

# Multi-lane Modelling using Convolutional Neural Networks and Conditional Random Fields

Ganesh Babu<sup>1</sup>, Ramchandra Cheke<sup>2</sup>, Ganesh Sistu<sup>3</sup> and Senthil Yogamani<sup>3</sup>

<sup>1</sup>School of Mathematics and Statistics, University College Dublin, Ireland

<sup>2</sup>University of Limerick

<sup>3</sup>Valeo Vision Systems, Ireland

## Abstract

Lane detection and modelling is a crucial module in autonomous driving which enables the vehicle to drive within the ego lane. Typically, CNN based semantic segmentation is used to segment lane markings and then a post processing algorithm fits polynomial models for the lanes based on the road geometry. Recently, direct regression of the lane polynomials were explored but it is still not a mature solution. In this paper, we propose a combination of deep learning based semantic segmentation and a graphical model based lane fitting. We use conditional random fields (CRFs) to effectively fit lane polynomials in the presence of noisy segmentation maps. The proposed method provides an accuracy improvement of 15% relatively to the conventional post processing baseline.

## INTRODUCTION

Over the years, progress in autonomous driving has grown due to the involvement of deep learning in computer vision. The introduction of neural networks to solve autonomous driving vehicle issues has expanded the horizons and brought the vehicles to reality. Autonomous Driving tasks such as perception which involves object detection [1, 2, 3], soiling detection [4, 5, 6], semantic segmentation [7, 8], weather classification [9, 10], depth prediction [11, 12, 13, 14, 15], moving object detection [16], SLAM [17, 18, 19], fusion [20] and multi-task learning [21, 22, 23] are challenging due to the highly dynamic and interactive nature of surrounding objects in the automotive scenarios [24]. However, lane modelling remains a challenging problem. Lane detection is a process of detecting lane areas or lane lines by camera or lidar [25] and then projecting them to localize the position of the vehicle for future decision making. In recent years, tremendous progress has been made in terms of detection accuracy but at the cost of heavy pre-processing and high computation for lane projection. Furthermore, the perception of deep learning being a 'black box' and the uncertainty on what drives the decision making makes it hard for autonomous driving to be approved by many.

In this paper, we propose a hybrid approach to multi-lane modelling which utilizes a convolutional neural network (DeepLabv3+) for the initial detection of lanes at a pixel level and then uses a statistical graphical model (conditional random fields) to model the lanes to further improve the accuracy of lane detection. This approach works on top of the CNN predictions and build a CRF graph based on the associated pixels. The energy of the graph is then minimized by removing the unwanted nodes and edges to find the best solution. Final connected nodes are then used to fit the polynomials to find the multiple lanes in the image.

## BACKGROUND

### Multi-lane Detection

Multi-lane detection is a process of detecting and projecting multiple lanes in the road to assist the autonomous driving vehicles. Over the years, extensive research has been performed on the former and the most conventional is to use a CNN to predict the lane pixels and then perform an extensive post-processing techniques like feature extraction, edge detection etc. on the CNN output. Once the post-processing is completed, RANSAC (Random sample consensus) is used to fit the lanes [26]. It is known that RANSAC is an iterative exhaustive search and computationally intensive. Several alternative approaches have been proposed to overcome the same and some of them are cascaded CNNs [27], end to end lane position estimation [28] and end to end segmentation methods like GCN [29] and SCNN [30]. Most of the proposed methods emphasize on increasing the complexity CNN architectures or using a stand alone network for multi-lane detection which is neither feasible nor cost-effective in the real world scenario with limited computational resources. Hence we propose a graphical approach on top of the existing CNN output to increase the prediction and decrease the complexity of the solution. The graph based approach is similar to [31].

