Survey of various road detection and extraction techniques for autonomous vehicles

[1]Poonam UttamShelke & [2]Dr. Parul S. Arora
[1]ME (Signal Processing) Student,[2]Professor
[1]E&TC, JSPM's ICOER,SPPU,
[1]JSPM's ICOER, Pune,India.

Received: January 23, 2019 Accepted: March 01, 2019

ABSTRACT: *Identification of the safe usable road is a crucial task for autonomous vehicles and robots. It is one of the main parameters which help in elevating the level of automation. The autonomous vehicle needs to understand the difference between various types of roads present in various conditions. External factors like lighting effects, shadows, weather conditions, occlusions, and mobile passengers also need to be considered. It helps in avoiding an obstacle, path planning, and decision making. It is useful in conditions where lane markings are not visible, example- road covered by snow or due to poor lighting conditions or lane markings not present (in certain rural and urban roads). There are many different ways to overcome these difficult situations using multiple sensors like monocular camera images, stereo images, RADAR sensor, LiDAR sensor, Inertial Measurement Unit (IMU), Global Positioning System (GPS) information and pre-loaded maps. Here we have presented a survey of various road extraction techniques.*

Key Words: : *Road Extraction, autonomous vehicles, LiDAR, Point Cloud, KITTI dataset, camera, maps*

I. Introduction

In recent years, the level of automation in road vehicles has increased tremendously. The deep learning technology advantage is effectively been used in extracting meaningful data from a large amount of real driving data. Automated vehicles are mounted with various sensors to help them assess the environment and to increase the robustness. There are many challenges, which the system needs to take care. Like- the difference in surface textures, illumination, moving or stationary vehicles, static or in motion people and pets. Roads having single or multiple lanes with marking or different marking or no markings. Roads with curbs or no curbs. Many different kinds of roads like forest, village, highway, and city are there. They all have different characteristic features, which makes us realize that for effective detection and extraction multiple algorithms need to be implemented. Rest of the paper is organized as follows, Section II contains the related work, Section III contain published results comparison table and Section VI concludes survey work.

II. RELATED WORK

LiDARs help to create a 3D map of the environment and hence can help in safe and accurate navigation. A large amount of data points generated by LiDAR along with accurate ego drive motion is required to register the points in the correct location.

Reference [1] proposed base FCN architecture with 21 layers. Experimented with early fusion, late fusion, and cross fusion to create a tensor having both the features of input camera and Lidar images. Cross fusion FCN is among the best MaxF score (96.03%) in urban category evaluated on KITTI road benchmark. Generative Adversarial Network (GAN) study was done by [2]. They have tried semi-supervised and weakly supervised semantic segmentation. Here along with unlabeled data, some supervision is provided i.e. some samples are labeled. GANs support semi-supervised segmentation by providing more useful information to the classification task. Performance analysis was done on various benchmarking datasets like Pascal VOC 2012, SiftFlow, StandfordBG and CamVid. StixelNet monocular approach was improved by [3]. Softmax-loss used during training for type-neurons. As clear road detection is complementary to identifying obstacles. This method can be used for road detection too.

Reference [4] used feed-forward architecture, MultiNet for real-time joint semantic reasoning with image classification and object detection. Weights are initialized to all layers of the encoder. During the back-propagation stage, only gradients are added. Softmax cross-entropy loss function is used for classification and segmentation. Future scope includes the use of compression methods to reduce the computational bottleneck and energy consumption.

MultiNet trains network that can segment the road quickly in real time. It can also detect and classify vehicles on the road. MultiNet architecture is three-headed. It begins with VGG16 network without three fully connected layers at the end. This forms encoder. "CNN encoder" reduces each input image to 512 set of features. For each region of the input image, Encoded features tensor captures the measure of how strongly each of 512 features is represented in that region. The new set of features specific to classification are formed using 1x1 convolutional layer. Detection has series of 1x1 convolutions that output a tensor with bounding box coordinates. Encoded Features only have 39x12 dimensions but the original input image is large 1248x384. Thus 39x12 is supposed to be too small to produce accurate bounding boxes. So, the network focuses on ROI align to have accurate bounding boxes. Segmentation output has fully-convolutional up sampling layers which increase the encoded features from 39x12x512 to the original image size of 1248x312x2. Digit 2 at the end shows that it's a binary mask, not the original image. It masks each pixel in the image as "road" or "not road" and its used in the score for the KITTI leader board.

