Image Prediction for Lane Following Assist using Convolutional Neural Network-based U-Net

Byung Chan Choi[1,2], Jaerock Kwon[3], and Haewoon Nam[1]
[1]Division of Electrical Engineering, Hanyang University, Ansan, Korea
[2]RF Seeker R&D, LIG Nex1, Yongin, Korea
[3]College of Engineering and Computer Science, University of Michigan-Dearborn, Dearborn, MI, USA

***Abstract*—Current autonomous driving systems compute steering and throttle control commands by running perception-decision-action pipeline at high frequency. Although human drivers cannot react or control the vehicles as quickly as the autonomous driving softwares, most drivers control their vehicles to stay in lane unless they intend to break away from the lane. According to forward internal model theory, human can choose an optimal action for the best outcome by internally simulating all the possible consequences of various actions. This means that humans drivers choose the optimal motor commands for lane following based on their internal simulation of near-future lane changes. This paper proposes a convolutional neural network-based U-Net as a state estimator for forward internal model-based lane following assist. This state estimator can predict the lane image of near-future based on current lane image and driving status data, such as speed and steering angle. This paper also explains how time difference between current lane image and the next one to be predicted will affect the training and prediction output of the estimator.**

***Index Terms*—Lane Following Assist, Deep Learning, Convolutional Neural Network, Internal Model**

I. Introduction

Lane Following Assist (LFA) is one of the basic functions that make the autonomous vehicle detect and follow the lanes on the road. Performance of LFA is determined by the vehicle's reaction time for wheel control. An autonomous vehicle needs time to process its sensor data and run lane detection algorithms before applying wheel controls. As a result, many automotive companies try to improve the performance of LFA by minimizing processing time for lane detection. However, this approach cannot make processing time for lane detection into zero. There will always be a delay between lane detection and wheel control.

Similar to LFA, human drivers also suffer the delay between lane detection and wheel control. However, although human drivers cannot control wheels as quickly or precisely as the autonomous vehicles, they can follow the lanes without any troubles. This is because human drivers use a different approach for lane following. Human drivers internally simulate how the lane will change in the future based on current driving status and choose the optimal action to achieve the best outcome for following the lanes on the road. One of the theoretical frameworks for a human to choose actions based on *internal simulation* is *forward internal model principle* of the cerebellum [1] [2] [3]. One key function of the cerebellum is to predict the sensory consequences of the motor outputs by comparing its prediction to the sensory feedback and minimizing its error [4].

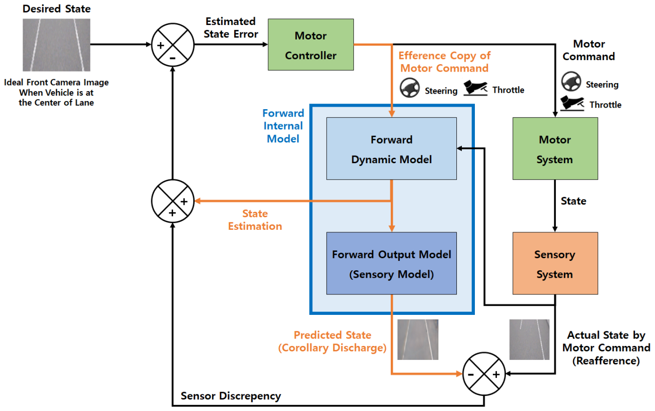
Desired State
Ideal Front Camera Image When Vehicle is at the Center of Lane
+
-
Estimated State Error
Motor Controller
Efference Copy of Motor Command
Steering
Throttle
Motor Command
Steering
Throttle
Forward Internal Model
Forward Dynamic Model
Motor System
State
+
+
State Estimation
Forward Output Model (Sensory Model)
Sensory System
Predicted State (Corollary Discharge)
Actual State by Motor Command (Reafference)
-
+
Sensor Discrepency

Fig. 1. Smith Predictor for Forward Internal Model-based LFA

Human driver's forward internal model compensates for the latency between lane detection and wheel control. This internal model-based control can be implemented as Smith predictor, a feedback control system with time-delay compensation scheme [5] [6]. Fig 1 is the diagram of vision-based LFA system with forward internal model and Smith predictor design. The system receives the desired state for its task. For LFA, the desired state is the ideal front camera image when the vehicle is driving at the center of lane. When a driver applies throttle and steering controls, the system feeds these control commands to forward internal model and motor system. Motor system changes the vehicle's speed and orientation based on the driver's motor commands. Sensory system perceives the changes in state by motor commands and produces the front camera image output of changed state. Forward internal model uses an efference copy of motor commands and produces the prediction of next lane image state. Sensor discrepancy between the next lane image prediction and the front camera image of changed state will be added with state estimation results from forward internal model in order to adjust the estimation. The actions that produce the lowest error between the desired state and the state estimation by current motor commands will be selected as next motor commands.

