FastDraw: Addressing the Long Tail of Lane Detection by Adapting a Sequential Prediction Network

Jonah Philion
ISEE.AI
jonahphilion@isee.ai

Abstract

The search for predictive models that generalize to the long tail of sensor inputs is the central difficulty when developing data-driven models for autonomous vehicles. In this paper, we use lane detection to study modeling and training techniques that yield better performance on real world test drives. On the modeling side, we introduce a novel fully convolutional model of lane detection that learns to decode lane structures instead of delegating structure inference to post-processing. In contrast to previous works, our convolutional decoder is able to represent an arbitrary number of lanes per image, preserves the polyline representation of lanes without reducing lanes to polynomials, and draws lanes iteratively without requiring the computational and temporal complexity of recurrent neural networks. Because our model includes an estimate of the joint distribution of neighboring pixels belonging to the same lane, our formulation includes a natural and computationally cheap definition of uncertainty. On the training side, we demonstrate a simple yet effective approach to adapt the model to new environments using unsupervised style transfer. By training FastDraw to make predictions of lane structure that are invariant to low-level stylistic differences between images, we achieve strong performance at test time in weather and lighting conditions that deviate substantially from those of the annotated datasets that are publicly available. We quantitatively evaluate our approach on the CVPR 2017 Tusimple lane marking challenge, difficult CULane datasets [8], and a small labeled dataset of our own and achieve competitive accuracy while running at 90 FPS.

1. Introduction

Previous models of lane detection generally follow the following three-step template. First, the likelihood that each pixel is part of a lane is estimated. Second, pixels that clear a certain threshold probability p_{min} of being part of a lane are collected. Lastly, these pixels are clustered, for instance with RANSAC, into individual lanes.

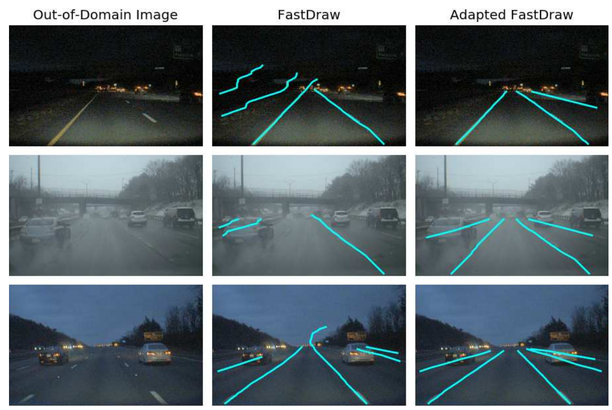

Figure 1. Best viewed in color. We train a novel convolutional lane detection network on a public dataset of labeled sunny California highways. Deploying the model in conditions far from the training set distribution (left) leads to poor performance (middle). Leveraging unsupervised style transfer to train FastDraw to be invariant to low-level texture differences leads to robust lane detection (right).

Because the second and third steps in which road structure is inferred from a point cloud of candidate pixels are in general not differentiable, the performance of models of lane detection that follow this template is limited by the performance of the initial segmentation. We propose a new approach to lane detection in which the network performs the bulk of the decoding, thereby eliminating the need for hyper-parameters in post-processing. Our model "draws" lanes in the sense that the network is trained to predict the local lane shape at each pixel. At test time, we decode the global lane by following the local contours as predicted by the CNN.

A variety of applications benefit from robust lane detection algorithms that can perform in the wild. If the detector is iterative, the detector can be used as an interactive annotation tool which can be used to decrease the cost of building high definition maps [6, 1]. For level 5 systems that depend on high definition maps, online lane detection is a useful lo-

