Robert Kaszubski

CSC 578 – Survey Paper

Deep Learning & Autonomous Driving – Lane Detection

Introduction

Autonomous driving has grown increasingly popular in the last several years with several companies and researchers attempting to reach fully autonomy. Full autonomy means that a vehicle can perform any driving function while under any condition without the requirement for the driver to have any input [1]. The vehicle is essentially capable of driving entirely by itself. Tremendous advancements have been made toward this goal thanks to deep learning. Deep learning algorithms are key in making autonomous driving possible. Cars capable of any form of autonomous driving consist of several cameras and sensors that constantly feed data into algorithms that can make decisions on the car's behavior [1]. An important facet of autonomous driving is the ability to detect lanes and road markings. Lanes and road markings must quickly be detected and abided by as cars need to stay in their lanes while driving and obey any markings on the road. Differentiating between different lane markings indicating edge markings, center markings, broken and solid white or yellow lines is key as it informs the algorithm what movement is possible and what restrictions are in place. Autonomous driving systems must likewise identify and abide by turn markers and detect bike lanes. This is only a small piece of the puzzle that is autonomous driving, yet much research and development has been conducted with several solutions proposed and recent improvements made.

Description

In 2019, researchers [2] proposed a method that would combine a convolutional neural network with a recurrent neural network for lane detection. The architecture of this network is in the form of one encoder and one decoder convolutional neural network with a LSTM recurrent neural network sandwiched between. The network takes a series of frames from on-board

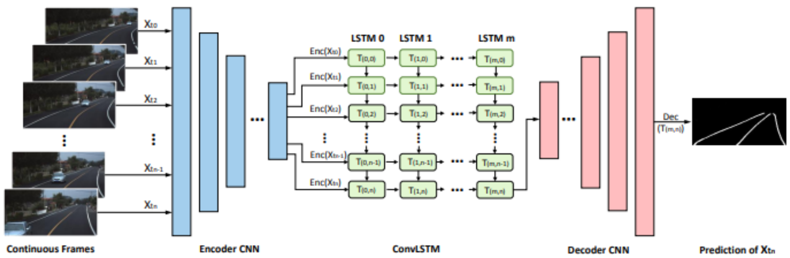

Figure 2 from [2] – CNN + RNN architecture

cameras as input, the encoder CNN processes them and returns feature maps as a time series. The time series feature maps are passed into the LSTM network to retrieve lane-information prediction to be passed into the decoder CNN to output probabilities for lane prediction. Two different CNNs architectures were tested based on SegNet and U-Net which are used for image segmentation, with SegNet functioning better [2]. The network was trained using two datasets of sequential images and was found to outperform more traditional existing models. The inclusion of a RNN and utilization of sequential frames allowed for a more robust model with better accuracy in difficult driving situations such as dim lighting.

A more recent study [3] published at the beginning of 2021 proposes a different solution to lane and road marking detection in difficult driving situations. As with the previous example,

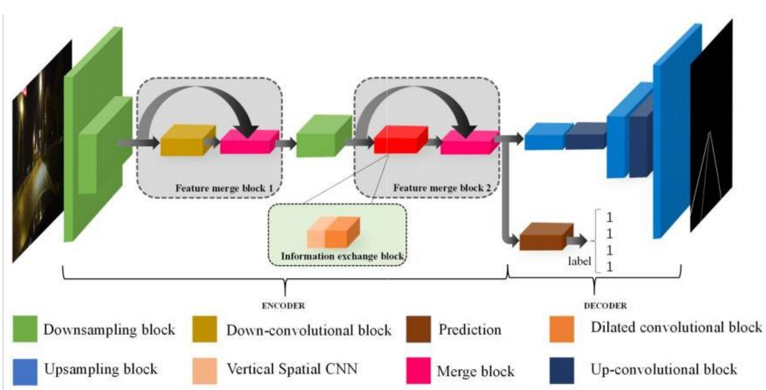

Figure 3 from [3] - Spatial CNN architecture

an encoder and decoder network architecture is proposed however without the inclusion of a RNN. Instead, this study utilizes vertical spatial convolution and contextual driving information. The data from the input image is first down sampled using convolution and pooling, the features are then merged using concatenation and convolution layers, and the dimensionality is reduced. To make the model more robust, vertical spatial convolution is added. Spatial CNNs are proven to be particularly effective at detecting and providing spatial information of objects with distinct properties such as thin, long, and continuous lane lines. The spatial CNN is used concurrently with multiscale dilated convolution to assist with finding features of different sizes by essentially allowing for an expanded kernel. The model was successful in testing being able to create lane maps in different driving conditions, however fell short in accurately predicting curves in the roads.

