lv.)

Siqi Fan and Hui Zhang are with the State Key Laboratory for Management and Control of Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China, and also with the School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China. (fansiqi2019@ia.ac.cn; zhanghui2015@ia.ac.cn)

Fenghua Zhu and Gang Xiong are with the State Key Laboratory for Management and Control of Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China, and also with the Clouding Computing Center, Chinese Academy of Sciences, Dongguan 523808, Guangdong, China. (fenghua.zhu@ia.ac.cn; Gang.Xiong@ia.ac.cn)

Yisheng Lv, Xiao Wang and Fei-Yue Wang are with the State Key Laboratory for Management and Control of Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China. (yisheng.lv@ia.ac.cn; x.wang@ia.ac.cn; feiyue.wang@ia.ac.cn)

robust to the diverse road scenes.

Another obstacle to the improvement of detection results is inadequate attention to key samples. Generally, hard samples are key samples. Hard samples are those that are easily misclassified. They should be paid more attention when training models. In addition, key samples in diverse usage scenarios are different for various needs. The users may have different tolerances for the two types of error detection. Therefore, different mount of attention should be given to missed and false detection during training. This requires us to solve the problem of the imbalance between positive and negative samples.

In this paper, a road detection method is proposed. The main contributions of this paper are as follows:

- A road detection network is designed and implemented, which integrates the classification results by weighted voting. The base classifiers are based on different feature combinations.
- A novel loss function based on key samples focusing is designed. The loss function introduces the focusing factor to achieve the mining of hard samples. At the same time, the loss function can be adjusted by changing the attention weighting coefficient to pay different attention to missed detection and false detection.

The experimental results on KITTI [7] dataset demonstrate that the road detection results can be improved using the proposed method.

II. RELATED WORK

Road detection has a certain degree of research foundation. Ensemble learning also has a lot of research results. Some work has been explored in hard samples mining.

A. Road Detection Methods

Traditional road detection methods use pixel-level features and super-pixel level features, and usually use color space features. Using histogram peaks and temporal filter responses, texture features from varied color space are described, then the lane areas are generated within flat regions [8]. The road is described as a linear combination of different color space in [9], and use the color distribution of each pixel to decide whether it belongs to the road. An AdaBoost classifier with super-pixel color features is trained by Li et al. [10] to enhance the road detection.

With the development of deep learning techniques, the segmentation methods based on convolutional neural networks (CNNs) has got impressive results[11][12][13]. Fully convolution neural network (FCN) [11] add upsampling layers to standard CNNs and introduce skip connections between the downsampling and upsampling paths. U-net [13] modify and extend FCN's architecture so that it can works with very few training images and get more precise detection results. On the basis of FCN and U-net, [6] is further improved and gets better results in terms of speed and accuracy of road detection. The encoder-decoder structure is used by many approaches [14] [15] [6] to enhance road detection. It is verified that such architectures are helpful for obtaining spatial details.

There are some approaches use the relationship between road and road boundary. To take advantage of road boundary information, [16] [17] adopt a sequential processing strategy. RBNet [18] formulate the relationship among road, road boundary and the input image in the same probabilistic graph using a unified Bayesian network model. [19] introduces an inter-link encoder to stream complementary information between the two decoders with the two different tasks.

In this paper, we use the encoder-decoder structure as the basis. We obtain two decoders that focus on different feature combinations using the information of the road and boundary. Different from others, the detection network relies on ensemble learning that contains four base classifiers, and the loss function makes the network focus on the key samples.

B. Ensemble Learning

The ensemble learning [20] methods use multiple learning algorithms to obtain better predictive performance. Multiple weak classifiers can achieve the effect of strong classifiers through integration. The two common types of ensembles are bagging [21] and boosting [22]. The weak classifiers in bagging are parallel and independent of each other. Each model are trained using a subset of the training data. Attribute bagging [23] , also called random subspace method, trains models using subsets in feature space. When this method is applied to decision tree, the resulting model is called a random forest [24]. Unlike bagging, weak classifiers in boosting are related. Boosting trains the new model to focus more on those samples that are misclassified by previous models. Adaboost [25] is the most common implementation, which pay more attention to the misclassified samples and obtain the result by weighted-voting. Boosting pays attention to hard samples that are easily misclassified during training, while bagging can reduce the overfitting of the models.

