# LaneNet: Real-Time Lane Detection Networks for Autonomous Driving

A Preprint

## Briefly

Authors suggest 2 staged approach: lane edge proposal and lane line localization(each involves an independent deep neural network). Stage one uses a lane edge proposal network for pixel-wise lane edge classification, and the lane line localization network in stage two then detects lane lines based on lane edge proposals. It is stated, that most approached previously where limited on lane patterns, sometimes even required strict assumptions on lane( parallel or approximately straight). In the lane edge proposal stage, the proposal network runs binary classification on every pixel of an input image for generating lane edge proposals, which are served as the input to the lane line localization network in the second stage. lane line localization network, consists of a point feature encoder and a LSTM decoder, localize the lane lines robustly under various scenarios.

The first proposal network, the lane edge proposal network, outputs a "pixel-wise lane edge proposal map."The paper states that this map is a binary classification where each pixel is classified as either a lane edge or not. Further, it specifies that "the value of each pixel indicates the confidence that the network decides this pixel lies on the edge of a lane segment in the input image."This suggests that while the ultimate goal is a binary classification, the raw output is a single-channel map (H x W) where each pixel value represents the probability or confidence of it being a lane edge. The second network, the lane line localization network, takes the coordinates of the lane edges as its input. This is a crucial design choice, as it reduces the input size from a full image (H x W) to a set of N x 2 coordinates (N being the number of lane edge points), contributing to a compact architecture and faster prediction. This coordinate-based input also facilitates a weakly supervised training method.

The encoder takes an IPM image of the front view of a vehicle as the input, and hierarchically extracts the features. The decoder progressively recovers the resolution of the feature map and produce a pixel-wise lane edge proposal map.

## Encoder

The encoder for the lane edge proposal network in LaneNet demonstrates a well-thought-out design that prioritizes both computational efficiency and feature extraction effectiveness. By adopting depthwise separable convolutions and 1x1 pointwise convolutions, it significantly reduces the computational burden compared to traditional convolutional networks, making it suitable for real-time vehicle-based deployment. A key innovation in this encoder is the strategic use of dilated convolutions within its feature extraction blocks. By progressively increasing the dilation rates (1, 2, and 4) across stacked depthwise separable layers, the network effectively expands its receptive field. This allows the encoder to gather broader contextual information from the input image without incurring additional parameters or computational costs. This enhanced contextual understanding is vital for robust lane edge detection, enabling the network to accurately differentiate actual lane markings from other visually similar road features and thereby reducing false positive detections. The design reflects a strong balance between performance and the stringent resource constraints of autonomous driving systems.

## Decoder

The decoder architecture for the lane edge proposal network is a testament to LaneNet's commitment to efficiency without sacrificing accuracy. Recognizing the computational and training challenges associated with deconvolution layers, the authors ingeniously opted for sub-pixel convolution layers, drawing inspiration from their proven efficacy in image super-resolution. This decision is a significant strength, as sub-pixel convolutions offer the crucial benefits of being entirely parameter-free and computationally inexpensive, perfectly aligning with the real-time requirements of vehicle-based systems. Furthermore, the strategic inclusion of skip connections ensures that high-resolution features from the encoder are effectively utilized during the decoding process, which is fundamental for precise lane edge localization. The culmination of these design choices is a highly efficient decoder that effectively recovers feature resolution to produce an accurate pixel-wise lane edge probability map, providing confidence scores for each potential lane edge pixel. This approach represents a clever solution to the common trade-off between model complexity and performance in embedded systems.

## Training

The training methodology for the lane line localization network in LaneNet presents an astute solution to the challenges of directly predicting complex geometric parameters. Instead of attempting to predict quadratic function coefficients, which suffer from problematic scale differences and hinder stable training, the network is ingeniously trained to predict "key values"—the X-coordinates where a lane intersects predefined horizontal lines at the top, middle, and bottom of the image. This transformation standardizes the output magnitudes, significantly improving training stability and convergence. A particularly noteworthy aspect is the introduction and integration of a weakly supervised training approach using a "min-distance loss."This addresses the costly and often imprecise nature of full lane annotations. By calculating the sum of horizontal distances from all detected lane edge points to their closest estimated lane line, this loss pushes the predictions to accurately align with true lane centers, requiring only the count of lanes as ground truth. While pure min-distance training from scratch is prone to poor local minima, the paper proposes effective strategies to overcome this: either by minimizing a weighted combination of $L_2$ and min-distance losses during initial training, or by using min-distance loss for fine-tuning an already optimized model on larger, weakly annotated datasets. This flexible training paradigm is a substantial advantage, enhancing the model's generalization capabilities and mitigating the impact of annotation inaccuracies, ultimately leading to more robust and accurate lane line localization.

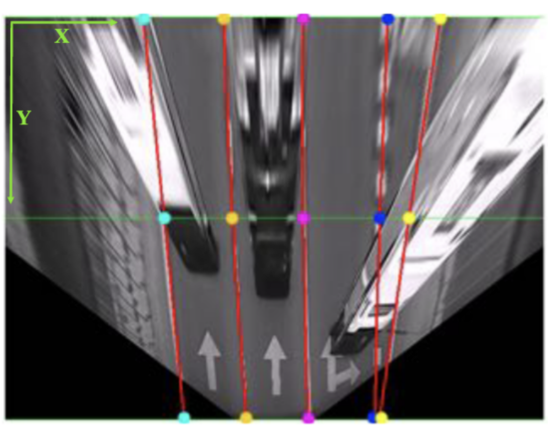
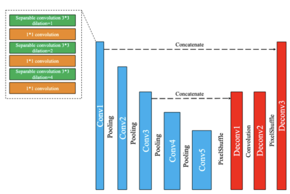


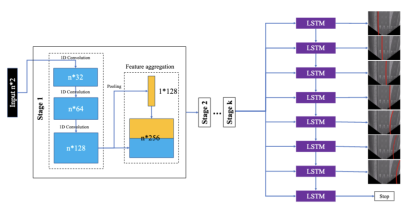

Figure 3. We present an illustration on how we derive the training objectives of the network. The green line indicates the middle line of the image. We use three points to locate a lane which is fitted using a quadratic function.

(a) Original images (b) Lane edge proposals (c) Final results

Figure 4. Illustrative results. The original images along with the corresponding lane edge proposal maps and final detection results are presented. In the final results, different lanes are marked in different colours.




(a) Lane edge proposal network. The layers of encoder and decoder are marked in blue and red, respectively.




(b) Lane line localization network.

Figure 1. The architectures of the lane edge proposal network in the first stage and the lane line localization network in the second stage.