Inspired by this idea, this paper proposes a deep neural network-based state estimator for forward internal model-

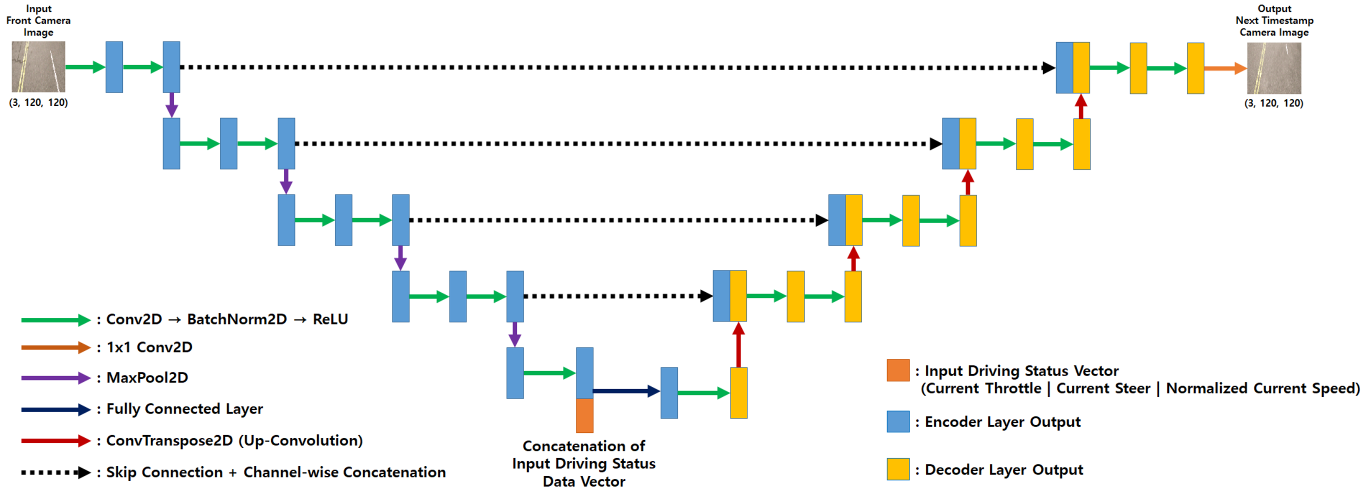

Fig. 2. CNN-based U-Net for State Estimation of Internal Model

based LFA. It utilizes Convolutional Neural Network (CNN)-based U-Net from [7] to predict next lane image from current lane image and driving status data. U-Net's skip architecture provides feature reusability that creates thorough gradient update flow and prevent gradient vanishing during the training. This approach can be later implemented in internal simulation of forward internal model-based LFA in order to compensate for the latency between lane detection and wheel control.

II. Backgrounds

A. Internal Model

In neuroscience, the internal model is a cognitive interpretation of organism motion planning. It claims that an organic agent, such as human, can anticipate the consequences of its actions without actually committing them [8] [9]. According to Wolpert et al. [10], there are two types of motion planning models : forward model and inverse model. In forward model, the agent predicts the sensory outcome of an action based on current state information and motion commands. With the proper forward model, it internally simulates actions and matches the outcome with the closest target outcome. In inverse model, the agent guesses which action has led to the current state.

The forward internal model is an intuitive interpretation of how an organic agent chooses its actions for its task. McNamee and Wolpert show that forward model consists of four stages [8]. First, in perception stage, an agent receives sensor input by monitoring the environment and its current action status. Sensory input will contain noise, because the agent cannot observe non-visible environment parameters, such as speed and spin. Also, there is noise in the agent's sensory system. Second, in simulation stage, the agent predicts how the environment state will change in near future. Third, in motion planning stage, the agent produces all the possible outcomes by its given action options. Among all the state and action predictions, the agent chooses the action that can lead to the outcome closest to its target outcome. Fourth, in optimal feedback control stage, the agent applies motor commands for its chosen action. The agent uses optimal feedback controller when applying actions in order to adjust the motor commands according to current sensory feedback and environment state.

