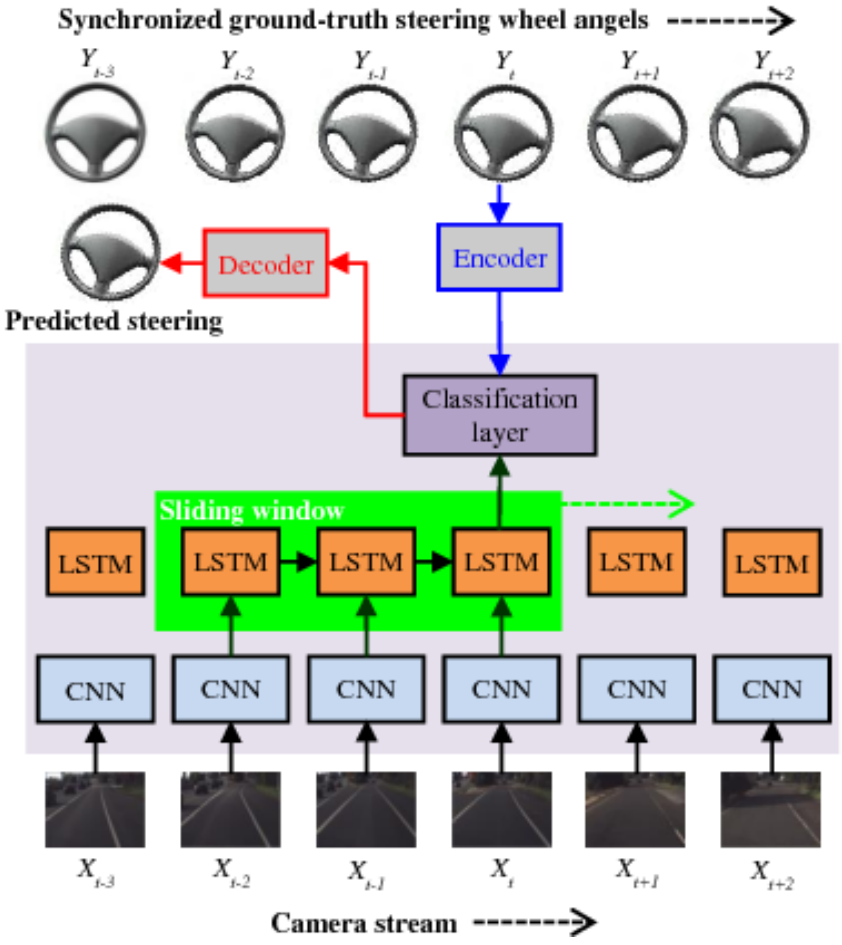
USE OF LSTM RECURRENT NEURAL NETWORK FOR STEERING IN AUTONOMOUS VEHICLES

Convolutional Neural Network allows us to extract features based on visual dependencies thereby ignoring the temporal relationship between the frames. But RNN such as **LSTM allows us to learn temporal dependencies along with visual features** of driving allowing us to automate steering in automobiles.



Loss Function : **RMSE(Root Mean Square Error)** between the predicted and ground-truth steering angle can be used to give high penalties to large errors since large values of errors are undesirable.

Optimizer : **Gradient Descent or Adam optimization.**

The model requires a **front facing camera** and a way to take the **steering angle measurement.** CNN takes the camera images frame by frame and extracts the useful visual features. These features are then given as input to the respective LSTM layers. LSTM then learns the temporal relations and gives directly or indirectly the steering angle as output. LSTM can be trained to directly output steering angle or the output of LSTM can be given as input to regression or classification layers to output steering angle. The predicted steering is transmitted to the steering wheel by some actuators.

The sliding window technique can be used to fix the number of LSTM layers to extract temporal information. Unlike a typical RNN training based on a fixed batch of inputs, the sliding window allows the use of the same feature for learning steering angles at different times(stack of LSTM layers).

While training the model the true steering angles are recorded by the human driver and are given to the LSTM layers as true values. In the above figure the red path from the LSTM classification layer is for prediction whereas the blue path is for training purposes wherein the ground truth steering angles are given to the model.

OVERTAKING PROBLEM

To take into account the overtaking problem of vehicles behind, we can have side cameras and a back camera along with a front facing camera. Then the input to the model will be a stack of synchronized images from all cameras frame by frame instead of only one image frame from the front facing camera. The rest of the model remains the same with only difference in the input. This will be a direct end-to-end learning approach for steering in automobiles.

END