In this paper, we integrate the results of base classifiers based on different feature combinations by weighted voting to obtain better road detection results.

C. Hard Sample Mining

Insufficient mining of hard samples has always been one of the important factors affecting the results of classification problems. The importance of hard sample mining is self-evident. Hard samples that are easy to be misclassified should be paid more attention during training. A common solution is to sample the hard samples during training [26] [27] [28]. These approaches are effective but complex. Focal loss introduces a focusing factor $(1-p)^{\gamma}$ to the cross entropy loss. Its effect on the results is obvious. However, the selection of γ is critical.

In this paper, a loss function is proposed. We introduce a focus factor that does not require numerical selection.

III. METHOD

In order to improve road detection results on the basis of the existing weak classifiers using the small training set,

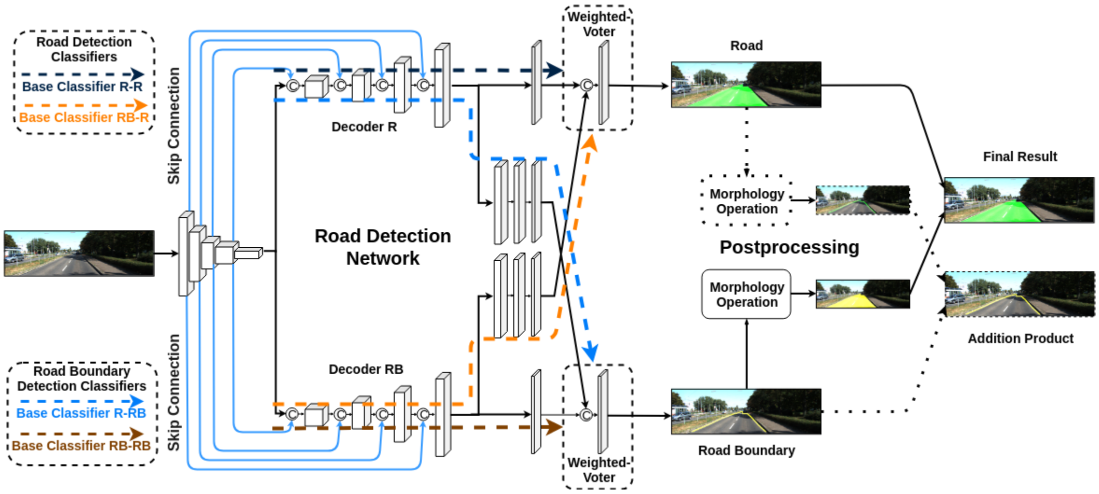

Fig. 2. Road Detection Framework: Road Detection Network + Postprocessing. The road detection network is designed based on ensemble learning, containing four base classifiers. The loss function is designed based on key samples focusing. The postprocessing usually can slightly improve the performance by regenerating road detection from the two output results of the network.

the proposed method is based on ensemble learning and key samples focusing. The framework is shown in Figure 2.

The road detection network is designed based on ensemble learning. It uses two base classifiers based on different features to identified whether the pixels belong to the road areas. The road detection results of the network are obtained by integrating the results of the base classifier R-R (Road-Road) and RB-R (Road Boundary-Road). At the same time, the results of the base classifier R-RB (Road-Road Boundary) and RB-RB (Road Boundary-Road Boundary) are integrated to obtain the detection results of road boundary, which is used in postprocessing.

The loss function is designed based on key samples focusing. We introduce a focusing factor to attach importance to hard samples. After solving the imbalance between positive and negative samples, an attention weighting coefficient is introduced to pay different attention to missed detection and false detection.

A. *Road Detection Network Based on Ensemble Learning*

The road detection network, like most CNN-based segmentation networks, includes encoder and decoder. To decode two different feature combinations, the network contains two decoders, which share one encoder. The shared encoder is based on the VGG structure. It extracts the features for both decoders from the input image.

