International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 03 | Mar 2020 www.irjet.net p-ISSN: 2395-0072

MULTI-LABEL ROAD SCENE PREDICTION FOR AUTONOMOUS VEHICLES USING DEEP NEURAL NETWORKS

KK. Shibna[1], R. Soundararajan[2]

[1]PG Scholar, Dept. of Computer Science and Engineering, SVS College of Engineering, Coimbatore, Tamilnadu, India
[2]Assistant Professor, Dept. of Computer Science and Engineering, SVS College of Engineering, Coimbatore, Tamilnadu, India

---

Abstract – Autonomous driving systems are becoming popularity nowadays. The detection of road conditions is vital for the efficient working of autonomous vehicles. Internet giants and academic institutes to spice up autonomous driving technologies; while progress has been witnessed in environmental perception tasks, like object detection and driver state recognition, the scene centric understanding and identification still remain a virgin land. This mainly encompasses two key issues: the shortage of shared large datasets with comprehensively annotated road scene information and the difficulty to find effective ways to coach networks concerning the bias of category samples, image resolutions, scene dynamics, and capturing conditions. We propose a multi-label neural network for road scene recognition, which includes both single- and multi-class classification modes into a multi-level cost function for training with imbalanced categories and utilizes a deep data integration strategy to enhance the classification ability on hard samples. Along side road conditions the situation of the vehicle and lane departure warning system is additionally included for the efficient autonomous system.

Key Words: Large-scale dataset, data imbalance, multi-label classification, road scene recognition.

1. INTRODUCTION

SELF-DRIVING has gotten immense capital inflow and huge research enthusiasm for both scholarly world and industry as of late. Analysts originating from different worldwide regarded establishments and top-level portable fabricates, join together to push the limit of self-driving difficulties. Self driving innovation can be isolated into a few sections, including restriction and mapping [1]–[3], movement arranging [4], [5] social choice [6], [7], and scene understanding [8], and so on. While the initial three sections have just been broadly looked into, scene understanding has not been all around contemplated or tackled. The essential two purposes behind this incorporate the absence of freely accessible huge scope scene-driven dataset in self driving space and the absence of successful preparing approach on these datasets to manage information awkwardness brought by classification tests and picture goals.

DCNNs are exceptionally subject to the preparing dataset to acquire an elite. Testing datasets regularly assume a significant job in approving the presentation of cutting edge profound models just as in animating new calculations. For instance, the ImageNet LSVRC-2010 dataset, which contains 1.2 million high-goals pictures of 1,000 unique classes, has given an incredible assistance in finding the limit of DCNN models in picture classification[18].

Right now, first present a largescale dataset, called Driving Scene, which is at first intended for enormous scope scene acknowledgment in self-driving space. It contains over 110K pictures with regular traffic scenes gathered by both vehicle run camera and web internet searcher, including diverse climate conditions, street structures, ecological occasions and driving spots, with fine-grained and deliberately commented on names. We ensure that each class has a preparation set of in excess of 400 examples, which is near the test standard of LSVRC. We trust our Driving Scene dataset can assist with learning more extravagant traffic scenes.

Despite the fact that object order and scene characterization have made extraordinary accomplishments in ImageNet and Places datasets, there are still a few issues unsolved. For instance, there are some mistaken names among the 280 million marked pictures of ImageNet, and there exist a ton of equivocal and comparative semantic classifications in the mark corpus of Spots. Every one of these focuses have carried more noteworthy difficulties to the preparation of profound systems. While in our Driving Scene, fine-grained and deliberately commented on marks have been given to maintain a strategic distance from the above issues. Nearly, the fundamental difficulties of the Driving Scene dataset lie in the accompanying focuses:

• Multi-class prediction. This dataset gives various labels for the street scenes. Henceforth, the street scene characterization will be included by perceiving various classifications of items simultaneously.

• Data imbalance. This dataset holds a huge assortment of traffic situations. Be that as it may, the size of every class isn't the equivalent, also, some once in a while showed up scenes have less pictures. Along these lines, how to prepare the profound convolutional systems to guarantee a high exactness on little classifications is a difficult issue.