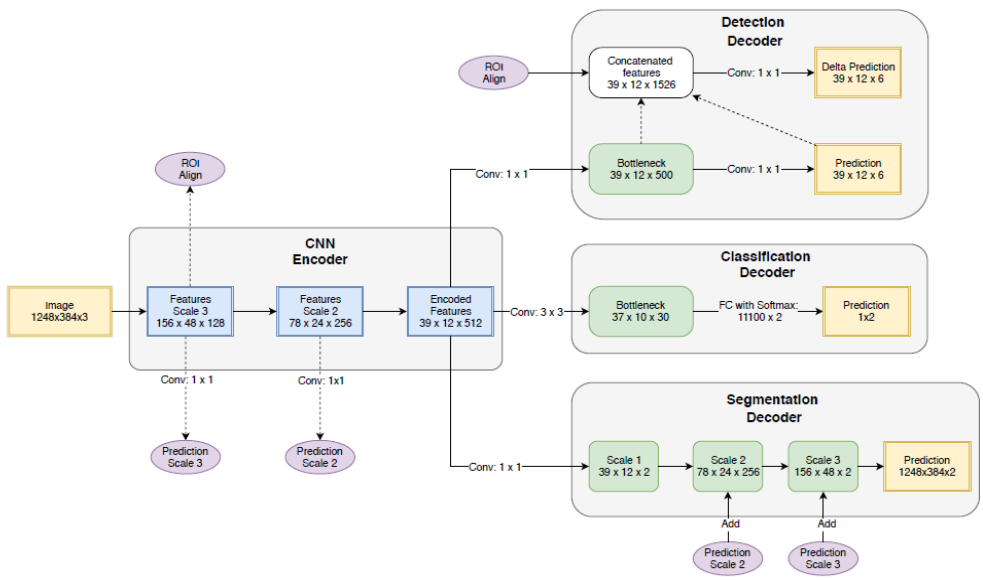

Fig. 1 MultiNet Architecture*[4]*

Fully Convolutional Neural Network (FCN) [5] for road detection used only LiDAR data. High-resolution feature maps are used to train FCN for pixel-wise semantic segmentation. From unstructured point cloud, top-view images of the vehicle's surroundings are generated. As the neural network is fully convolutional, it can process images of any size. Thus for road detection regions of interest (ROIs) can be dynamically changed and can even span 360 degrees in case of rotating Lidars view around thevehicle. To avoid over fitting and to improve generalization, each training example was rotated about LiDAR z-axis for angles in the range [-300, 300], using steps of three degrees. After rotation, each example was also mirrored about the x-axis. Adam optimization algorithm used for training.

Deep Fully Convolutional Network [6]was proposed using ResNet-101 network. This network has increased learning capacity. The overfitting can be prevented by use of data augmentation. It also reduces the gap between training and validation losses and is useful in Bird's Eye View (BEV). Future work suggested is the addition of post-processing layer into the system to get smoother results and to include high-level information provided by digital navigation maps. Up_Conv_Poly [7] developed as an efficient deep model for monocular road segmentation. Parameter reduction is achieved by using VGG-16 classification network as a basis for the contraction side of the network. Further, the numbers of parameters are reduced by reducing the number of FC-conv filters. The width of the up-convolution side of the network is increased to improve system accuracy. U-nets used for the distribution of parameters (it has a variable number of filters which are the same between the contraction and expansion side). The network is trained by backpropagation using stochastic gradient descent (SGD) with momentum. The proposed network has an advantage with respect to speed and segmentation accuracy.

Deep deconvolutional networks and CNN's based network [8] used multipatch training. Output images are split into multiple patches. During learning, eachpatch is looped over and training happens on the entire input image only to predict the pixels in that specific patch. This method is effective in breaking down problems in challenging situations.