Based on the theory of ensemble learning, we hope that there is a significant difference between the feature combinations decoded by the two decoders. They should be also related to the road. Therefore, we choose road detection and road boundary detection as targets, and train two decoders. Different training targets make the decoder focus on different features, and then we can obtain different feature combinations. For example, with road detection as the target, the resulting decoder puts much emphasis on the fine-grained color features. In contrast, the resulting decoder focuses more on edge-like structural features, if the target is to detect the road boundary. Due to the large differences in features, the two base classifiers have a certain degree of differences. So integrating the two classification results can get better comprehensive results.

As for the base classifier R-R and RB-RB, since the feature combinations obtained by the decoder are highly relevant to the training targets, there is only a $3 \times 3 \times 2$ convolutional layer behind decoder R and RB. However, for base classifier R-RB and RB-R, the feature combinations they are based on have relatively little correlation with the training targets, so the decoding results are sequentially subjected to three convolutional layers, respectively $3 \times 3 \times 32$, $3 \times 3 \times 16$ and $3 \times 3 \times 2$. In order to classify pixels into two categories, each base classifier uses a sigmoid classification layer at the end.

Then, concatenate the results of the two base classifiers with the same target, and achieve weighted voting by using a convolutional layer whose activation function is sigmoid. Considering the relevance of the neighborhood pixels in the image, the classification results within the 5×5 neighborhood around the pixel are taken into consideration while voting. Therefore, the probability that the pixel (x, y) belongs to the road is as follows

$$P((x, y) \in r) = \sum_{c=1}^{2} \sum_{i,j=-2}^{2} \omega_{c_{ij}} \times P_c((x+i, y+j) \in r)$$

where $\omega_{c_{ij}}$ is the weight of classifier c at (x+i, y+j), which is obtained through training. $P_c((x+i, y+j) \in r)$ is the probability that the pixel (x+i, y+j) belongs to the road, which is predicted by classifier c.

B. *Loss Function Based on Key Samples Focusing*

In general, the road detection results can be divided into four parts, as shown in Fig. 1: correct road detection, correct non-road detection, missed detection and false detection. The correct road detection and missed detection make up the actual road area, while the other two make up the actual non-road area. In order to pay different attention to missed

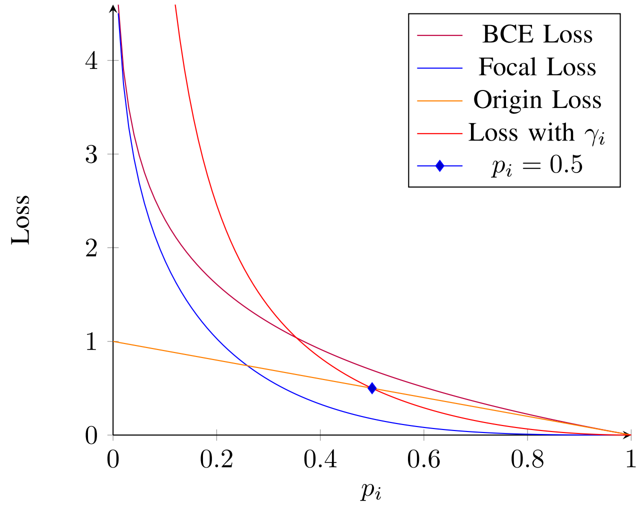

Fig. 3. Diagram of loss function with γ_i for positive samples: when $p_i > 0.5$, γ_i tends to reduces the effect of easy samples, and it tend to improve the impact of hard samples while $p_i < 0.5$.

detection and false detection, the loss function is mainly divided into two parts: the first part is for missed detection (FN), and the second is for false detection (FP).

$$L_{FN} = \sum g_i - \sum g_i p_i = \sum g_i(1 - p_i)$$

$$L_{FP} = \sum(1 - g_i) - \sum(1 - g_i)(1 - p_i) = \sum(1 - g_i)p_i$$

In the above $g_i \in \pm 1$ specifies the ground-truth class and $p_i \in [0, 1]$ is the models estimated probability for the class with label $g_i = 1$. Assume that $g_i = 1$ for pixels in road area, and $g_i = 0$ for pixels in non-road area.