• Varied image resolution. All pictures right now caught in reality. Not withstanding, the information source incorporates both the Internet and the genuine vehicle

driving. It is hard to keep up a uniform goals for them. Subsequently, it requires the system to have the option to smother the impact of differed picture goals of the data sources.

With respect to above focuses, right now, additionally present another multi-label neural network as a benchmark on the Driving Scene dataset. The engineering misuses cross breed marks which incorporates both multi-and single-labels. The multi-labels are for the most part utilized for multi-classification expectation learning while the single labels are utilized to implement the administered learning of hard examples or little classes which should be more painstakingly dealt with during the preparation strategy. Moreover, we propose a profound information incorporation technique, which employments a boosting strategy to ensure a versatile inspecting of scene pictures, particularly for imbalanced class tests. As picture quality fluctuates as for the pressure approach, we further utilize goals versatile system in our system to improve the power against the commotion from shifted picture goals.

2. RELATED WORK

Here, we present researches related to multi label scene recognition from two aspects: the dataset of self driving scenes and the classification with biased data.

A. Dataset of Self-Driving Scenes

Immense measures of labeled dataset are frequently compared as the fuel to profound learning rocket, without which can the enormous advancement of vision based research more outlandish be accomplished. Different models for huge preparing set-profited scene acknowledgment can likewise be found, e.g., the SUN dataset gives a wide inclusion of scene classifications containing 397 classifications with in excess of 100 pictures for each class. Notwithstanding, there is no such dataset in the self-sufficient driving field. Some traffic-scene datasets for the most part center around ecological discernment, with oneself driving scene acknowledgment nearly disregarded. For instance, KITTI contains a wide scope of difficulties like sound system vision, odometry, object location and following, and so on. CompCar dataset explicitly centers around fine-grained vehicle characterization/check and trait forecast. CityScapes [8] gives an enormous scope dataset got from sound system successions, focusing on both pixel-and occasion level semantic naming. Oneself driving scene acknowledgment is straightforward for human cerebrums however amazingly hard for PCs to address. To pull in and spur more research on self driving scene recognition, 1 we along these lines present another enormous scaled dataset of over 110k scene pictures cutting across 52 classes, which is right now, to our best information, the biggest dataset regarding scene acknowledgment in self-driving space. The most related work to our own is the FM2 dataset, yet it contains an all out number of 6,237 pictures from eight scene classes.

B. Classification with Biased Data

A single image usually associates with multiple scene labels is used in scene recognition,. Thus, most prior works train a deep neural network to assign the multi-class label to the query image. Some deep structures such as VggNet [19], Inception V3 and Res Net have demonstrated higher performance with deeper layers in classification, the training still suffers from the negative impact of data imbalance.