Drivable road area detection in monocular images using Convolutional Neural Network (CNN) was proposed in [9]. The training road annotations are automatically generated using OpenStreetMap, vehicle pose estimation sensors and camera parameters. Generation of road annotations without any human intervention was one of the main motives, which can be used to train road classifier. This helps in having scalability and reduces cost. Use of maps islimited to label images taken from ground and classifier is not dependent on it during testing. For robustness – H and S channels from HSI space and Cb and Cr channels from YCbCr space are taken as color features (representation of the appearance of pixels). K-means clustering is used along with L method to determine K value. During back-propagation soft-max loss used. Further work includes – use of temporal information present in videos to improve the labeling step, use of motion information present in videos to remove false positives on the moving vehicles and extending the capability of the system for extreme weather conditions like rain and snow.

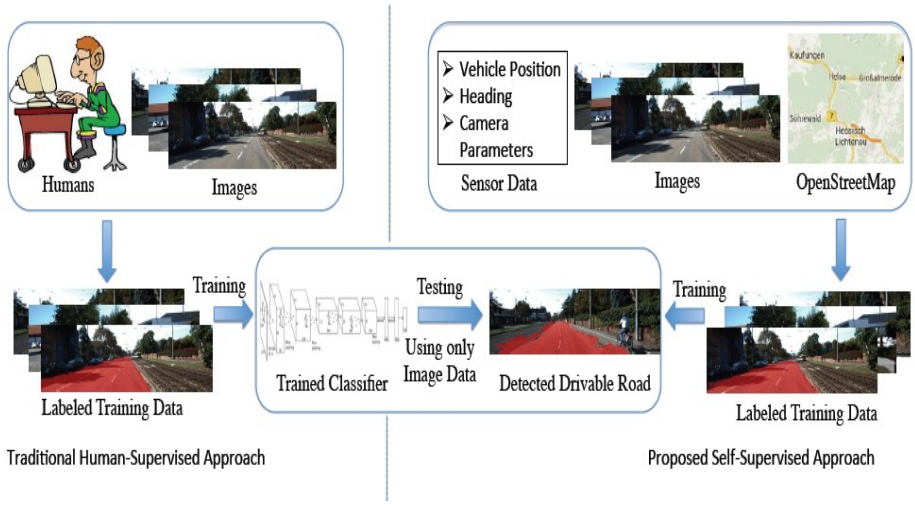

Fig.2. Flowchart depicting human-supervised and proposed map-supervised road detection approaches.[9]
Human supervised approaches use human to label the training data. Here, the author uses publicly available OpenStreetMap data, vehicle pose, camera parameters, and the pixel appearance feature to label the training data. During testing, only image data is used.

Thus the approach includes two steps:

1. Automatically build a set of noisy labeled images using maps, localization sensor data, and camera parameters. Annotation noise reduced using pixel appearance features.
2. Next, Fully Convolutional Network is trained using above automatically generated labels.

Machine learning approach for road detection which can extract important details from huge available information with speed based on Convolutional Neural Network (CNN) model is presented [10]. Input is a patch from the image and the result is in the form of class as road or non-road. CNN is then converted into a Fully Convolutional Network (FCNs) by converting the fully connected layers into convolutional layers. Conversion at inference time allows the use of large contextual windows and maintains real-time inference time. Further improvement needed in classifying different types of road surfaces or regions under extreme lighting conditions

Reference [11] proposed a hierarchical approach for labeling semantic objects and regions in scenes. Low-level information is combined with higher-order reasoning using a hierarchical representation. The problem is simplified into a series of sub-problems which are specifically trained to perform well for the concerned task. Thus, the test-time structured prediction is a sequence of predictions. As the labels are modeled over regions the method is robust to imperfect segmentation. Features defined over large regions can be used and no hard commitments are there during inference. Over fitting is avoided by taking unseen data while prediction.