### DeepLabv3+

DeepLabv3+ [32] is the improved version of DeepLabv3 [33] by leveraging the advantage of encoder-decoder architecture. DeepLabv3+ provides various advantages for the semantics segmentation task such as dense prediction using Atrous convolution [34], memory optimisation using depth-wise separable convolution [35] and multi-scale processing using Atrous Spatial Pyramid Pooling (ASPP) module. These important elements are discussed as follows:

**Atrous convolution:** Atrous convolution increases the spatial resolution of feature maps while using deep convolution neural networks. In Atrous convolution, dilation rate defines a spacing between the adjacent values in the kernel. Therefore, multi-scale information is captured by controlling dilation rate, hence enhancing generalization ability of the network.

**Depth-wise separable convolution:** Depthwise separable convolution, breaks down a standard convolution operation into two parts. First it performs depthwise convolution followed by a pointwise convolution. Precisely, the depthwise convolution performs a spatial convolution on each input channel separately by creating different kernels for each input chan-

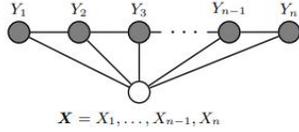


Figure 1: A graphical representation of chain-structured CRFs [40].

nel. The output of depthwise convolution is then convoluted with a 1x1 kernel to obtain a single channel output. Therefore, the number of 1x1 kernels decides the number of output channels in the depth-wise separable convolution. This creative approach not only significantly reduces computation complexity but also improves performance.

**Atrous Spatial Pyramid Pooling:** Several architectures have been proposed to extract features at multiple scales [36, 34]. DeepLabv3+ uses Atrous Spatial Pyramid Pooling (ASPP) with atrous rates of 6,12 and 18 to process the DCNN output.

**Network backbone:** In this work, we used Resnet-50 [37] as a backbone. We performed the experiment using a PyTorch [38] based implementation.

### Conditional Random Fields

Conditional random fields (CRFs) are a probabilistic framework for labeling and segmenting sequential data, based on a conditional approach. It is a form of undirected graphical model globally conditioned on  $X$ , the random variable representing observation sequences and defines a single log-linear distribution over label sequences for the particular observation sequence [39].

We define  $G = (V, E)$  to be an undirected graph such that there is a node  $v \in V$  corresponding to each of the random variables representing an element  $Y_v$  of  $Y$  [40], where  $Y$  is the corresponding label sequence for the observation sequence  $X$  and  $Y_v$  is the label of the node  $v$  in the graph. If each random variable  $Y_v$  obeys the Markov property with respect to  $G$ , then  $(Y, X)$  is a conditional random field. A simple graphical representation while modelling the sequence is shown in Figure 1.

The graphical structure of a conditional random field may be used to factorize the joint distribution over elements  $Y_v$  of  $Y$  into a normalized product of strictly positive, real-valued potential functions, derived from the notion of conditional independence [40]. Each potential function operates on a subset of the random variables represented by vertices in  $G$  and as per the conditional independence, the absence of an edge between two vertices in  $G$  implies that the random variables represented by these vertices are conditionally independent given all other random variables in the model. It is the responsibility of the potential functions to ensure that it is possible to factorize the joint probability such that conditionally independent random variables do not appear in the same potential function. In order to do so, each potential function is made to operate on a set of random variables whose corresponding vertices form a maximal clique within  $G$ . In Figure 1, each potential function will operate on pairs of adjacent label variables  $Y_i$  and  $Y_{i+1}$ .

The probability of a given label sequence  $y$  given an obser-

vation sequence  $x$  can be written as,

$$p(y|x, \lambda) = \frac{1}{Z(x)} \exp \sum_j \lambda_j F_j(y, x) \quad (1)$$

Where

$$\frac{1}{Z(x)} = \text{Normalization Factor}$$

Similarly, the log-likelihood for the CRF can be written as,

$$L(\lambda) = \sum_k \left[ \log \frac{1}{Z(x^{(k)})} + \sum_j \lambda_j F_j(y^{(k)}, x^{(k)}) \right] \quad (2)$$

The major advantage of CRFs over hidden Markov models (HMMs) is their conditional nature which relaxes the assumption of independence required by the latter to ensure tractable inference. Furthermore, CRFs also avoid the label bias problem which is a major weakness exhibited by the maximum entropy Markov models (MEMMs) and other conditional Markov models based on directed graphical models. CRFs outperform both MEMMs and HMMs on a number of real-world sequence labeling tasks [40].