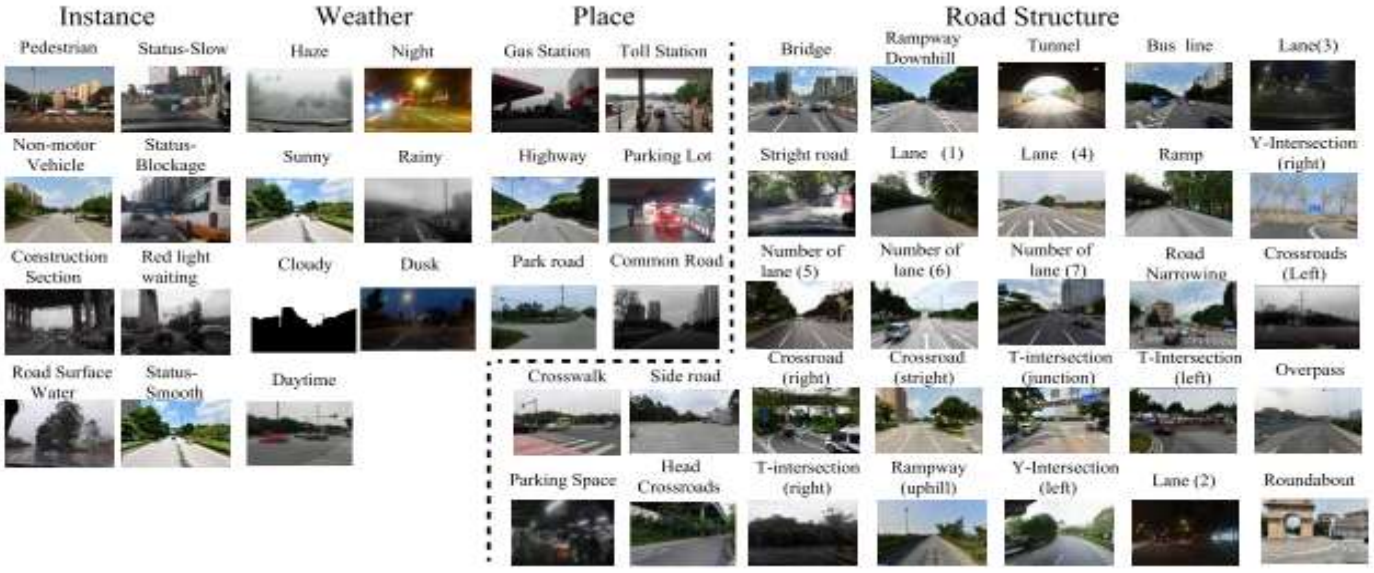


Fig -1: The Driving Scene dataset. Subclasses of each super class are displayed together.

Attempting to solve this problem, two approaches are well studied in past years. The first approach is re-sampling, to rebalance class priors during training through under- and oversampling. The second is the cost-sensitive learning, along with the Triple-Header Hinge Loss to assign different costs for misclassification on the majority and minority classes. Despite a good performance, both methods are proposed mainly for single-label classification. Moreover, the over- and under-sampling may introduce undesired noise or remove valuable sample information while the cost-sensitive learning usually requires utilization of additional features.

3. DEEP DRIVING SCENE DATASET

To highly strengthen the robustness and completeness of our deep Driving Scene dataset, we take the whole geographical location, temporal variation, static and dynamic characteristics into consideration when collecting the dataset. We define each driving scene by combing the world-wide transportation construction rule and human-biased understanding towards all scenes. In sum, we provide 52 different kinds of driving scenes, cutting across common driving instances, weather conditions, places and road structures. The scene category structure is shown in Fig. 1, in which each image is tagged with fine-grained and carefully annotated labels.

1) Driving Instance: Here Driving instance represents the instances currently happening on traffic roads and they are temporally short, including traffic congestion, road construction, red light waiting, pedestrian crossing and road with surface water. In terms of traffic congestion, we classify it into three status according to the congestion level: smooth, slow and fully-blocked. It involves the pedestrian crossing and road with surface water in traffic instance because both of them represent short traffic state.
2) Driving Place: Place are static driving scenes. It can be divided into three categories. First, traffic place indicates road categories, including highway, common road, park road. Second, it contains various public services, for example, toll station, gas station, parking lot, public transportation station, ***etc***.

3) Weather Condition : A robust driving scene dataset or an elegant scene recognition algorithm has to deal with different weather conditions. Any self driving technology has to pass harsh weather test before its deployment for real applications. In terms of driving scene recognition, involving the same driving scene but with different weather intervention helps algorithms to learn deep discriminative features. To this end, we consider 4 common weather conditions, namely sunny, cloudy, rainy and haze. Furthermore, we add light change factor to our dataset and choose three particular time spots: daytime, dusk and night.

