Supplementary Material: Hierarchical Recurrent Attention Networks for Structured Online Maps

Namdar Homayounfar, Wei-Chiu Ma, Shrinidhi Kowshika Lakshmikanth, Raquel Urtasun
Uber ATG Toronto
{namdar,weichiu,klshrinidhi,urtasun}@uber.com

1. Videos

We have attached a video that displays how for one example, our model attends to the initial regions of the lane boundaries and traces them sequentially until there are no more left. Furthermore, the video contains the results of our model applied to a couple of sequences from the test set.

2. Model Details

In this section, we present a detailed diagram of the architecture used in our experiments. The basic building blocks are defined below:

- $Conv2D(kernel\_size \times kernel\_size \times out\_channels, stride, padding)$ corresponds to a 2D convolution kernel.
- $BRC(kernel\_size \times kernel\_size \times out\_channels, stride, padding)$ corresponds to batch normalization followed by a $ReLU$ and a $Conv2D(kernel\_size \times kernel\_size \times out\_channels, stride, padding)$.
- $BRUC(kernel\_size \times kernel\_size \times out\_channels, stride, padding)$ corresponds to batch normalization followed by a $ReLU$, followed by a nearest neighbour upsampling and finally a $Conv2D(kernel\_size \times kernel\_size \times out\_channels, stride, padding)$.
- $Residual(kernel\_size \times kernel\_size \times out\_channels, stride, padding)$ corresponds to consecutive $BRC(kernel\_size \times kernel\_size \times out\_channels, stride, padding)$ followed by another $BRC(kernel\_size \times kernel\_size \times out\_channels, 1, 1)$.
- ConvRNN is a vanilla RNN where matrix multiplications are replaced by convolutions. Moreover, we use a hardtanh instead of the usuall tanh non-linearity.
- Similarly, the ConvLSTM module has convolutions instead of matrix multiplications.
- $Crop(heightxwidthxchannels)$ corresponds to cropping a height x width region.

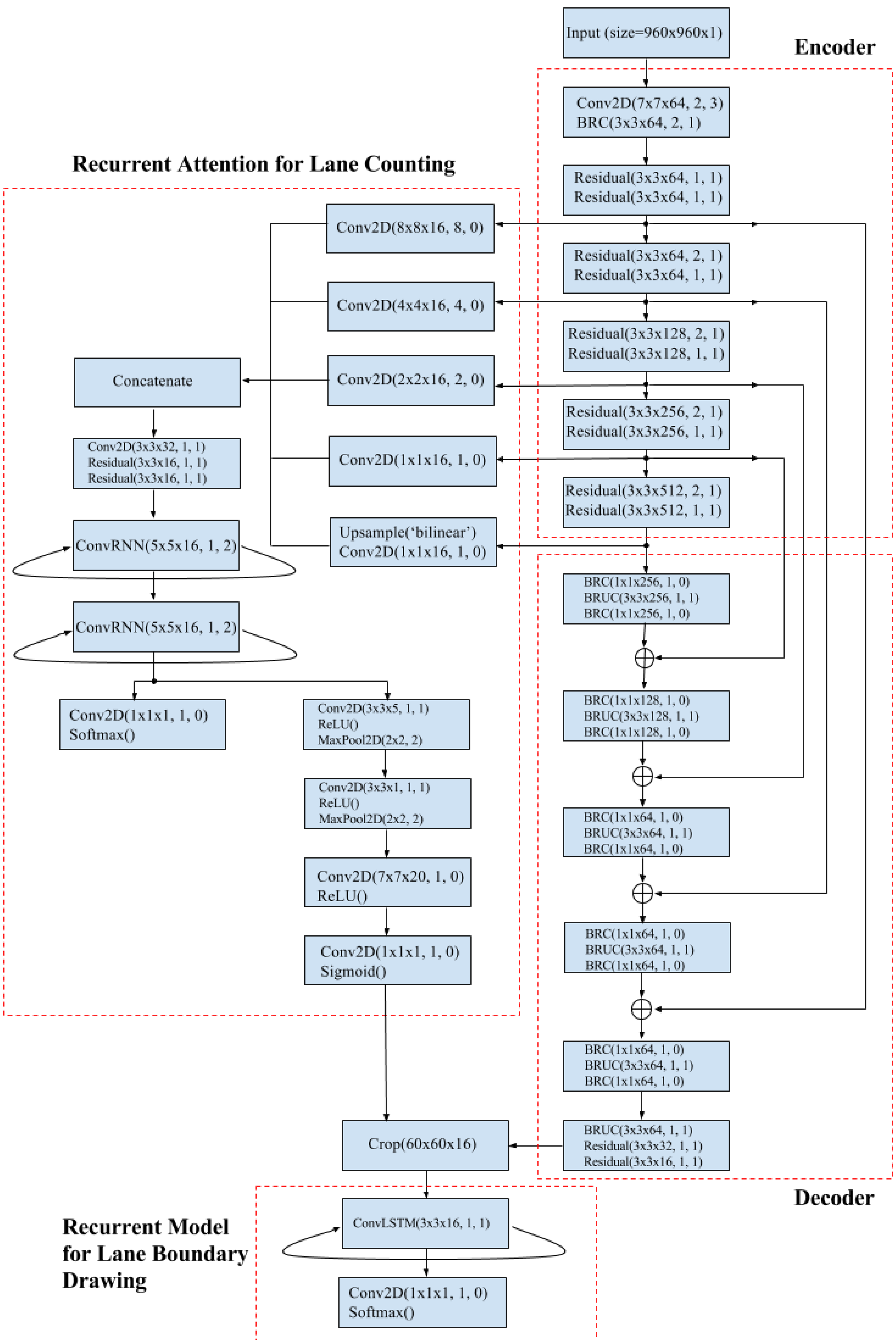

Figure 1. Model Details