Monocular vision and analyzing problem in column-wise regression was achieved in [12]. Which is further solved using a deep convolutional neural network (CNN). The network is trained on data generated from a laser-scanner point cloud. The idea is based on “Stixel-World” called “Stixel”. The goal is to find a bottom pixel of each “Stixel”. Input is a single RGB image vertical stripe.

Block scheme method based on contextual information, image features and classifier study was carried out in [13]. In monocular road detection, images are segmented into the road and non-road regions. Contextual

blocks provide information about the surroundings of the classification blocks. During training mini-batch stochastic gradientdescent used with momentum. Particle swarm optimization (PSO) algorithm is used for optimizing parameters like a number of neurons, learning rate, hidden layer maximum norm, and output layer maximum norm. Effect of using contextual blocks and their radius parameter using all image features is tested up to a radius of 3. The number of blocks and features increases with increase in radius but it's doesn't affect the processing time due to the implementation of contextual blocks. (i.e. features are pre-calculated for the whole image and then appropriately concatenated for each classification block).

Fusion of LiDAR and monocular image in the framework of Conditional Random Field (CRF) was achieved in [14]. The Lidar points are aligned with pixels in the image by cross-calibration. Boosted decision tree classifier is used for both image and point cloud. Image-based features are texture, Dense Histogram of Oriented Gradients (HOG), Color and location. Point Cloud Feature used are simple geometric feature like normalized 3D location (w.r.t. the Euclidean distance) and the direction of the local normal vector. For energy minimization open source library Darwin is used for getting most probable labeling. As the random field model is built on the pixel lattice, it takes more time. Speed can be increased by grouping the image into superpixels or patches, using Lidar points projected into the units(to reduce the uncertainty), label each lidar point as a random variable and extend the points with the features of the aligned pixels. Further improvements in results are possible by use of higher-order CRF or fully connected CRF.

Probabilistic distribution based on Texton (2D texture and color) and Dispton maps (3D information) [15] was used to model weak classifiers like Joint Boosting classifier. Watershed Transformation is used to calculate the superpixel and feature maps are created based on Textons and Disptons. Further improvement can be made by using Self-Organizing Maps to have better recognition for different classes like vehicles, buildings, and sidewalks.

Reference [16] proposed SPatial RAY (SPRAY) features to enhance local classification decisions. It has three parts: base classification, SPRAY feature generation, and road terrain classification. Each base classifier creates a map of confidence values, wherein each location corresponds to acertain location in the metric space which is internally obtained using inverse perspective mapping. Confidence map value shows that corresponding cell in metric space has a certain property or not. Value continuous spatial representation is created by combining all confidence maps. Base boundary classifier generates low confidences on the road-like area and high confidences at locations that correspond to boundaries. During SPRAY feature generation, confidence map from a base classifier is taken as input for a defined number of base points (BP) in the metric representation. The spatial layout with respect to the confidence map is captured at each individual base points by radial vectors called rays. Ego spray feature indicates if a base point is located on the ego-lane. For training GentleBoost classifier is used. The algorithm generates a sequentially weighted set of weak classifiers that build a strong classifier in combination. During each training iteration, it attempts to find an optimal classifier considering input signal distribution weights. Further suggested improvements are – finding optimal configurations for different scenarios like highway and inner city and using larger dataset during the training process.

Reference [17, 18] proposed method based on dense 3D lidar data thus it is independent of road markings. [13] First projected sparse lidar points on 2D reference plan and computed dense height map using the upsampling method. Next probability distributions of neighboring regions are compared according to a similarity measure and morphological operations used to enhance performance. This method overcomes challenges like the unknown number of lanes or slopes but further improvements are needed in terms of speed and accuracy. [18] Projected Lidar point clouds into the images and obtained original sparse height images. Then, they extracted the image based features (textons), Lidar-based features (LDD) and location features to train an AdaBoost classifier. Finally, Conditional Random Field (CRF) framework was used to get the road detection results.

III. Results and discussion

Table 3.1KITTI Road Benchmark results (in percentage %) on the urban road category (only published reports).