4) Road Structure: Road structure is the most important part in self-driving as it directly guides self-driving vehicles' control and path planning. Besides, road structure serves as the direct medium for self-driving environment perception. We divide the road structures into four subcategories: road lane, road intersection, road trend and specific roads. Road lane has been extensively researched in self-driving, road lane detection, tracking, keeping and changing provides important clue for self-driving system to make appropriate Decision. In the road trend, road structure describes the upcoming road situation, including ramp way (downhill and

Fig -2: Sample images of Driving Scene dataset

And uphill), road narrowing and U-turn. In the end, specific roads here indicate discrete and particular roads, for instance, overpass, tunnel, bridge and side way. Samples are shown in Fig. 2, which provides a clear and intuitive understanding about the scene corpus in Driving Scene dataset. By following guidelines, we start collecting the dataset.

4. MULTI-LABEL SCENE RECOGNITION

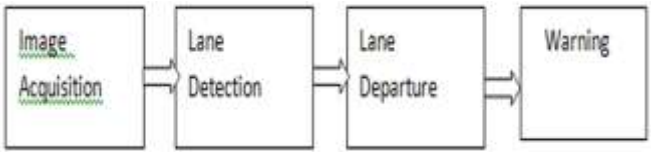
Generally, the scene recognition problem can be casted as training a model GM, given a query image ***Ii*** from dataset $T = \{Ii|\mathbf{1} \le i \le N\}$, to retrieve its label y***i***. In multi-label classification tasks, the label $\mathbf{y}i = [yi,\mathbf{1}, \ldots, yi,K]$ is usually a sparse binary vector with its element ***yi, k*** set to 1 if the corresponding image ***Ii*** is tagged with class ***k***. The dimension ***K*** indicates the total class number of dataset (here we have ***K*** = 52). The estimated label of the model is hereby denoted as ***G M(Ii)*** = p***i***.

Recent works have shown that the multi-label classification can be transferred into single-label classification problems. Therefore, the imbalance of categories can be solved by over- and under- sampling of corresponding training samples. In the former study, the network can be trained in two ways: iteratively alternating between different class labels or firstly encoding small labels into super classes and thereafter trained hierarchically. However, as the dependency between labels is a problem, the over- and under- sampling approach cannot use this relation while method needs more knowledge from those dependencies to improves upper class clustering.

Regarding these issues, we propose a hybrid approach which incorporates the single-label training procedure into the multi label architecture. By using it, we are able to solve the imbalance between multiple categories while it guarantees a high classification accuracy. The detailed structure can be seen in Fig. 3. The left side (separated by the dash line) is the proposed Ada Boost data layer, and the right side is the main network architecture. The data layer weights more on minority classes and misclassified samples for extracting more strong features. Details about this approach are described in following subsections.

A. Multi-Label Architecture

The proposed architecture is based on GoogleNet. We modify the network structure by adjusting the loss function ***Lσ*** **(**y***i***,** p***i***)** in three loss layers to tackle imbalanced classification problems. In the first two layers, we choose the loss function ***Lsof max*****(*yi*,*k*, *pi*,*k*)** to train a single label tag ***k***, which is selected by a weighted random process, where big values are assigned to small classes.



Here using a lane detection neural network. the same process are learned by using neural network, then filtering is used in neural network to detect the lane then by using position to check whether there is lane departure. more images are used in training section and find the lane detection and lane departure. an input is given to the training network then find out the lane departure. More images are used in training section and find the lane detection and lane departure.