B. Forward Internal Model for Autonomous Driving

According to Plebe et al, current autonomous driving algorithm loop is strictly divided into perception-decision-action [11]. Deep neural network is often applied in perception stage as a single module, because it is well suited for its generalization in object detection and classification task. However, Plebe et al. suggest that if this rigid division between perception, decision, and action can be collapsed, deep neural network itself can be implemented as the complete perception-decision-action loop [11]. Inspired by neuroscience, the entire autonomous driving algorithm loop can be re-defined with three pathways. First pathway, dorsal stream, is the sensor data tensor flow through the entire deep neural network. It is based on the visual pathway of the primate brain [12]. Second pathway, cerebellum loop, also known internal simulation, represents the network's capability to internally predict the consequences of various actions. Third pathway, action selection loop, is the network's capability to select the optimal action decision based on its internal simulation results. In order to implement internal model for autonomous driving, the system needs deep neural networks for internal simulation and action decision selection.

III. Proposed Method

Forward internal model requires an internal simulation mechanism that can predict near-future state based on current sensory state input and system action output. Therfore, in order to integrate forward internal model into vision-based LFA system, it requires a state estimator that can predict the lane image of near-future based on current lane image and driving status data. This paper utilizes CNN-based U-Net as a state estimator for forward internal model-based LFA.

A. *CNN-based U-Net for State Estimation of Internal Model*

U-Net is a convolutional networks for biomedical image segmentation [7]. In U-Net, Ronneberger et al. implement skip connections between matching encoding layers and decoding layers [7]. These skip connections can prevent gradient vanishing by allowing gradient information to be maintained all the through the layers. In next lane image prediction, it is important that the network learns to use the features from current lane image and driving status data to produce next lane image. U-Net's skip connections can achieve this by creating a thorough gradient information flow among the layers.

CNN-based U-Net is trained to produce a lane image of next timestamp from a current lane image and driving status data vector. Current lane image and driving status data are used as input to the network. Current lane image is processed into latent feature vectors by CNN-based encoders. Driving status data is appended in the encoder's latent feature vector. In order to produce the output image with the same shape as input image, the appended latent feature vector is reshaped into 1x1000 shape by fully connected layer. The output of fully connected layer is then processed into an image through CNN-based decoder. During the decoding process, output vectors of matching encoders will be appended into the input of decoders in order to establish feature resusability and maintain gradient flow through skip connections. Along with vision-based steering control, this deep learning-based next lane image prediction can play as a state estimation pathway for forward internal model-based LFA.

B. *Effect of Time Difference for Lane Prediction*

Longer the time difference between current and next timestamps, there will be greater displacement between current and next lane images. In order to determine the capability of next lane image prediction, it is necessary to figure out how much displacement the neural network can be trained to handle. In this paper, we tested the network under four timestamp differences. We observed how the length of timestamp difference affects the neural network's training for next lane image prediction.

IV. Experiment

A. *Dataset Collection and Preparation using CARLA*

This paper uses CARLA, open-source autonomous driving simulator, in order to collect an extensive amount of lane image and driving status data [13]. We added a dataset recording function on top of the autnomous driving example provided by CARLA. This function records lane images from the vehicle's front camera. It also collects driving status data, wheel steering, throttle, and speed, with the matching simulation timestamp. Dataset was collected from four maps, Default Town, Town01, Town 06, and Town 04, with different weather settings. Dataset was collected in the driving environment and weather conditions, where the lanes are clearly visible.

