Using Segmentation Masks in the ICCV 2019 Learning to Drive Challenge

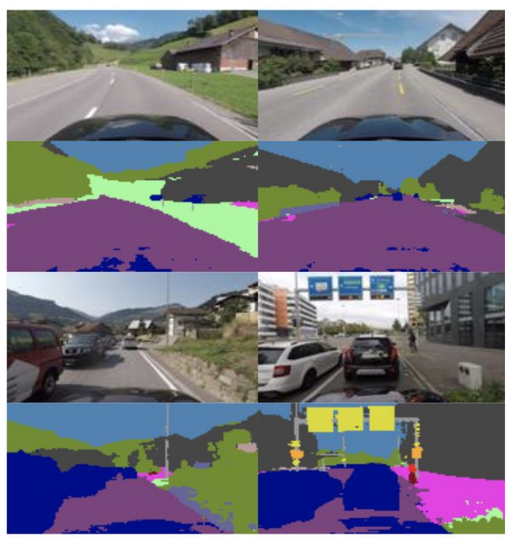
Antonia Lovjer, Minsu Yeom, Benedikt Schifferer, Iddo Drori

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

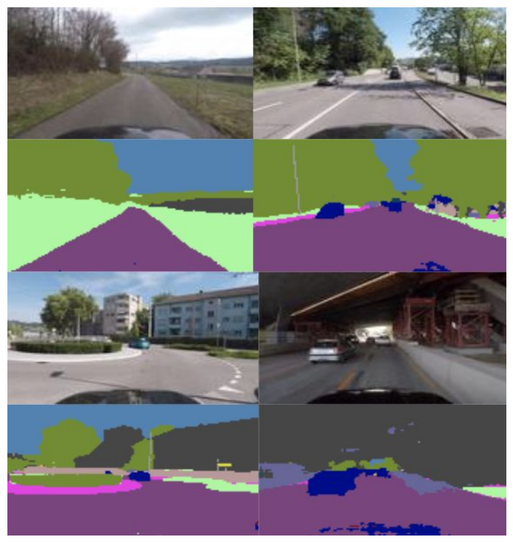
Models and Code: github.com/AntoniaLovjer/learntodrive

Data and Models

In this work we predict vehicle speed and steering angle given camera image frames. Our key contribution is using an external pre-trained neural network for segmentation. We augment the raw images with their segmentation masks and mirror images. We ensemble three diverse neural network models (i) a CNN using a single image and its segmentation mask, (ii) a stacked CNN taking as input a sequence of images and segmentation masks, and (iii) a bidirectional GRU, extracting image features using a pre-trained ResNet34, DenseNet121 and our own CNN single image model. We achieve the second best performance for MSE angle and second best performance overall, to win 2nd place in the ICCV 2019 Learning to Drive challenge.