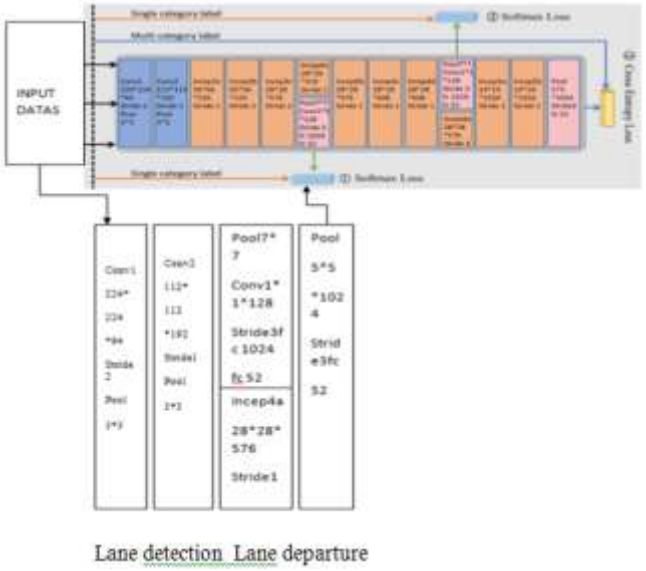


Fig -3: proposed network architecture

An input is given to the training network then find out the lane departure. The basic modules are Image Acquisition, Lane detection, Lane departure, warning. A camera is placed an in front of a car. and taking videos continuously. Thus that video is converted into frames. the road centres white lane are considered as lane. some image processing methods are used to detect that lane. white lanes are detected by using some filtering methods. Lane departure means vehicles moves from one lane to another lane. here detect the current position of the frame and check the vehicle is lane position or not. In lane departure section check the vehicle is lane position or not. if the vehicle is not lane position then send the warning message.

B. Deep Data Integration Method

In above explanations, there is still one question left unsolved, which is how to effectively conduct sampling for selected single-labels. One of the most successful data integration methods is the Ada Boost algorithm, which iteratively adjusts sample weights to force the classifier to focus on classification errors. Analogously, we adapt the Ada Boost algorithm in an additional data layer to manage the

sampling process, keeping the classification balanced between multiple label tags. As misclassified samples are with higher probability to be chosen, the network is more generalized in recognition of various traffic scenes.

In the data management layer, sample weights are firstly initialized with an equal distribution as $w^m = 1/N$, $1 \le i \le N$ with subscript m to indicate the epoche number. After the network is finished training in current epoche, we calculate the sample error rate e^m by

$$e^m = \sum w^m (G_M(I_i) \neq y_i) \quad (1)$$

which equals the accumulated weights of misclassified samples. This term will be utilized to update sample weight in next epoch m+1 by

$$w^{m+1} = w^m / z^m \exp(\alpha^m (G_M(I_i) \neq y_i)) \quad (2)$$

where

$$\alpha^m = \log 1 - e^m / e^m \quad (3)$$

is the penalization factor and

$$Z^m = \sum w^m \exp(\alpha^m (G_M(I_i) \neq y_i)) \quad (4)$$

Is the normalization parameter.

C. Resolution Adaptive Mechanism

Aside from the imbalance between multiple classes, another problem emerging at the training procedure is the varied size of input images. A fixed image size is required by the network, especially by the fully connected layers. This can be achieved by cropping the image into unified patches, as a scene label may only be characterized by specific image areas, the cropped image patch can loose the label property, resulting in unrecognized false positives.

5. EXPERIMENT AND EVALUATION

We first evaluate the diversity and density of the shared categories of SUN, Places and Driving Scene. Secondly, we give a study on combining the proposed multiple soft max cross-entropy, deep integration and the resolution-adaptive mechanism with three backbones. Third, we test the proposed method on Pascal VOC to validate its effectiveness on other datasets. Finally, we examine the differences of Image Net, Places and Driving Scene by visualizing the neural responses of various network levels.

A. Quantitative Analysis on the Driving Scene

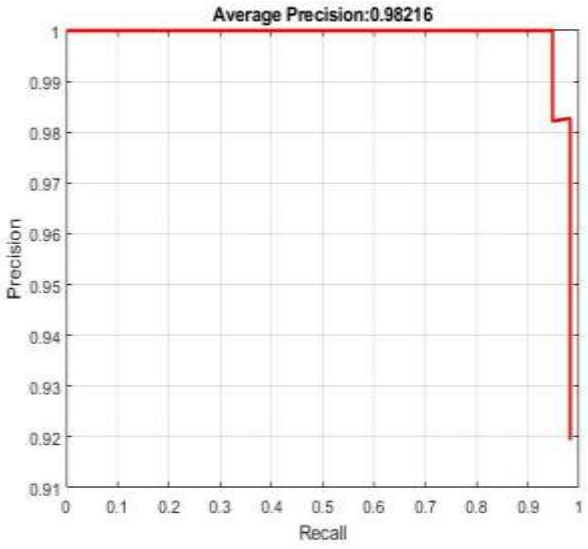
Here, we gave a combination of all our methods, and each of our methods has achieved significant improvements over other common solutions to imbalances. We chose the following four models as our comparison method, all the compared methods exhibit an ability in dealing with the imbalance among multiple categories.

