Automated Valet Parking using Optical Flow

Harsh Sharma IIT2021131, Parmar Milan IIT2021121, Aryan Pandey IIT2021153, Saransh Yadav IIT2021162, Kushagra Jain IIT2021191,

Indian Institute of Information Technology Allahabad
India

Abstract: This model utilizes both RGB information and optical flow as inputs. The key idea is that by incorporating both types of information, the model aims to enhance the accuracy of steering angle predictions compared to a single-frame-based end-to-end steering prediction method.

Index Terms: self-driving cars, CNN model, temporal information

1. Introduction

Self-driving cars have the potential to revolutionize the transportation industry by reducing accidents and increasing efficiency. However, developing a reliable and accurate steering behavior imitation system is a challenging task. In this project, we propose a CNN model that significantly improves steering angle prediction accuracy by incorporating temporal information. With its ability of learning deep hierarchical representations and features with high non-linearity, deep convolutional neural network (CNN) enables the system to learn the steering behavior based on the raw pixels obtained by the vision system, which is also known as end to end driving. The method of establishing a mapping relations between steering angle and a single frame by using CNN has gained a remarkable success. The problem at hand is precisely articulated as follows: How can a Convolutional Neural Network (CNN) attain a comprehensive understanding of steering behavior for precise valet parking, leveraging temporal information extracted from the video stream? In the pursuit of this research objective, we propose a sophisticated two-stream framework meticulously designed to process inputs from both RGB data and temporal information. The envisioned framework aims to produce accurate steering outputs tailored to navigate the intricacies of valet parking. Notably, the temporal information crucial to this endeavor encompasses the integration of optical flow, capturing the nuanced motion dynamics essential for effective autonomous navigation in such constrained environments.

2. Related Work

The field of autonomous driving systems has seen significant advancements, particularly with the integration of Convolutional Neural Networks (CNNs) for direct steering command regression based on raw images[1] . This paradigm involves training a CNN regressor with spatial cues extracted from forward-facing camera footage, incorporating vehicle cinematic measurements. The success of this approach lies in its efficiency in spatial cue extraction[7], although it operates on individual frames without considering temporal information. Two-Stream Convolutional Networks, introduced by Symonyan and Zisserman (2014), have emerged as a preferred method for action recognition in videos. These networks excel in capturing temporal dynamics through the optical flow between consecutive frames.[1] However, their application in regression tasks, especially in autonomous steering, remains less explored. This paper aims to bridge this gap by demonstrating the effectiveness of Two-Stream Convolutional Networks in leveraging temporal cues for end-to-end learning. Pioneering work in this field has introduced a deep convolutional network architecture designed explicitly for video action recognition. This architecture integrates two primary streams: the spatial stream and the temporal stream. The spatial stream captures the appearance of individual frames, providing crucial information about scenes and depicted objects. Simultaneously, the temporal stream focuses on motion across frames, conveying the dynamic aspects of both the observer (camera) and the objects within the video.

3. Methodology

3.1. Two-Stream Model

We employ a two-stream CNN model to capture both spatial and temporal information for replicating human steering behavior. The model takes inputs of three consecutive RGB frames and one or two optical flow frames. RGB frames undergo processing in the spatial stream, while the optical flow frames are processed in the temporal stream. The outputs are concatenated and fed into fully connected layers to predict the steering angle.

3.2. Optical Flow - Farneback Method

The Farneback algorithm is utilized to compute optical flow between consecutive frames. Flow vectors obtained from this method serve as inputs to the temporal stream of the two-stream CNN.

3.3. Performance Metrics

We evaluate model accuracy using mean square error (MSE) and root mean square error (RMSE). MSE measures the average squared difference between predicted and ground truth steering angles, while RMSE represents the square root of MSE. To quantify the disparity between predicted and actual angles, we utilize the mean square error (MSE) as the chosen loss function:

L = $\sum_{i=1}^{N}(\theta_i - \bar{\theta}_i)^2 \frac{}{N}$

Here, N represents the batch size, θ_i and $\bar{\theta}_i$ denote the predicted angle and ground truth, respectively.

4. Dataset

We use a public shared dataset created by USC researcher Sully Chen and used it to train and test our two-stream CNN model. There are two subsets in this dataset: 2017 dataset and 2018 dataset. Both of them are made up with a number of image files and time-stamped car steering angle logs. The image data was recorded by a camera mounted on the Honda Civic front windshield at 20 frames per second. 2017 dataset (Sully Chen 1) contains 45567 frames and corresponding steering angle labels. The dataset records a trajectory of approximately 4km around the Rolling Hills in LA, USA. 2018 dataset (Sully Chen 2) contains 63825 frames and corresponding steering angle label as well as timestamps. This dataset records a trajectory of approximately 6km along the Palos Verdes Dr in LA, USA.

In the dataset, we incorporated Sully Chen 1 for the training of our model and we have taken 10,000 inital frames from Sully Chen 2 for testing our model.