### PROPOSED APPROACH

The architecture of the proposed solution is shown in Figure 2. First we trained the DeepLabv3+ model using the 3623 training images of TuSimple data for 80k iterations. The trained model predicts whether a pixel in a given image is lane or not (see Figure 2 (top left sub-figure)). The predictions from the CNNs are noisy and cannot be directly used to formulate the CRF graph or model the lanes similar to [31]. Hence we first perform a low level association to reduce the noise and group the pixels to supermarkings, in turn reducing the computational cost on the CRF. Supermarking is a collection of closely available lane pixels grouped into a single large pixel as illustrated in the Figure 2 (bottom left sub-figure). Each coloured region in the image is a supermarking.

The low level association is performed as two steps i.e pixel level and supermarking level. At first, the centre of continuous pixels classified as lane in each row of the image is extracted i.e. if pixel 1 to 10 in row 1 of the image is classified as lane, then pixel 5 is taken as one of the lane points in row 1. If there is another group of pixels from 51 to 60 in row 1 that are classified as lane then pixel 55 is taken as another lane point present in the same row and this process is repeated for all the rows in the image. Then for each extracted lane point  $i$  in each row, the distance and orientation [31] from lane point  $j$  in subsequent rows is calculated as shown below.

$$dist_{geo}(s_i, s_j) = | (x_{t_j} - x_{h_i}) \sin \theta_{h_i} - (y_{t_j} - y_{h_i}) \cos \theta_{h_i} | + | (x_{t_j} - x_{h_i}) \sin \theta_{h_i} - (y_{t_j} - y_{h_i}) \cos \theta_{h_i} | \quad (3)$$

$$dist_{dir}(s_i, s_j) = | \theta_{t_j} - \theta_{h_{t_j}} | + | \theta_{h_i} - \theta_{h_{t_j}} | \quad (4)$$

where  $dist_{geo}$  = Geometric distance between the two lane pixels  $i, j$  or supermarkings  $s_i, s_j$

$dist_{dir}$  = Angle between the two lane pixels  $i, j$  or supermarkings  $s_i, s_j$

$\theta$  = Angle of the corresponding lane pixel with reference to

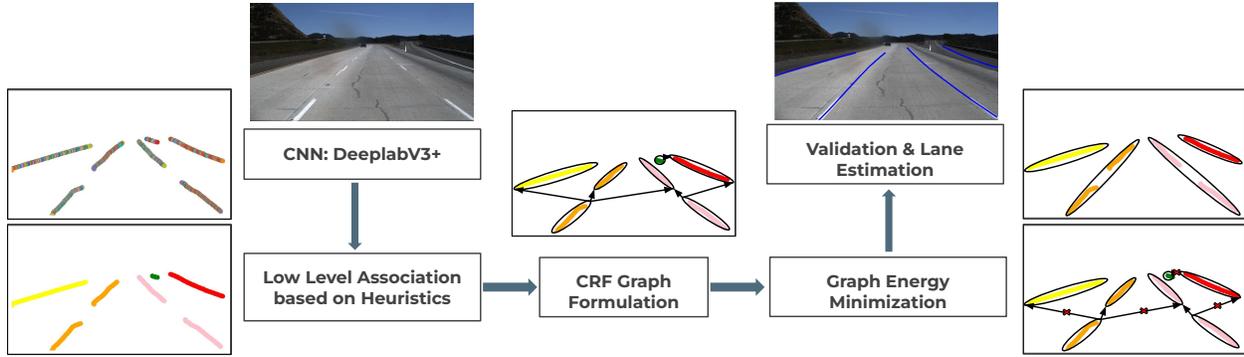


Figure 2: Architecture of the proposed approach

the  $x$  or  $y$  axis

$x_h, y_h = x$  and  $y$  coordinate of the head of the lane pixel or supermarking