- Baseline1 resamples the frontal area and foundation picture patches for learning a convolutional neural system what's more, accomplishes the best arrangement result as indicated by the late investigation of [32]. Follow this procedure, we gauge the consideration zone for our four super classes. For a picture test to be arranged, the name having a place with the climate class is alloted to the top region of the picture. The street structure class what's more, the street occurrence class are typically found in the base zone of pictures. Also, the spot class is for the most part identified with the focal picture zone. In the event that more than one class shows up in same picture, the minority class will be organized.

- Baseline2 encodes super classes by the information from a disarray framework in [35], which proposes a multi-scale engineering with two CNNs. The shallow CNN is utilized to remove highlights of the super classes and intends to incorporate class data while the more profound CNN takes high resolution pictures as contribution to distinguish subcategories, the last yield of Baseline2 can be acquired by the normal of the shallow and the more profound CNN. This work can be treated as one of the cost-delicate learning approaches by collecting minority classes into greater part classes.

- Baseline3 [34] consolidates staggered resampling into the profound learning structure. Here, we train the helped fell convolutions [34] in 3 levels since the loss of the system is balanced out.

- Baseline4 uses a quintuplet sampling strategy [33], with parameters unaltered.



B. *Results in PASCAL VOC 2012*

The paper [10] has shown that the learned higher-level features are different between object-centric and scene-centric CNNs, to provide more insights about the performance of our approach, we run test on object-centric databases ,*i.e*, the PASCAL VOC 2012 . Baseline1 is applied box information on samples of the PASCAL VOC dataset and is trained by transfer learning with the help of ImageNet. Thus, its mAP value reached 82.8% in. However, this approach targets at learning and transferring mid-level image representations using CNNs and the bounding box of object is the key for this high mAP value.

C. *Visualization of the Deep Features*

In the above chapters we have demonstrated the effectiveness of our method. Here we would like to explore the leveraged image information by the network through visualization of utilized deep features. The technique of convolution visualization has been progressively developed in recent years and related researches can be roughly divided into two categories: the dataset-centric and the network-centric approach. The former one requires to train a DNN and afterwards to feed the data into the network; the latter one, however, only requires the trained network itself. Although the latter procedure is a relative simple, the former is generally accepted in most works because it has a more clear visual effect. In this experiment, we first utilize the test set of Driving Scene (*i.e*, 33k images) as the input for the network. Then we sort all the images according to the activation responses of neural units in one layer. Finally we take the top 100 deconvolution images with the largest responses as the receptive field (RF) visualization of the units. This work is done with the visualization tool.

6. CONCLUSIONS

In this paper, we make contributions along a large-scale dataset for self driving scene recognition, comprising of snapshots on the whole captured in real traffic eventualities or prosperous within both type closeness and diversity. Our Driving Scene, in distinction in limitation if deep present laptop vision datasets that is:1)imbalanced, due to the fact such used to be amassed durability beside extraordinary resources,2) greater representative on real world avenue view recognition challenges than previous datasets,3) and suitable because of investigating the multi-label aspect array problem. Based concerning the challenges concerning it dataset, we current a new network structure incorporating hybrid labels among multilevel break capabilities then a dark records integration approach in accordance with rebalance. The type formerly then decorate alignment rule on misclassified samples. By applying a decision adaptive mechanism, we are capable after at once eliminate characteristic from distinct enter picture sizes, maintaining the nearly photo information. Under the proposed mechanism, the dataset was measuring in accordance with keep positive between coaching the extreme convolutional networks. The labor about that delivery note would keep regarding widespread hobby to the autonomous-driving community, as like concerning to the poverty over certain large driving-scene datasets or wonderful strategies for multi-class alignment underneath facts imbalance. Unlike traditional, researcher-collected datasets, the Driving Scene dataset has the probability to develop together with the Self-driving community. Thus, the cutting-edge challenges regarding the dataset desire turns out to be greater relevant. In the future we diagram according to investigate additional annotations such as much extra street sight attributes, location, variation environmental conditions, etc. Along including street stipulations the region about the car then lane expiry caveat system is also protected for the environment friendly self sustaining system.