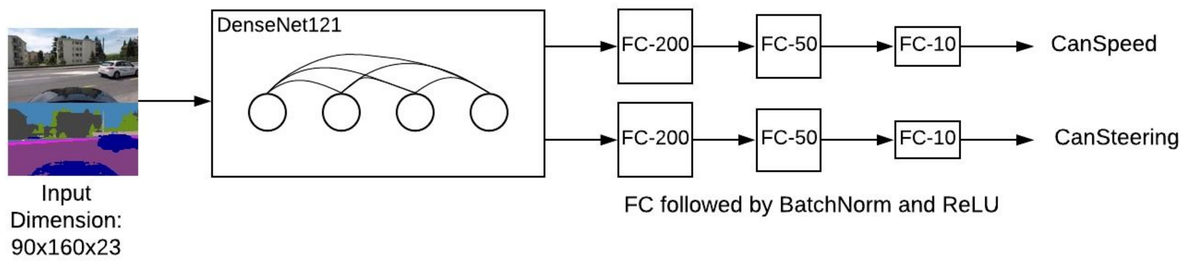


Example training images



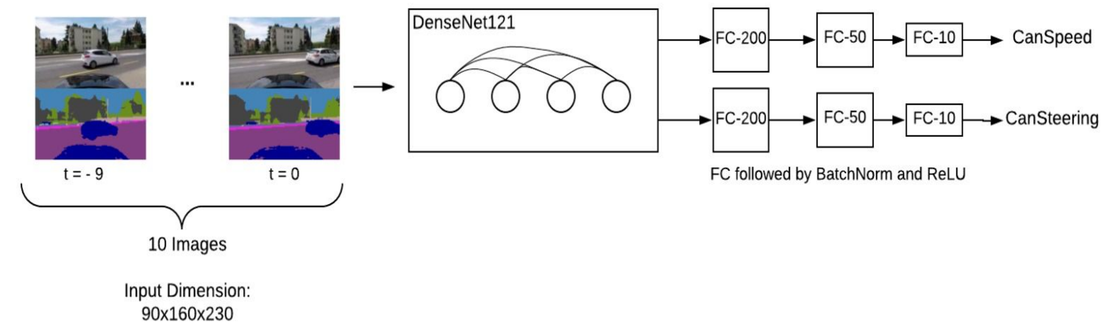
Example test images

Model A_CNN_SINGLE_IMAGE



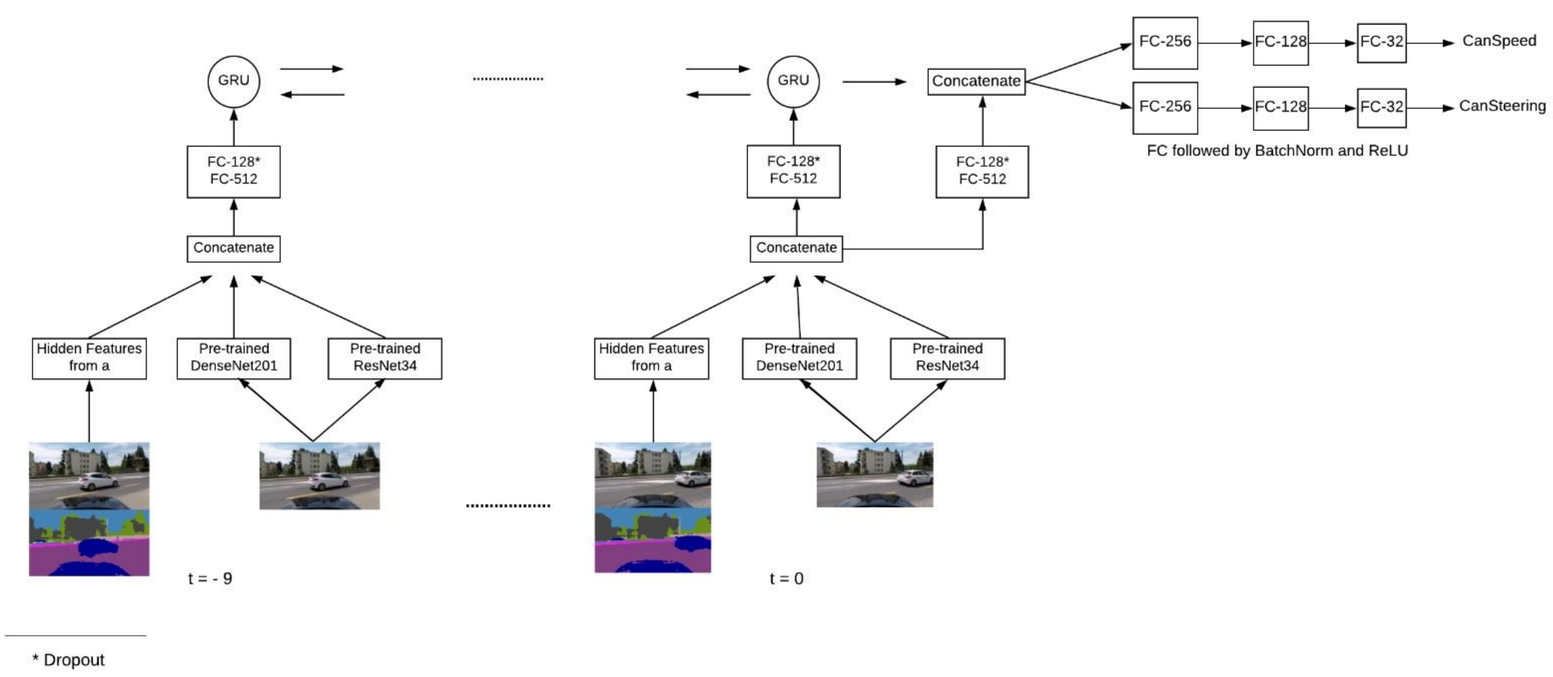
Model A takes as input a single image and its corresponding segmentation mask. The input is passed through a DenseNet121 architecture, then into two fully connected towers, one for predicting the speed, one for the steering angle. The towers contain three dense layers of size 200, 50 and 10. The model used Adam optimizer with an initial learning rate of 0.0003, decaying to 0.0001, 0.00005, and 0.00003 after the first 5, 15, and 20 epochs.

Model B_CNN_STACKED



Model B was trained in an identical way as the Model A. Given the prediction at t=0, the input is the sequence of images from -9 to 0 for every second one image. The inputs are the 10 consecutive images with their hot-n encoded segmentation masks. All images and masks are stacked to a width×height×230 (3 RGB channels + 20 classes)×10.

Model C_BI_GRU



Model C takes as input a sequence of 10 images, and their corresponding segmentation masks, similar to model B, except that they are not concatenated together. Each image in the input sequence is passed individually through a pre-trained ResNet34 model, a pre-trained DenseNet201, and model A. The resulting outputs are concatenated and passed through an intermediate layer which contains two dense layers of size 512 and 128 (with dropout). The output is then passed into a bi-directional GRU cell. The output of the final GRU cell is concatenated with the input to the previous fully connected layer. This is then fed into the two towers for prediction. The model was trained using the Adam optimizer with an initial learning rate of 0.003, which is halved after epochs 20, 30 and 40.

Results

Model	MSE Speed	MSE Angle
A_CNN_SINGLE_IMAGE	7.440	1,140.875
B_CNN_STACKED	7.036	925.926
C_BI_GRU	6.115	1,075.497
Avg:	**5.312**	**901.802**

Our best performing single model for speed was model C_BI GRU with the lowest MSE on the test set at 6.115, while the best performing single model for steering angle was model Model B_CNN_STACKED with MSE loss at 925.926. The results are summarized in the table. Model B performed the best for steering as a result of the combination of data augmentation which included horizontal flipping of images in a sequence, and the full view of the sequence to the input of the model.

The best performing model overall is B_CNN_STACKED, which achieved the lowest MSE for steering and was second for speed. The individual model that had the best performance for speed was C_BI_GRU. The average of the predictions generated by models B and C provided a significant decrease in the MSE of both speed and steering angle. This combination was our best submission to the competition which awarded us second place in steering angle prediction and second place overall.

Acknowledgements: We would like to thank the Columbia University students of the Fall 2019 Deep Learning class for their participation in the challenge. Specifically, we would like to thank Xiren Zhou, Fei Zheng, Xiaoxi Zhao, Yiyang Zeng, Albert Song, Kevin Wong, and Jiali Sun for their participation.