$x_t, y_t = x$  and  $y$  coordinate of the tail of the lane pixel or supermarking

If two lane points in different rows are within a certain distance and angle (orientation), then we combine the two points into one and call them a supermarking. This process is repeated for all the lane points in the image. Once the supermarkings are created we compute the angles at the top, bottom and centre for each supermarking before performing the second step of the low level association. In the second step, we perform the same process followed in the first step but on a supermarking level. The distance and orientation of each supermarking with relevance to the rest of the supermarking is computed. Alongside, we also compute the probability of association [31] between the two supermarkings as shown below,

$$P(s_i, s_j) = \frac{1}{z} \exp\left(-\frac{|\theta_{c_i} - \theta_{c_j}|^2 + |\theta_{c_j} - \theta_{c_i}|^2}{\sigma^2}\right) \quad (5)$$

We combine the two supermarkings into a superior supermarking, if

- the distance and orientation between two supermarkings are within the allowed threshold.
- the starting point of supermarking 1 is greater than the end point of supermarking 2 and vice versa.
- the probability of association between the supermarkings is greater than 0.98

Once the low level association is performed, we then create the CRF graph  $G = (V, E)$  based on the supermarking created. In the CRF graph, the superior supermarkings become the vertex of the graph and the edges are drawn between such vertices, if they share a common supermarking (see Figure 2 (center sub-figure)).

In a CRF graph, the energy of the graph is a combination of the unary and pairwise potential. The unary potential defines the probability of the occurrence of the vertex (superior supermarking) in the graph and the pairwise potential defines the probability of a relationship between two vertices connected by an edge in the graph. The unary and pairwise (clique) potential [31] can be represented as follows,



Figure 3: Sample Images from TuSimple dataset.

$$U(l_k | s_k) = -\ln(P(l_k | s_k)) \quad (6)$$

$$\phi_C(l_C | s_C) = -\ln(P(l_C | s_C)) \quad (7)$$

where  $C$  is the clique set  $cl(G)$  in graph  $G$  and  $l_k$  is the label of the node  $s_k$ .

The best solution is the graph with the minimal energy. In order to achieve that, the CRF graph is pruned in such a way that the edges with low probability scores are removed and only the cliques with high probability of occurrence are retained (see Figure 2 (bottom right sub-figure)). The superior supermarkings in the cliques are then combined to form even larger supermarkings.

The final set of superior supermarkings and the supermarkings not involved in the superior supermarkings are then evaluated based on length, number of pixels and the orientation. If they are all within the allowed limits, then a high order polynomial is fitted on the pixels of the eligible supermarkings to model the final lanes in the image.

Thus the best label set  $L^*$  obtained by minimizing the energy of the graph can be represented as,

$$L^* = \arg \max_L \sum_{l_k \in L} \ln(P(l_k | s_k)) + \sum_{C \in cl(G)} \ln(P(l_C | s_C)) \quad (8)$$

## RESULTS

For the scope of the work, we used TuSimple data. TuSimple data is extensively used in several lane detection research works

Model	Actual Lanes	Predicted Lanes	Accuracy	FPR	FNR
Contour Level Regression	9938	12074	0.714	0.234	0.354
CRF based Regression	9938	10010	<b>0.946</b>	<b>0.106</b>	<b>0.081</b>

Table 1: Quantitative comparison of the proposed approach and the baseline. (FPR is False Positive Rate & FNR is False Negative Rate)