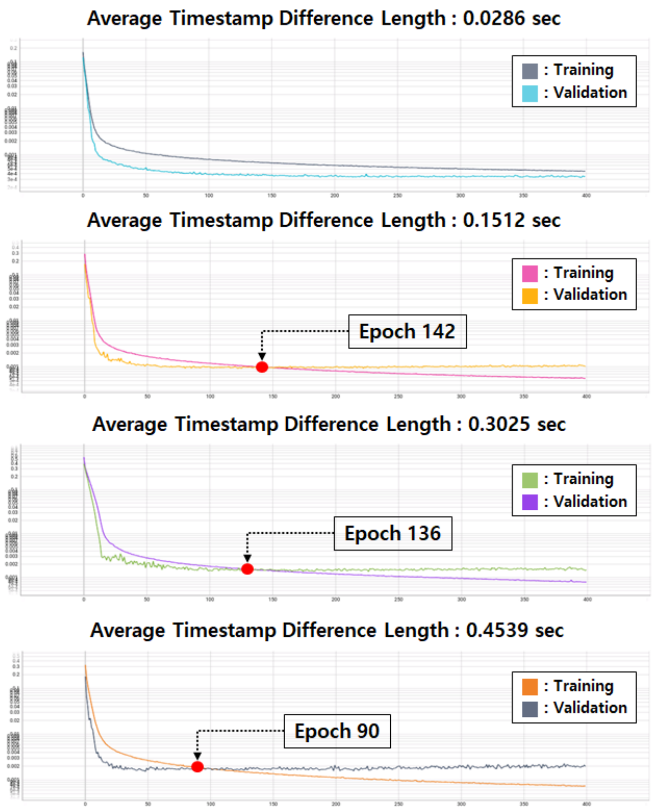

Fig. 3. Loss Graph Comparison under Different Timestamp Differences

For stable training, it is imperative that both input data and output data are normalized with same range. Original lane images are scaled between 0 and 1 by being multiplied by 1/255. Original wheel steering and throttle are already normalized between 0 and 1 from CARLA. Original speed data is scaled between 0 and 1 by being multiplied by 1/100. This assumes that the maximum driving speed is 100km/h. Normalizing input lane image data with the same range as driving status data can prevent unstable gradient backpropagation during the training process.

B. *Training and Experiment Setup*

CNN-based U-Net proposed in Fig 2 is implemented using PyTorch 1.9.0. It is trained for 400 epochs on Nvidia RTX 3090. It is trained by Adam optimizer with learning rate of 1e-5. Mean Squared Error (MSE) is used as the loss function between target next lane image and the network's prediction output image.

C. *Training Result Comparison*

Fig 3 is the compilation of training and validation loss function graphs of CNN-based U-Net trained with four different timestamp difference lengths. It is log-scaled on y-axis in order to clearly show how a loss function graphs changes in different timestamp length. Fig 3 shows that the network trained to predict the image with longer timestamp difference suffers faster overfitting. This is because under longer timestamp

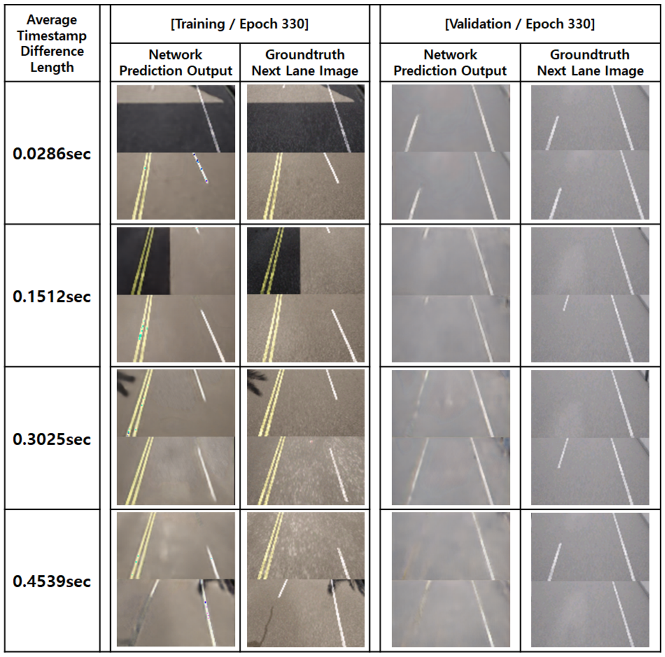
Average Timestamp Difference Length
[Training / Epoch 330]
Network Prediction Output
Groundtruth Next Lane Image
[Validation / Epoch 330]
Network Prediction Output
Groundtruth Next Lane Image
0.0286sec
0.1512sec
0.3025sec
0.4539sec

Fig. 4. Prediction Output Image Comparison

difference, there will be greater displacement between between current and next lane images. As a result, a larger portion of input lane image will be considered unrelated to next lane image groundtruth.

Fig 4 shows the prediction output image of CNN-based U-Net with different timestamp differences. In Fig 4, the network trained with longer timestamp difference produces more blurry prediction output image. Based on the analysis from Fig 3 that a larger part of current lane image input is considered unrelated to next lane image prediction in longer timestamp difference condition, the level of feature reuse will decrease. This results in more blurriness between current lane image input and next lane prediction output. However, even in longer timestamp difference, CNN-based U-Net's output image still contains lanes that can used for vision-based steering. This result shows that CNN-based U-Net can be trained to estimate next lane image and implemented as a state estimator for forward internal model-based LFA.