Ref. No.	Method	Details	MaxF (%)	AP (%)	PRE (%)	REC (%)	Time (s)
1	LidCamNet	FCN, Camera and Lidar images, Cross fusion	96.03	93.93	96.23	95.83	0.15
2	SSLGAN	Image, Generative Adversarial Network (GAN)	95.53	90.35	95.84	95.24	0.70

3	StixelNet II	Lidar, camera Images, StixelNet monocular approach, Caffe framework	94.88	87.75	92.97	96.87	1.20
4	MultiNet	Image, GPS information with open-street map data, Feed – forward architecture, MultiNet, Softmax	94.88	93.71	94.84	94.91	0.17
5	LoDNN	Only Lidar data, FCN, Feature maps, Pixel-wise semantic segmentation	94.07	92.03	92.81	95.37	0.018
6	DEEP_DIG	Images, ResNet-101 Network, Caffe framework	93.98	93.65	94.26	93.69	0.14
7	Up_Conv_Poly	Image, Monocular road segmentation, VGG-16 Classification network	93.83	90.47	94.00	93.67	0.08
8	DNN	Images, Deep deconvolutional networks, CNN with multipatch training	93.43	89.67	95.09	91.82	2
9	FTP	CNN, monocular images, Generation of automatic road annotations using maps	91.61	90.96	91.04	92.20	0.28
10	FCN_LC	Images, CNN , FCNs	90.79	85.83	90.87	90.72	0.03
11	HIM	Images, Hierarchical approach	90.64	81.42	91.62	89.68	7
12	StixelNet	Images, CNN , Stixel base	89.12	81.23	85.80	92.71	1
13	CB	Contextual information, Partical swarm optimization (PSO) algorithm	88.97	79.69	89.50	88.44	2
14	Fused CRF	Lidar, monocular vision, Conditional Random Field (CRF), Boosted decision tree classifier	88.25	79.24	83.62	93.44	2
15	ProbBoost	Images, Self-Organizing Maps, Texton (2D texture), Dispton maps (3D information)	87.78	77.30	86.59	89.01	150
16	SPRAY	Images, Spatial RAY	87.09	91.12	87.10	87.08	0.04
17	Road detection using High resolution LiDAR	3D Lidar data converted into 2 D reference plane , dense maps and morphological operations	82.72	87.58	85.44	80.17	NA

Kitti Dataset offer pixel-based evaluation and behavior-based evaluation with following metrics: PRE (Precision), REC (Recall), MaxF (Maximum value of F-measure), AP (Average Precision), FPR (False Positive Rate) and FNR (False Negative Rate)

F value is the measure of effectiveness. It's a trade-off using weighted harmonic mean between precision and recall. High precision indicates many road pixels are correctly classified. The recall is the ability to detect road surface.

From above table, we can see majority of the algorithms are based on Image data as it is sufficient to extract all the required information in well illuminated condition. Having additional sensor like Lidar can reinforce the decision and can help in adverse conditions.

LidCamNet [1], it's based on the integration of camera images and LiDAR point clouds is among the top-performing algorithms on KITTI road benchmark with MaxF score of 96.03% in urban category. Hence, the advantages of both the sensors are considered and they both can be employed for better functioning. Having automatically annotated images [9] can help in path planning and decision making. Having only Lidar [5, 17]based system is also coming under top performing algorithms but its scope is limited due to heavy equipment costs compared to camera system.

IV. Conclusion and future scope

Integrating multiple sensors information (like LiDAR sensor, camera, and digital navigation maps) canbe considered for obtaining robust and accurate segmentation in various external conditions. The trade-off between performance and computing capacity need to be taken care when dealing with real-time data. The advantage provided by neural networks can be utilized further to have higher efficiency.