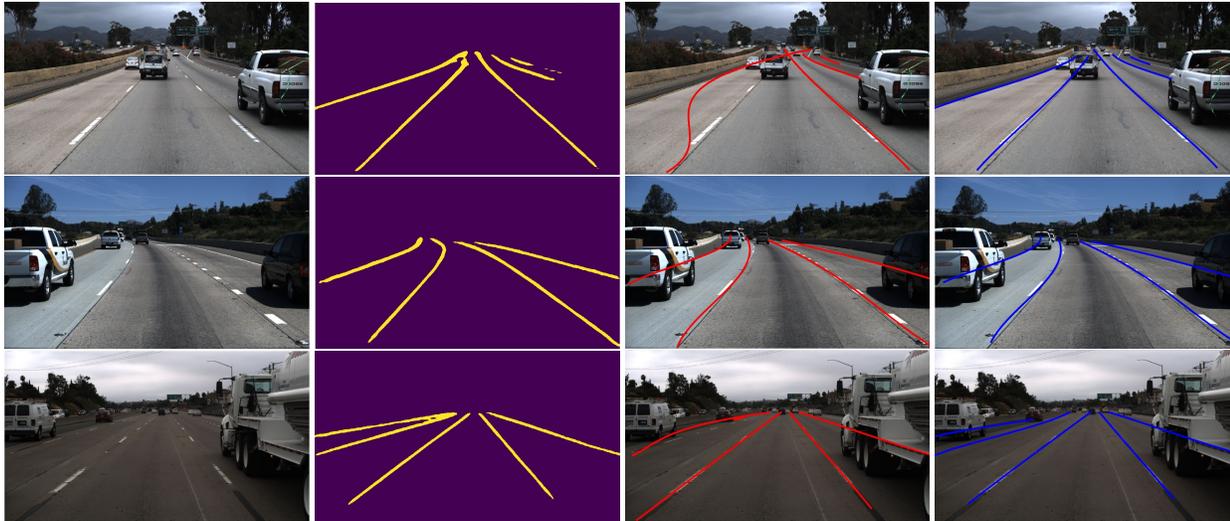


Figure 4: Comparison of CRF based Multi-lane modelling with Contour based Multi-lane Modelling. First Column: Actual Image; Second Column: Deeplabv3+ Model Prediction; Third Column: Contour Based Lane Modelling; Fourth Column: CRF Based Lane Modelling.

and gives a much better standard for comparison. The TuSimple dataset consists of 6,408 road images on US highways captured using the dashboard camera. The resolution of image is  $1280 \times 720$ . The dataset is composed of 3,626 images for training, 358 for validation, and 2,782 for testing [27]. These test dataset is captured under different weather conditions. Some sample images in the dataset are shown in Figure 3.

First a DeepLab V3 model with ResNet-50 [37] encoder is trained for 64 epochs on the TuSimple training dataset. The images are resized to  $640 \times 360$  to maintain an aspect ratio. Model was trained on Nvidia V100, 16GB dual GPU machine. The model is trained to classify the lanes at pixel level in a binary classification fashion. The reported accuracy of this model on the validation data is 66.32% mIoU.

In the second stage the test set images are processed through the DeepLab V3 [34] model for pixel level semantic classification of lanes and then the lane binary images are processed through the low-level association stage followed by the graphical model stage to extract the multi-lane labels. No morphological or handcrafted filters are used to enhance the results.

A traditional contour based model is chosen as baseline for comparison with the graphical model based approach. The traditional pipeline involves smoothing of binary labels from DeepLab V3 model followed by contour extraction. These individual contours then processed iterative through a second order lane polynomial fitting.

The quantitative results are presented in the Table 1. The three evaluation metrics accuracy, False Positive Rate (FPR) and

False Negative Rates (FNR) are generated by the scripts provided by the authors of TuSimple[27]. Proposed approach has shown a significant improvement of 10% in accuracy over the baseline method. The FPR and FNR also reduced significantly. Though methods like [29] and [30] have shown slightly higher performance than the proposed method, they use a dedicated CNN network followed by the standard pixel level classification module and these two are tightly coupled. Where as the proposed solution is agnostic to the pixel classification module and it is easily extendable to other classes like curbs, sidewalks etc.

The qualitative results are shown in Figure 4. Each row of the image comprises of the actual image, neural network output, prediction based on contour level regression and prediction based on CRF based regression. In row 1 and 3, we can clearly see that the lanes predicted by the proposed approach is precise and has outperformed the contour based approach significantly.