To increase the attention on hard examples, a focusing factor γ_i is introduced. Take positive samples as an example,

$$\gamma_i = tan(\frac{\pi}{2} \times (1 - p_i))$$

For positive samples, when $p_i > 0.5$, $\gamma_i < 1$. γ_i reduces the effect of easy samples on the value of the loss function. When $p_i < 0.5$, $\gamma_i > 1$. γ_i improves the impact of hard samples, and the more likely the sample is misclassified, the greater the γ_i is. When $p_i = 0.5$, $\gamma_i = 1$, which will not change the original value. Comparison between the loss with γ and the original one is shown in Figure 3. It also shows the comparison between focal loss and the bce loss at the same time. Generally, the robustness of the unbounded loss function is insufficient due to the noise data. Therefore, in order to avoid the situation of $\gamma_i \rightarrow \inf$ when $p_i \rightarrow 0$, we replace $\frac{\pi}{2}$ with a number slightly smaller than it, such as 1.5. After introducing γ_i, there is

$$L_{FN} = \sum g_i(1 - p_i) \times tan(1.5 \times (1 - p_i))$$

$$L_{FP} = \sum(1 - g_i)p_i \times tan(1.5 \times p_i)$$

The imbalance between positive and negative samples often occurs in classification problems, which causes the value of the loss function is more affected by the large proportion of samples. Since the value of the loss function depends on the misclassified samples, that is, it is only affected by missed and false detection. The false detection of the positive samples is the missed detection of the negative one. Therefore, the imbalance problem can be solved to a certain extent by calculating the missed detection rates of the positive and negative samples, respectively.

Finally, an attention weighting coefficient α is introduced to adjust the attention of the two parts.

$$\begin{aligned} L =& \alpha \times \frac{L_{FN}}{\sum g_i} + (1 - \alpha) \times \frac{L_{FP}}{\sum 1 - g_i} \\ =& \alpha \times \frac{\sum g_i(1 - p_i) \times tan(1.5 \times (1 - p_i))}{\sum g_i} \\ &+ (1 - \alpha) \times \frac{\sum(1 - g_i)p_i \times tan(1.5 \times p_i)}{\sum 1 - g_i} \end{aligned}$$

The loss function is used for both road detection and road boundary detection. The only difference is the value of α.

C. *Postprocessing*

The two output results of the network can be further processed by morphological transformation. The relationship between road and its boundary make it possible to generate one with another. The detected road boundary is supposed to be the edge of the road area. So the problem is simplified to fill the area based on the edge, which can be achieved by morphological transformation. The union of the generated road area and the road detection is regarded as the final detection result. Simultaneously, the results of road boundary detection can also be output as the addition products.

IV. EXPERIMENT

In this section, we evaluate the proposed method on KITTI dataset and test on KITTI road benchmark [29].

A. *Setup*

1) Dataset: The KITTI dataset contains 289 labeled images and 290 testing images, including three subsets of road scenes: urban unmarked road (UU), urban multiple marked road (UMM) and urban marked road (UM). For the KITTI benchmark does not provide groundtruth labels for the testing set, we implemented our loss experiments and framework experiments on the labeled images.

2) Training Detail: While training, we randomly do data augmentation like horizontal flips to the training images. The input images are resized into a uniform size of 256×512. The shared encoder network is initialized by ImageNet [30] with VGG structure [31].

First, we start with training base classifier R-R and RB-RB respectively, to ensure that the decoder R and RB can decode different feature combinations as we expected. Then, train the base classifier R-RB and RB-R based on the other feature combinations with the weights of the previous parts fixed. Finally, train the two weighted-voters and get the final trained network.

The overall network utilizes a batch normalization with the batch size of 2. We use the Adam optimizer [32] and train the weighted-voters with a learning rate of 10^{-4}. While training other parts, the learning rate is set as 10^{-3}

3) Evaluation Metrics: The metrics for offline evaluation include the mean of each of these values : F1-measure, precision rate (PRE), recall rate (REC), false positive rate (FPR) and false negative rate (FNR). We use the result images directly. Our method is tested on the KIITI road benchmark. It turns the images into bird-eye view space.

B. Loss Study

There are two parts of the loss function worth studying: the focusing factor γ for hard samples and the attention weighting coefficient α.