V. Acknowledgment
Special thanks to all the staff of JSPM's ICOER and AIT, Pune for their continued support.

References

1. L. Caltagirone, M. Bellone, L. Svensson, and M. Wahde, "LIDAR–camera fusion for road detection using fully convolutional neural networks" ,Robotics and Autonomous Systems,Volume**111**,pp. **125-131, 2019**
2. N. Souly, C. Spampinato and M. Shah, "Semi Supervised Semantic Segmentation Using Generative Adversarial Network," 2017 IEEE International Conference on Computer Vision (ICCV), Venice, pp. **5689-5697, 2017**.
3. N. Garnett, S. Silberstein, S. Oron, E. Fetaya, U. Verner, A. Ayash et al , "Real-time category-based and general obstacle detection for autonomous driving", in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. **198-205, 2017**
4. Teichmann, M., Weber, M., Zöllner, J. M., Cipolla, R. &Urtasun, R., "MultiNet: Real-time Joint Semantic Reasoning for Autonomous Driving.", CoRR abs/**1612.07695** ., **2016**
5. L. Caltagirone, S. Scheidegger, L. Svensson, and M. Wahde, "Fast LIDAR- based Road Detection Using Fully Convolutional Neural Networks", in Intelligent Vehicles Symposium (IV), IEEE, pp**1019-1024, 2017**
6. J. Munoz-Bulnes, C. Fernandez, I. Parra, D. Fernandez-Llorca, and M. A. Sotelo, "Deep Fully Convolutional Networks with Random Data Augmentation for Enhanced Generalization in Road Detection", Workshop on Deep Learning for Autonomous Driving on IEEE **20**th International Conference on Intelligent Transportation Systems, **2017**.
7. G. L. Oliveira, W. Burgard and T. Brox, "Efficient Deep Models for Monocular Road Segmentation", IEEE RSJ International Conference on Intelligent Robots and Systems (IROS), pp. **4885-4891. 2016**
8. R. Mohan, "Deep Deconvolutional Networks for Scene Parsing", arXiv:**1411.4101v1** [stat.ML] Nov **2014**
9. L. Ankit, M. K. Kocamaz, L. E. Navarro-Serment, and M. Hebert, "Map-Supervised Road Detection", in IEEE Intelligent Vehicles Symposium (IV), Gothenburg, Swede, June 19-22, pp. **118-123, 2016**
10. C. C. T. Mendes, V. Frémont and D. F. Wolf, "Exploiting fully convolutional neural networks for fast road detection," **2016** IEEE International Conference on Robotics and Automation (ICRA), Stockholm, pp. **3174-3179, 2016**
11. D. Munoz, J A Bagnell and M. Herbert, "Stacked Hierarchical Labeling", in European Conference on Computer Vision (ECCV) , pp. **57-70, 2010**
12. D. Levi, N. Garnett and E. Fetaya, "StixelNet: A Deep Convolutional Network for Obstacle Detection and Road Segmentation",**26**th British Machine Vision Conference (BMVC) ,**2015**
13. C. C. T. Mendes, V. Fremont and D. F. Wolf,"Vision-Based Road Detection using Contextual Blocks", arXiv:**1509.01122v1 [cs.CV] 3 Sep 2015**
14. L. Xiao, B. Dai, D. Liu, T. Hu and T. Wu ,"CRF based Road Detection with Multi-Sensor Fusion", in IEEE Intelligent Vehicles Symposium (IV) June **2015**, pp. **192-198**
15. G.B. Vitor, A.C. Victorino and J.V. Ferreira, "A probabilistic distribution approach for the classification of urban roads in complex environments", IEEE Workshop on International Conference on Robotics and Automation (ICRA), May **2014**
16. T. Kuhnl, F. Kummert and J. Fritsch, "Spatial Ray Features for Real-Time Ego-lane Extraction",**15**th International IEEE Conference on Intelligent Transportation Systems (ITSC), September **2012**, pp. **288-293**
17. R. Fernandes, C. Premebida, P. Peixoto, D. Wolf and U. Nunes, "Road Detection Using High Resolution LIDAR," 2014 IEEE Vehicle Power and Propulsion Conference (VPPC), Coimbra, pp. 1-6, 2014
18. X. Han, H. Wang, J. Lu, and C. Zhao,"Road Detection Based on the Fusion of Lidar and Image Data." International Journal of Advanced Robotic Systems, pp**1-10, 2017**