## CONCLUSION

Multi-lane detection is a classical problem that was addressed in many research works previously. While conventional lane fitting (RANSAC) algorithms use a pixel-level classifier followed by heavy heuristics based post-processing and lane fitting algorithms, majority of the modern approaches focused on solving this problem by a dedicated neural networks with complex task specific computing blocks. In this work, we explored a new probabilistic strategy as an alternative to this. The proposed approach combines CNNs with CRFs for polynomial lane modelling. Compared to the conventional methods, this removes the

need for manually designed post-processing and provides modelling of prior information instead of an end-to-end deep learning solution.

## ACKNOWLEDGEMENTS

This publication has emanated from research conducted with the financial support of Science Foundation Ireland under Grant number 18/CRT/6049/. The authors wish to acknowledge the Irish Centre for High-End Computing (ICHEC) for the provision of computational facilities and support.

The authors would like to thank Science Foundation Ireland and Valeo Vision Systems for providing the extraordinary opportunity to perform the research. The author also thanks Dr. Isobel Claire Gormley and Dr. Michael Fop for the continued support while performing the research.

## References

- [1] H. Rashed, E. Mohamed, G. Sistu, V. Ravi Kumar, C. Eising, A. El-Sallab, and S. Yogamani, "Generalized Object Detection on Fisheye Cameras for Autonomous Driving: Dataset, Representations and Baseline," in *Proceedings of the Winter Conference on Applications of Computer Vision*, pp. 2272–2280, 2021. 1
- [2] A. Dahal, V. R. Kumar, S. Yogamani, and C. Eising, "An online learning system for wireless charging alignment using surround-view fisheye cameras," *arXiv preprint arXiv:2105.12763*, 2021. 1
- [3] R. Hazem, E. Mohamed, V. R. K. Sistu, Ganesh and, C. Eising, A. El-Sallab, and S. Yogamani, "FisheyeYOLO: Object Detection on Fisheye Cameras for Autonomous Driving," *Machine Learning for Autonomous Driving NeurIPS 2020 Virtual Workshop*, 2020. 1
- [4] M. Uricar, G. Sistu, H. Rashed, A. Vobecky, V. Ravi Kumar, P. Krizek, F. Burger, and S. Yogamani, "Let's get dirty: Gan based data augmentation for camera lens soiling detection in autonomous driving," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 766–775, 2021. 1
- [5] A. Das, P. Křížek, G. Sistu, F. Bürger, S. Madasamy, M. Uříčář, V. Ravi Kumar, and S. Yogamani, "TiledSoilingNet: Tile-level Soiling Detection on Automotive Surround-view Cameras Using Coverage Metric," in *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, pp. 1–6, IEEE, 2020. 1
- [6] M. Uricar, J. Ulicny, G. Sistu, H. Rashed, P. Krizek, D. Hurych, A. Vobecky, and S. Yogamani, "Desoiling dataset: Restoring soiled areas on automotive fisheye cameras," in *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, pp. 0–0, 2019. 1
- [7] I. Sobh, A. Hamed, V. Ravi Kumar, and S. Yogamani, "Adversarial attacks on multi-task visual perception for autonomous driving," *Journal of Imaging Science and Technology*, vol. 65, no. 6, pp. 60408–1, 2021. 1
- [8] A. Dahal, E. Golab, R. Garlapati, V. Ravi Kumar, and S. Yogamani, "RoadEdgeNet: Road Edge Detection System Using Surround View Camera Images," in *Electronic Imaging*, 2021. 1
- [9] M. M. Dhananjaya, V. R. Kumar, and S. Yogamani, "Weather and light level classification for autonomous driving: Dataset, baseline and active learning," in *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, pp. 2816–2821, 2021. 1