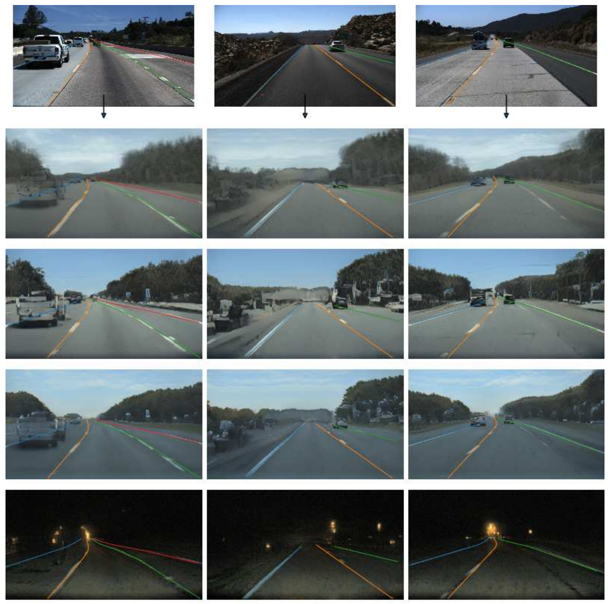

Figure 2. The top row shows three images $\mathbf{x}_i$ from the Tusimple dataset and their annotations. The bottom four rows display samples from $G(\mathbf{x}_i)$ with the adjusted Tusimple annotations overlaid. We use these additional training samples to bias the network towards shape instead of texture [4].

calization signal. Level 2 systems that are not equipped to handle the computational load required of high definition maps depend on models of lane detection equipped with principled methods of determining when to notify the driver that the lane detection is uncertain. In pursuit of solutions for these applications, we identify three characteristics that a lane detection module should possess.

First, the lane detection algorithm must be able to represent any number of lanes of any length. Whereas variability in the number of instances of an object in an image is an aspect of any kind detection problem, variability in the dimensionality of a single instance is a more unique to the lane detection problem; unlike bounding boxes which have a precise encoding of fixed dimensionality, lane segments can be arbitrary length. Solutions that reduce lanes to a constant dimensionality - such as by fitting them with polynomials - lose accuracy on tight curves where accurate lane detection or localization is important for safe driving.

Second, the detection algorithm must run in real-time. Therefore, although there is variability in the number and size of lanes in an image, whatever recursion used to identify and draw these lanes must be fast. Solutions to the variable dimensionality problem that involve recurrent cells [5] or attention [9] are therefore a last resort.

Finally, the detection algorithm must be able to adapt quickly to new scenes. Sensors such as cameras and lidar that are used in self-driving carry with them a long tail in the distribution of their outputs. A lane detection algorithm should be able to adapt to new domains in a scalable way.

We present an approach which addresses these problems and is competitive with other contemporary lane detection algorithms. Our contributions are

- A lane detection model that integrates the decoding step directly into the network. Our network is autoregressive and therefore comes equipped with a natural definition of uncertainty. Because decoding is largely carried out by the convolutional backbone, we are able to optimize the network to run at 90 frames per second on a GTX 1080. The convolutional nature of FastDraw makes it ideal for multi-task learning [7] or as an auxiliary loss [2].

- A simple but effective approach to adapt our model to handle images that are far from the distribution of images for which we have public annotations. Qualititative results are shown in Figure 1 and Figure **??**. While style transfer has been used extensively to adapt the output distribution of simulators to better match reality [3], we use style transfer to adapt the distribution of images from publicly available annotated datasets to better match corner case weather and environmental conditions.

References

[1] D. Acuna, H. Ling, A. Kar, and S. Fidler. Efficient interactive annotation of segmentation datasets with polygon-rnn++. *CoRR*, abs/1803.09693, 2018. 1

[2] M. Bansal, A. Krizhevsky, and A. S. Ogale. Chauffeurnet: Learning to drive by imitating the best and synthesizing the worst. *CoRR*, abs/1812.03079, 2018. 2

[3] A. Dundar, M. Liu, T. Wang, J. Zedlewski, and J. Kautz. Domain stylization: A strong, simple baseline for synthetic to real image domain adaptation. *CoRR*, abs/1807.09384, 2018. 2

[4] R. Geirhos, P. Rubisch, C. Michaelis, M. Bethge, F. A. Wichmann, and W. Brendel. Imagenet-trained cnns are biased towards texture; increasing shape bias improves accuracy and robustness. *CoRR*, abs/1811.12231, 2018. 2

[5] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural Comput.*, 9(8):1735–1780, Nov. 1997. 2

[6] N. Homayounfar, W.-C. Ma, S. Kowshika Lakshmikanth, and R. Urtasun. Hierarchical recurrent attention networks for structured online maps. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018. 1

[7] A. Kendall, Y. Gal, and R. Cipolla. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. *CoRR*, abs/1705.07115, 2017. 2

[8] X. Pan, J. Shi, P. Luo, X. Wang, and X. Tang. Spatial as deep: Spatial CNN for traffic scene understanding. *CoRR*, abs/1712.06080, 2017. 1

[9] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. Attention is all you need. *CoRR*, abs/1706.03762, 2017. 2