We start with γ and verify its validity. In the case of $\alpha = 0.5$, the experiment is performed with base classifier R-R as an example, which is the most basic road detection network. The results are shown in Table I. It can be seen that the introduction of γ can effectively improve the road detection results.

TABLE I

LOSS STUDY OF FOCUSING FACTOR γ (F1-MEASURE)

	UU	UMM	UM
Original Loss	84.87%	90.85%	87.88%
Loss with γ	**86.21%**	**91.36%**	**88.29%**

In order to explore the effect of the attention weighting coefficient α in the loss function, we adjust the value of α for experiments. The results are shown in Table II.

It can be seen that the direct influence of α on the final detection results is the change in FNR and FPR. As α increases, the FNR tends to decrease, indicating that the loss function's attention to the missed detection part has increased, which is in line with expectations. At the same time, it can be noticed that with the increased attention to the missed detection part, the false detection part has received less attention, so FPR has increased. The results are ploted in Figure 4 for the convenience of observation. Therefore, when selecting α, there is a trade-off between missed detection and false detection. If the detection result is evaluated by F1-measure, $\alpha = 0.6$ is a better choice. So in the subsequent experiments, we set $\alpha = 0.6$ under the road detection training target.

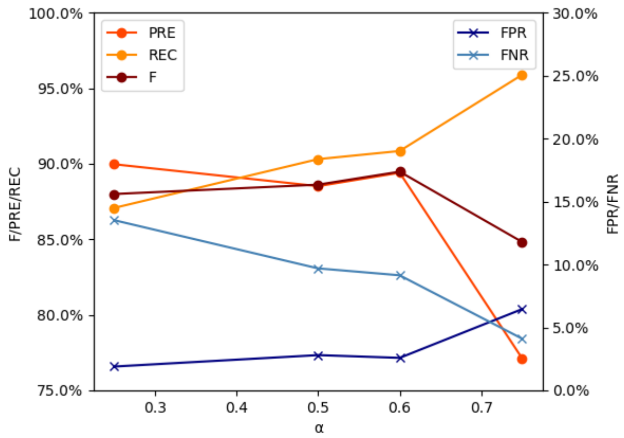

Fig. 4. Loss Study of weighting coefficient α: The increase of α directly leads to the decrease of FNR and the increase of FPR.

C. Framework Study

This section will explore the effects of road detection network and postprocessing separately.

As for the effect of the ensemble in road detection network, we compare the road detection results before and after integrations. In the experiment, a better value of 0.1 was selected as the value of α under the road boundary detection training target. The results are shown in Table III.

Integrating the base classifier R-R and RB-R can effectively improve the detection results of the network. The F-measure value is increased by 0.14%, 0.21% and 0.56% in the three road scenarios. Observing the PRE and REC values in the table, it can be found that there is an effective improvement of the detection accuracy (PRE) because of the integrations. The PRE in the three road scenarios is increased by 1.18%, 1.16% and 2.01%. However, the value of REC has decreased to some extent.

TABLE II

LOSS STUDY OF ATTENTION WEIGHTING COEFFICIENT α

	UU					UMM					UM				
	F	**PRE**	**REC**	**FPR**	**FNR**	**F**	**PRE**	**REC**	**FPR**	**FNR**	**F**	**PRE**	**REC**	**FPR**	**FNR**
0.25	84.72%	**91.76%**	81.76%	**1.38%**	18.24%	90.95%	**91.63%**	91.16%	**2.71%**	8.84%	88.32%	86.53%	88.32%	**1.57%**	13.47%
0.50	86.21%	87.81%	87.06%	2.21%	12.94%	91.36%	88.97%	94.71%	3.86%	5.29%	88.29%	88.78%	89.15%	2.33%	10.85%
0.60	**87.98%**	89.55%	88.29%	1.92%	11.71%	**91.72%**	89.71%	94.60%	3.53%	5.39%	**88.73%**	**88.96%**	89.67%	2.28%	10.31%
0.75	80.65%	71.47%	**94.90%**	6.64%	**5.10%**	88.85%	82.54%	**97.19%**	6.63%	**2.81%**	85.00%	77.37%	**95.58%**	6.09%	**4.41%**