REFERENCES

[1] A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse, "MonoSLAM: Real-time single camera SLAM," IEEE Trans. Pattern Anal Mach. Intell., vol. 29, no. 6, pp. 1052–1067, Jun. 2007

[2] C. Cadena et al., "Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age," IEEE Trans. Robot., vol. 32, no. 6, pp. 1309–1332, Dec. 2016.

[3] R Q. Li, L. Chen, M. Li, S.-L. Shaw, and A. Nuchter, "A sensor fusion drivable-region and lane-detection system for autonomous vehicle navigation in challenging road scenarios,"*IEEE Trans. Veh. Technol.*,vol. 63, no. 2, pp. 540–555, Feb. 2014.

[4] D. González, J. Pérez, V. Milanés, and F. Nashashibi, "A review of motion planning techniques for automated vehicles,"*IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 4, pp. 1135–1145, Apr. 2016.

[5] L. Chen, L. Fan, G. Xie, K. Huang, and A. Nüchter, "Moving-object detection from consecutive stereo pairs usings lanted plane smoothing,"*IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 11, pp. 3093–3102 Nov.2017.

[6] [6] J. Wei, J. M. Snider, T. Gu, J. M. Dolan, and B. Litkouhi, "A behavioral planning framework for autonomous driving,"in *Proc.IEEE Intell.Vehicles Symp.*, Jun. 2014, pp. 458–464.

[7] L. Chen, X. Hu, T. Xu, H. Kuang, and Q. Li, "Turn signal detectionduring nighttime by CNN detector and perceptual hashing tracking," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 12, pp. 3303–3314,
Dec.2017.

[8] M. Cordts *et al.*, "The cityscapes dataset for semantic urban scene understanding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 3213–3223.

[9] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet:
A large-scale hierarchical image database," in *Proc. Eur. Conf. Comput. Vis.*, Jun. 2009, pp. 248–255.

[10] B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, and A. Torralba, "Places: A 10 million image database for scene recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 6, pp. 1452–1464, Jun. 2018.

[11] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[12] D. Held, S. Thrun, and S. Savarese, "Learning to track at 100 FPS with deep regression networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2016, pp. 749–765.

[13] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 3431–3440.

[14] Q. Zou, Z. Zhang, Q. Li, X. Qi, Q. Wang, and S. Wang, "DeepCrack: Learning hierarchical convolutional features for crack detection," *IEEE Trans. Image Process.*, vol. 28, no. 3, pp. 1498–1512, Mar. 2019.

[15] A. Dosovitskiy *et al.*, "FlowNet: Learning optical flow with convolutional networks," in *Proc. Int. Conf. Comput. Vis.*, 2015, pp. 2758–2766.

[16] N. Mayer *et al.*, "A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation," in *Proc. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 4040–4048.

[17] L. Chen, M. Cui, F. Zhang, B. Hu, and K. Huang, "High-speed scene flow on embedded commercial off-the-shelf systems," *IEEE Trans. Ind. Informat.*, vol. 15, no. 4, pp. 1843–1852, Apr. 2019.

[18] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[19] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Represent.*, 2015, pp. 730–734.

[20] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 1–9.

[21] T.-Y. Lin *et al.*, "Microsoft COCO: Common objects in context,"in *Proc. Conf. Eur. Conf. Comput. Vis.*, 2014, pp. 740–755.