- [10] L. Yahiaoui, M. Uříčář, A. Das, and S. Yogamani, "Let the sunshine in: Sun glare detection on automotive surround-view cameras," *Electronic Imaging*, vol. 2020, no. 16, pp. 80–1, 2020. 1
- [11] V. R. Kumar, S. A. Hiremath, M. Bach, S. Milz, C. Witt, C. Pinard, S. Yogamani, and P. Mäder, "FisheyedistanceNet: Self-supervised scale-aware distance estimation using monocular fisheye camera for autonomous driving," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 574–581, 2020. 1
- [12] R. K. Varun, S. Yogamani, S. Milz, and P. Mäder, "FisheyeDistanceNet++: Self-Supervised Fisheye Distance Estimation with Self-Attention, Robust Loss Function and Camera View Generalization," in *Electronic Imaging*, 2021. 1
- [13] V. Ravi Kumar, M. Klingner, S. Yogamani, S. Milz, T. Fingscheidt, and P. Mader, "Syndistnet: Self-supervised monocular fisheye camera distance estimation synergized with semantic segmentation for autonomous driving," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 61–71, 2021. 1
- [14] R. K. Varun, M. Klingner, S. Yogamani, M. Bach, S. Milz, T. Fingscheidt, and P. Mäder, "SVDistNet: Self-supervised near-field distance estimation on surround view fisheye cameras," *IEEE Transactions on Intelligent Transportation Systems*, 2021. 1
- [15] R. K. Varun, S. Yogamani, M. Bach, C. Witt, S. Milz, and P. Mäder, "UnRectDepthNet: Self-Supervised Monocular Depth Estimation using a Generic Framework for Handling Common Camera Distortion Models," in *IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS*, 2020. 1
- [16] M. Yahiaoui, H. Rashed, L. Mariotti, G. Sistu, I. Clancy, L. Yahiaoui, and S. Yogamani, "FisheyeMODNet: Moving object detection on surround-view cameras for autonomous driving," in *Proceedings of the Irish Machine Vision and Image Processing (IMVIP)*, 2019. 1
- [17] L. Gallagher, V. R. Kumar, S. Yogamani, and J. B. McDonald, "A hybrid sparse-dense monocular slam system for autonomous driving," in *Proc. of ECMR*, pp. 1–8, IEEE, 2021. 1
- [18] V. Ravi Kumar, S. Milz, C. Witt, and S. Yogamani, "Near-field depth estimation using monocular fisheye camera: A semi-supervised learning approach using sparse LiDAR data," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshop*, vol. 7, 2018. 1
- [19] V. R. Kumar, S. Milz, C. Witt, M. Simon, K. Amende, J. Petzold, S. Yogamani, and T. Pech, "Monocular fisheye camera depth estimation using sparse lidar supervision," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pp. 2853–2858, 2018. 1
- [20] K. El Madawi, H. Rashed, A. El Sallab, O. Nasr, H. Kamel, and S. Yogamani, "Rgb and lidar fusion based 3d semantic segmentation for autonomous driving," in *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, pp. 7–12, IEEE, 2019. 1
- [21] V. Ravi Kumar, S. Yogamani, H. Rashed, G. Sitsu, C. Witt, I. Leang, S. Milz, and P. Mäder, "Omnidet: Surround view cameras based multi-task visual perception network for autonomous driving," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 2830–2837, 2021. 1
- [22] G. Sistu, I. Leang, and S. Yogamani, "Real-time joint object detection and semantic segmentation network for automated driving," *NeurIPS 2018 Workshop on Machine Learning on the Phone and other Consumer Devices*, 2019. 1
- [23] P. Maddu, W. Doherty, G. Sistu, I. Leang, M. Uříčář, S. Chennupati, H. Rashed, J. Horgan, C. Hughes, and S. Yogamani, "Fisheyemulti-net: Real-time multi-task learning architecture for surround-view automated parking system," in *Proceedings of the Irish Machine Vision and Image Processing Conference (IMVIP)*, 2019. 1
- [24] S. Houben, S. Abrecht, M. Akila, A. Bär, et al., "Inspect, Under-