TABLE III

THE EFFECT OF ENSEMBLE IN ROAD DETECTION NETWORK

	UU			UMM			UM		
	F	**PRE**	**REC**	**F**	**PRE**	**REC**	**F**	**PRE**	**REC**
Base Classifier R-R	87.98%	89.55%	**88.29%**	91.72%	89.71%	**94.60%**	88.73%	88.96%	**89.67%**
Ensemble Result	**88.12%**	**90.73%**	87.38%	**91.93%**	**90.87%**	93.78%	**89.29%**	**90.94%**	88.64%

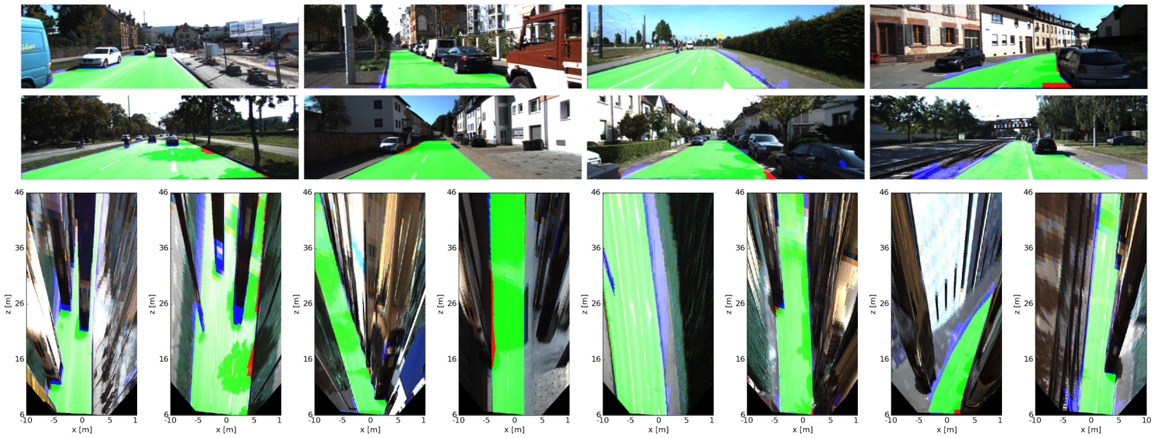

Fig. 5. Road Detection Results on KIITI Road Benchmark: The top is the road detection result, and the bottom is the corresponding results in bird-eye view space. The green area in the figure is the correct detection. The red and blue areas correspond to missed detection and false detection, respectively.

TABLE IV

THE EFFECT OF POSTPROCESSING

	UU	UMM	UM
Result of Network	88.12%	**91.93%**	89.29%
Result of Postprocessing	**88.22%**	91.92%	**89.42%**

Postprocessing based on morphological transformation can also achieve a certain degree of improvement, except that the results in the UMM scene have slightly decreased. The F-measure value in the UU and UM scenes has increased by 0.1% and 0.13%, respectively. We expect to obtain more robust detection results through post-processing. Thus the final result is the union of road detection and the road area generated from road boundary detection. But it also inevitably introduces errors of the boundary detection. These introduced errors sometimes affect the detection performance. The effect is obvious when the original road detection results are good, like UMM.

D. Test Result

We test our method on KITTI road benchmark. Table V. shows the results of the evaluation in the KITTI benchmark. Some demos are shown in Figure 5.

TABLE V

ROAD DETECTION RESULTS ON KITTI SUBSETS(UM/UMM/UU)

	MaxF	AP	PRE	REC	FPR	FNR
UM	92.56%	88.18%	93.86%	91.30%	2.72%	8.70%
UMM	95.71%	90.80%	94.61%	96.83%	6.07%	3.17%
UU	91.11%	82.68%	88.50%	93.88%	3.98%	6.12%

The encoder of this method is based on VGG, and the base classifiers are based on the idea of U-net. Therefore, we compare with the results of Up_Conv [6], whose encoder and classifier are based on the same things as ours. By combining the results of all the subsets (UM, UMM and UU), overall performance of urban road scene (UR) are illustrated in the Table VI. Compared with the baseline approach, our method is better overall. It outperforms Up-Conv with a gain of 1.26% on max-F measure.