V. Conclusion

This paper presents CNN-based U-Net as a state estimator for forward internal model-based LFA system. It shows how timestamp difference length between current and next lane image affects training characteristics and outuput prediction quality. Although longer timestamp difference length results in overfitting by increasing the displacement between current and next lane image, the lanes in prediction output image produced by the network trained with longer timestamp difference are visible enough for vision-based LFA.

CNN-based U-Net can be later integrated into the forward internal model-based LFA system from Fig 1 as a state estimator in forward internal model. This implementation can provide a deep learning-based LFA pipeline that can mimic human driver behaviors. Training the network with additional dataset with more diverse weather conditions, illumination changes, and traffic elements can further improve its performance as a state estimator.

Acknowledgment

This research was supported by the MSIT(Ministry of Science, ICT), Korea, under the High-Potential Individuals Global Training Program(2020-0-01513) supervised by the IITP(Institute for Information and Communications Technology Planning & Evaluation) and in part by the National Research Foundation of Korea (NRF) grant funded by the Korean Government (MSIT) (No. 2019R1A2C109009612).

References

[1] M. Kawato, "Internal models for motor control and trajectory planning," *Current Opinion in Neurobiology*, vol. 9, no. 6, pp. 718–727, Dec. 1999.

[2] J. Stein, "Cerebellar forward models to control movement," *The Journal of Physiology*, vol. 587, no. Pt 2, p. 299, Jan. 2009. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2670044/

[3] Q. Welniarz, Y. Worbe, and C. Gallea, "The forward model: A unifying theory for the role of the cerebellum in motor control and sense of agency," *Frontiers in Systems Neuroscience*, vol. 15, p. 22, 2021. [Online]. Available: https://www.frontiersin.org/article/10.3389/fnsys.2021.644059

[4] L. S. Popa and T. J. Ebner, "Cerebellum, Predictions and Errors," *Frontiers in Cellular Neuroscience*, vol. 12, p. 524, 2019. [Online]. Available: https://www.frontiersin.org/article/10.3389/fncel.2018.00524

[5] O. J. M. Smith, "A controller to overcome dead time," in *ISA Journal*, vol. 6, 1959, pp. 28–33.

[6] N. Abe and K. Yamanaka, "Smith predictor control and internal model control - a tutorial," in *SICE 2003 Annual Conference (IEEE Cat. No.03TH8734)*, vol. 2, 2003, pp. 1383–1387 Vol.2.

[7] O. Ronneberger, P.Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, ser. LNCS, vol. 9351. Springer, 2015, pp. 234–241, (available on arXiv:1505.04597 [cs.CV]). [Online]. Available: http://lmb.informatik.uni-freiburg.de/Publications/2015/RFB15a

[8] D. McNamee and D. M. Wolpert, "Internal models in biological control," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 2, no. 1, pp. 339–364, 2019. [Online]. Available: https://doi.org/10.1146/annurev-control-060117-105206

[9] S. J. Blakemore, S. J. Goodbody, and D. M. Wolpert, "Predicting the consequences of our own actions: The role of sensorimotor context estimation," *Journal of Neuroscience*, vol. 18, no. 18, pp. 7511–7518, 1998. [Online]. Available: https://www.jneurosci.org/content/18/18/7511

[10] D. M. Wolpert, Z. Ghahramani, and M. I. Jordan, "An internal model for sensorimotor integration," *Science*, vol. 269, no. 5232, pp. 1880–1882, 1995. [Online]. Available: https://www.science.org/doi/abs/10.1126/science.7569931

[11] A. Plebe, M. Da Lio, and D. Bortoluzzi, "On reliable neural network sensorimotor control in autonomous vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 2, pp. 711–722, 2020.

[12] B. R. Sheth and R. Young, "Two Visual Pathways in Primates Based on Sampling of Space: Exploitation and Exploration of Visual Information," *Frontiers in Integrative Neuroscience*, vol. 10, p. 37, 2016. [Online]. Available: https://www.frontiersin.org/article/10.3389/fnint.2016.00037

[13] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "CARLA: An open urban driving simulator," in *Proceedings of the 1st Annual Conference on Robot Learning*, 2017, pp. 1–16.