The most recent proposal [4] published in October 2021, involves the use of inputs from a single monocular camera to generate a lane-level road marking map also known as an RM. Traditionally, lane level maps were generated using expensive and inefficient mobile mapping systems consisting of various sensors such as LiDAR thus a system operating with a single camera is a far more attractive option. The system works by first extracting or segmenting road markings at a pixel level using a semantic segmentation network which the researchers dubbed RMNet. RMNet is built using an encoder-decoder architecture consisting of several convolution and deconvolution layers. The image is first processed in an 18-layer deep convolutional neural network in which features of lanes and road markings are generated then the network branches and separately trains for lane markings and road markings. These

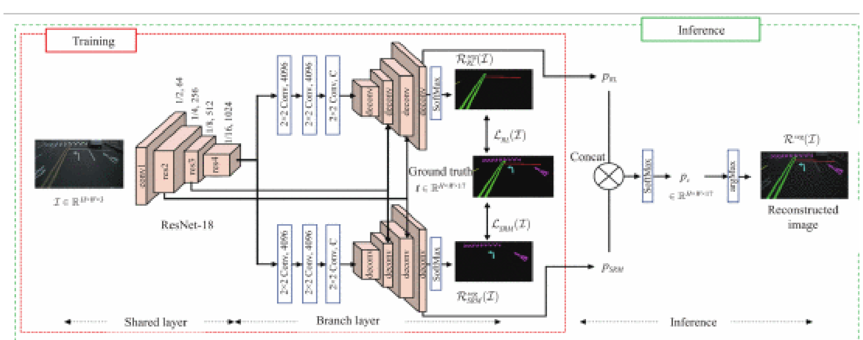

Figure 6 from [4] – RMNET architecture.

features are then used to construct a graph SLAM (Simultaneous localization and Mapping), to create a map of the environment and plot the location of lane and road markings as nodes on the graph in relation of the vehicles position [4]. Unlike the previous model detailed, this not only generated a map of the road lanes but also accurately detected and represented all road markings.

Discussion

Current approaches to lane and road marking detection rely heavily on deep learning encoder decoder neural networks. While the architecture of the networks is somewhat similar, the contents and techniques vary. At the heart of it remains the convolutional neural network as it has for years, there has just been layers added on top of it. The implementation of a RNN in conjunction with a CNN [2], I expect will become far more common particularly as computers with faster processing power become more accessible. Looking at previous sequential frames makes sense for detecting continuous lane lines and could likely be applied to detecting predicting movement of other vehicles. I assume the main hinderance is the lack of processing power. In the spatial convolution model [3], NVIDIA Titan GPUs were utilized, and real time processing was still not possible, I expect this was the case with the other models as well. Lane detection after all is only one part of autonomous driving and in these trials that alone is struggling to operate at real time. The most promising model appears to be the final most recent one which, unlike other models, can simultaneously detect road markings as well. It is likewise capable of adjusting when faced with difficult situations or irregular terrain such as road bumps making it even more impressive.

There has likewise been a heavy departure from previously popular radar-based mapping systems and a larger focus on autonomous driving with input from cameras only. All three lane detection models utilized input from a single or multiple cameras and did not require radar input despite the various problems cameras may have such as light bloom [5]. This is consistent with the direction several car manufactures are heading towards. In 2021, Tesla stopped including radar in their cars and are opting for a camera only autonomous driving system called Tesla Vision [6]. This isn't very different from the models described here and it clearly can work while making the technology accessible in more cars and showcasing the incredible improvements in deep learning allowing them to eschew radar altogether.

Conclusion

Three recently developed models have shown the progress made in lane detection for autonomous driving courtesy of deep learning. These models are all capable of correctly identifying lane markings in various and difficult driving conditions based on inputs from a car's onboard cameras. However, there is still a long road of further improvement before autonomous driving can reach full automation. The results are promising in most situations, but more improvement must be made in difficult driving scenarios such as low visibility, and improvements in hardware and performance that would allow these neural networks to operate in real time.

References:

[1] M. R. Bachute and J. M. Subhedar, “Autonomous Driving Architectures: Insights of machine learning and deep learning algorithms,” *Machine Learning with Applications*, vol. 6, p. 100164, 2021.

[2] Q. Zou, H. Jiang, Q. Dai, Y. Yue, L. Chen, and Q. Wang, “Robust lane detection from continuous driving scenes using Deep Neural Networks,” *IEEE Transactions on Vehicular Technology*, vol. 69, no. 1, pp. 41–54, 2020.

[3] W. Liu, F. Yan, J. Zhang, and T. Deng, “A robust lane detection model using vertical spatial features and contextual driving information,” *Sensors*, vol. 21, no. 3, p. 708, 2021.

[4] W. Jang, J. Hyun, J. An, M. Cho, and E. Kim, “A lane-level road marking map using a monocular camera,” *IEEE/CAA Journal of Automatica Sinica*, vol. 9, no. 1, pp. 187–204, 2022.

[5] D. Babić, D. Babić, M. Fiolić, A. Eichberger, and Z. F. Magosi, “A comparison of Lane marking detection quality and view range between daytime and night-time conditions by Machine Vision,” *Energies*, vol. 14, no. 15, p. 4666, 2021.

[6] “Transitioning to Tesla Vision,” *Tesla*, 13-Jul-2021. [Online]. Available: https://www.tesla.com/support/transitioning-tesla-vision. [Accessed: 22-Nov-2021].