TABLE VI

ROAD DETECTION RESULTS ON KIITI ROAD BENCHMARK

	UM	UMM	UU	UR
PLARD[33]	97.05%	97.77%	95.95%	97.03%
RBANet[34]	95.78%	97.38%	94.91%	96.30%
RBNet[18]	94.77%	96.06%	93.21%	94.97%
DDN[35]	93.65%	94.17%	91.76%	93.43%
ALO-AVG-MM[36]	91.15%	94.05%	89.45%	92.03%
Up-Conv[6]	90.48%	93.89%	91.89%	92.39%
Ours	92.56%	95.71%	91.11%	93.65%

It also achieves a competitive performance when compared to the state-of-the-art methods in benchmark. PLARD[33] is the top of the leader-board, which introduces LiDAR information into visual image-based road detection.

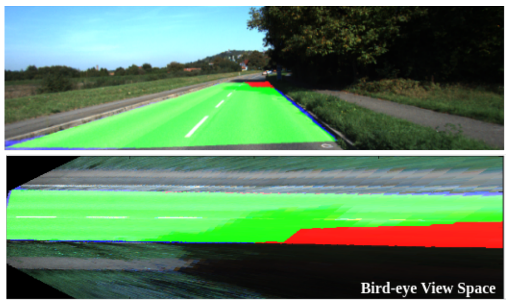

Fig. 6. A Common Error Case: The top is the road detection result, and the bottom is the corresponding result in bird-eye view space.

Figure 6 shows a common error case. The missed detection is caused by the shadow of the tree. Although ensemble learning can enhance the robustness of the detection network, error detection still cannot be avoided when shadows affect both road detection and road boundary detection.

V. CONCLUSION

In this work, we propose a road detection method based on ensemble learning and key samples focusing.

A road detection network based on ensemble learning is designed. The detection network integrates the results of the base classifiers based on different feature combinations by weighted voting. Its benefits for the road detection results have been proved by experiments.

A novel loss function based on key sample focusing is designed. The focusing factor is introduced to attach importance to hard samples. To adapt to different scenarios, the attention weighting coefficient is introduced to pay different attention to missed detection and false detection.

The proposed method is evaluated on KITTI dataset and its effectiveness is verified. The test results show that it also achieves a competitive performance when compared to the state-of-the-art methods in benchmark.

For the future research, the basic classifiers can be replaced with stronger classifiers to obtain better results. Obtaining more base classifiers by selecting feature combinations more randomly is also a direction that can be further studied.

REFERENCES

[1] J. Levinson, J. Askeland, J. Becker, J. Dolson, D. Held, S. Kammel, J. Z. Kolter, D. Langer, O. Pink, V. Pratt, *et al.*, "Towards fully autonomous driving: Systems and algorithms," in *2011 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2011, pp. 163–168.

[2] X. Yang, L. Chen, C. Long, H. Wang, W. Hong, D. Cao, E. Velenis, and F. Y. Wang, "Advances in vision-based lane detection:algorithms, integration, assessment, and perspectives on acp-based parallel vision," *IEEE/CAA Journal of Automatica Sinica*, vol. 005, no. 3, pp. 645–661, 2018.

[3] S. Zhou, J. Gong, G. Xiong, H. Chen, and K. Iagnemma, "Road detection using support vector machine based on online learning and evaluation," in *2010 IEEE Intelligent Vehicles Symposium*. IEEE, 2010, pp. 256–261.

[4] L. Xiao, B. Dai, D. Liu, D. Zhao, and T. Wu, "Monocular road detection using structured random forest," *International Journal of Advanced Robotic Systems*, vol. 13, no. 3, p. 101, 2016.

[5] R. Mohan, "Deep deconvolutional networks for scene parsing," *arXiv preprint arXiv:1411.4101*, 2014.

[6] G. L. Oliveira, W. Burgard, and T. Brox, "Efficient deep models for monocular road segmentation," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2016, pp. 4885–4891.

[7] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.

[8] I. Katramados, S. Crumpler, and T. P. Breckon, "Real-time traversable surface detection by colour space fusion and temporal analysis," in *International Conference on Computer Vision Systems*. Springer, 2009, pp. 265–274.