- stand, Overcome: A Survey of Practical Methods for AI Safety,” *CoRR*, vol. abs/2104.14235, 2021. 1
- [25] J. Tang, S. Li, and P. Liu, “A review of lane detection methods based on deep learning,” *Pattern Recognition*, vol. 111, p. 107623, 2021. 1
- [26] J. Kim and M. Lee, “Robust lane detection based on convolutional neural network and random sample consensus,” in *International conference on neural information processing*, pp. 454–461, Springer, 2014. 1
- [27] F. Pizzati, M. Alodi, A. Barrera, and F. García, “Lane detection and classification using cascaded cnns,” 2019. 1, 4
- [28] A. Gurghian, T. Koduri, S. V. Bailur, K. J. Carey, and V. N. Murali, “Deeplanes: End-to-end lane position estimation using deep neural networks,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 38–45, 2016. 1
- [29] W. Zhang and T. Mahale, “End to end video segmentation for driving: Lane detection for autonomous car,” *arXiv preprint arXiv:1812.05914*, 2018. 1, 4
- [30] X. Pan, J. Shi, P. Luo, X. Wang, and X. Tang, “Spatial as deep: Spatial cnn for traffic scene understanding,” in *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018. 1, 4
- [31] J. Hur, S.-N. Kang, and S.-W. Seo, “Multi-lane detection in urban driving environments using conditional random fields,” in *2013 IEEE Intelligent Vehicles Symposium (IV)*, pp. 1297–1302, IEEE, 2013. 1, 2, 3
- [32] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, “Encoder-decoder with atrous separable convolution for semantic image segmentation,” in *Proceedings of the European conference on computer vision (ECCV)*, pp. 801–818, 2018. 1
- [33] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, “Rethinking atrous convolution for semantic image segmentation,” *arXiv preprint arXiv:1706.05587*, 2017. 1
- [34] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, “Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 4, pp. 834–848, 2017. 1, 2, 4
- [35] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, “Mobilenets: Efficient convolutional neural networks for mobile vision applications,” 2017. 1
- [36] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, “Pyramid scene parsing network,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2881–2890, 2017. 2
- [37] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016. 2, 4
- [38] M. Contributors, “MMSegmentation: Openmmlab semantic segmentation toolbox and benchmark.” <https://github.com/open-mmlab/mms Segmentation>, 2020. 2
- [39] J. Lafferty, A. McCallum, and F. C. Pereira, “Conditional random fields: Probabilistic models for segmenting and labeling sequence data,” 2001. 2
- [40] H. M. Wallach, “Conditional random fields: An introduction,” *Technical Reports (CIS)*, p. 22, 2004. 2

## Author Biography

*Ganesh Babu is a second-year PhD student at University College Dublin (UCD) as part of the program funded by Science Foundation Ireland (SFI) Centre for Research Training in Foun-*

*datations of Data Science. His research interests are in statistical machine learning, unsupervised learning, neural networks, natural language and image processing, and high dimensional data. He also holds a master’s degree in Data and Computational Science from UCD and a bachelor’s degree in Electrical and Electronics Engineering from Anna University, India.*

*Ganesh Sistu is leading a global team of research engineers in developing large-scale deep learning models for Level-3 autonomous driving vehicles. His area of interest includes multitask deep learning, self-supervised learning, multitask learning, object detection, geometric vision and active learning. He has 20 patents in a.i. related domains and published research in various top tier a.i. conferences.*

*Ramchandra Cheke is a third-year PhD student at the University of Limerick under SFI Centre for Research Training in Foundations of Data Science. His research area includes machine learning, computer vision, deep learning applied to medical imaging. Before that, he completed his bachelors of engineering in electronics and telecommunication from Mumbai University.*

*Senthil Yogamani is an Artificial Intelligence architect and holds a director level technical leader position at Valeo Ireland. He leads the research and design of AI algorithms for various modules of autonomous driving systems. He has over 15 years of experience in computer vision and machine learning including 13 years of experience in industrial automotive systems. He is an author of 100+ publications and 100+ inventions with 70 filed patent families. He serves on the editorial board of various leading IEEE automotive conferences including ITSC and IV and advisory board of various industry consortia including Khronos, Cognitive Vehicles and IS Auto. He is a recipient of the best associate editor award at ITSC 2015 and best paper award at ITST 2012.*