[9] J. M. Alvarez, M. Salzmann, and N. Barnes, "Learning appearance models for road detection," in *2013 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2013, pp. 423–429.

[10] J. Li, L. Jin, S. Fei, and J. Ma, "Robust urban road image segmentation," in *Proceeding of the 11th World Congress on Intelligent Control and Automation*. IEEE, 2014, pp. 2923–2928.

[11] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[12] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.

[13] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[14] D. Barnes, W. Maddern, and I. Posner, "Find your own way: Weakly-supervised segmentation of path proposals for urban autonomy," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2017, pp. 203–210.

[15] Q. Wang, J. Gao, and Y. Yuan, "Embedding structured contour and location prior in siamesed fully convolutional networks for road detection," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 1, pp. 230–241, 2017.

[16] Y. Alon, A. Ferencz, and A. Shashua, "Off-road path following using region classification and geometric projection constraints," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 1. IEEE, 2006, pp. 689–696.

[17] G. Zhang, N. Zheng, C. Cui, Y. Yan, and Z. Yuan, "An efficient road detection method in noisy urban environment," in *2009 IEEE Intelligent Vehicles Symposium*. IEEE, 2009, pp. 556–561.

[18] Z. Chen and Z. Chen, "Rbnet: A deep neural network for unified road and road boundary detection," in *International Conference on Neural Information Processing*. Springer, 2017, pp. 677–687.

[19] J. Zhang, Y. Xu, B. Ni, and Z. Duan, "Geometric constrained joint lane segmentation and lane boundary detection," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 486–502.

[20] D. Opitz and R. Maclin, "Popular ensemble methods: An empirical study," *Journal of artificial intelligence research*, vol. 11, pp. 169–198, 1999.

[21] L. Breiman, "Bagging predictors," *Machine learning*, vol. 24, no. 2, pp. 123–140, 1996.

[22] Y. Freund, "Boosting a weak learning algorithm by majority," *Information and computation*, vol. 121, no. 2, pp. 256–285, 1995.

[23] R. Bryll, R. Gutierrez-Osuna, and F. Quek, "Attribute bagging: improving accuracy of classifier ensembles by using random feature subsets," *Pattern recognition*, vol. 36, no. 6, pp. 1291–1302, 2003.

[24] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[25] Y. Freund, R. Schapire, and N. Abe, "A short introduction to boosting," *Journal-Japanese Society For Artificial Intelligence*, vol. 14, no. 771-780, p. 1612, 1999.

[26] A. Shrivastava, A. Gupta, and R. Girshick, "Training region-based object detectors with online hard example mining," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 761–769.

[27] P. F. Felzenszwalb, R. B. Girshick, and D. McAllester, "Cascade object detection with deformable part models," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, 2010, pp. 2241–2248.

[28] P. Viola, M. Jones, *et al.*, "Rapid object detection using a boosted cascade of simple features," *CVPR (1)*, vol. 1, no. 511-518, p. 3, 2001.

[29] J. Fritsch, T. Kuehnl, and A. Geiger, "A new performance measure and evaluation benchmark for road detection algorithms," in *International Conference on Intelligent Transportation Systems (ITSC)*, 2013.

[30] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.

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

[32] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[33] Z. Chen, J. Zhang, and D. Tao, "Progressive lidar adaptation for road detection," *IEEE/CAA Journal of Automatica Sinica*, vol. 6, no. 3, pp. 693–702, 2019.

[34] J.-Y. Sun, S.-W. Kim, S.-W. Lee, Y.-W. Kim, and S.-J. Ko, "Reverse and boundary attention network for road segmentation," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2019, pp. 0–0.

[35] R. Mohan, "Deep deconvolutional networks for scene parsing," 2014.

[36] F. A. L. Reis, R. Almeida, E. Kijak, S. Malinowski, S. J. F. Guimaraes, and Z. K. G. do Patrocinio Jr., "Combining convolutional side-outputs for road image segmentation," in *2019 International Joint Conference on Neural Networks (IJCNN) - **Accepted***, July 2019.