4.1. Data Pre-processing

We use the OpenCV's inbuild Farneback method to compute the optical flow. The 2 channel results is then encoded into HSV space to form a single 3 channels image. Direction of the optical flow is encoded by Hue and the magnitute is encoded by the V value. Before feeding the images into the model, we resize the images as 80 × 320 to expand the road and shrink the background.

5. Results

5.1. Formula for Accuracy within Tolerance

The accuracy within a tolerance range is given by:

$$\text{Accuracy} = \frac{\sum_{i=1}^{n} \mathrm{I}\left(|\text{actual}_i - \text{pred}_i| \leq \text{tolerance}\right)}{100}$$

where n is the total number of predictions, and $\mathrm{I}(condition)$ is the indicator function that returns 1 if the condition is true and 0 otherwise.

5.2. Formula for Mean Square Error

The mean square error (MSE) is given by:

$$\text{MSE} = \frac{\sum_{i=1}^{N} (\theta_i - \bar{\theta}_i)^2}{N}$$

where N is the total number of observations, θ_i represents the actual values, and $\bar{\theta}_i$ represents the predicted values.

Threshold	Accuracy (%)
5	74.42
7	88.18
10	95.20
15	97.90
20	98.78
30	99.50
40	99.80

TABLE I
ACCURACY VALUES AT DIFFERENT THRESHOLDS

Paper	Root Mean Square Error
Our Paper	5.88
Fernández, Nelson. “Two-stream convolutional networks for end-to-end learning of self-driving cars.” ArXiv abs/1811.05785 (2018)	12.52
[1710.03804] End-to-End Deep Learning for Steering Autonomous Vehicles Considering Temporal Dependencies (arxiv.org)	16.01

TABLE II
COMPARISION BETWEEN DIFFERENT PAPERS

6. Discussion

6.1. CNN: A challenging neural network

The analysis of the activation map of the Convolutional Neural Network (CNN) reveals a notable emphasis on the edges of the road, indicating their significant importance in the prediction process. However, it is observed that the CNN also marginally activates some background areas, such as the edges near trees and the sky. These activations in non-road regions may not contribute positively to steering prediction and could potentially introduce noise.

One plausible interpretation is that redundant edge areas may interfere with the accuracy of predictions. This interference becomes particularly evident when the image contains a relatively large amount of edge information. The inclusion of non-informative background areas could potentially lead to suboptimal steering predictions.

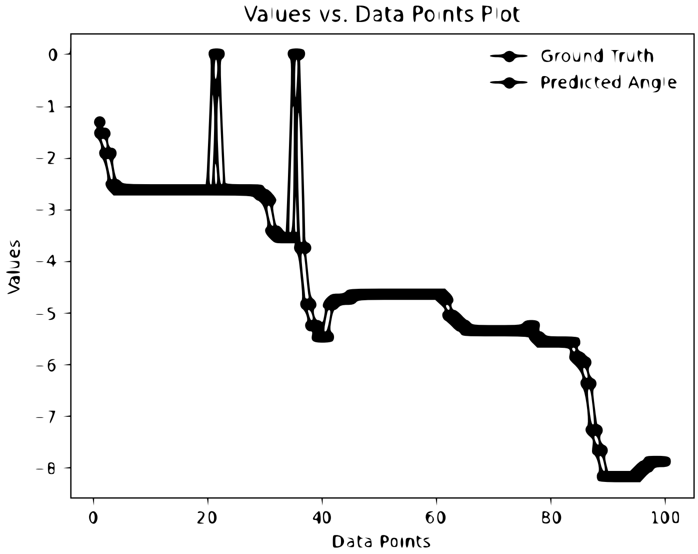

Fig. 1. Results and comparison table

A proposed strategy for mitigating this challenge is to consider the cropping of background information during the model training phase. By selectively focusing on the essential features, such as the edges of the road, and excluding less relevant background details, the CNN may achieve improved performance and more accurate predictions.

This observation underscores the importance of carefully curating the input data and preprocessing strategies in training CNN models for tasks such as steering prediction. Fine-tuning the model's sensitivity to relevant features while minimizing the impact of extraneous information is crucial for enhancing overall performance.

6.2. Limitations: Assumption of Constant Flow and Limited Scene Understanding

The use of optical flow methods, such as the Farneback method, is constrained by their assumption of a constant flow within a local neighborhood and their limited capacity to comprehend complex scene dynamics. The assumption of constant flow may falter in scenarios characterized by rapidly changing motion patterns or the presence of occlusions, impacting the accuracy of motion estimation. Optical flow methods may face challenges adapting to situations where objects move in and out of the scene or encounter significant changes in illumination. This limitation hinders their effectiveness in dynamic and unpredictable environments, where the assumption of constant flow does not align with the true dynamics of the scene. The reliance on this assumption can introduce inaccuracies, limiting the methods' ability to capture nuanced variations in motion.

7. Conclusions

We have demonstrated that through detailed processing of the reconstructed holographic images, performed by changing the object-hologram distance in the reconstruction code, it is possible to discriminate depth in the object. Using a specially fabricated object composed of spherical markers 465 nm in diameter spread on a tilted transparent surface, the reconstruction and analysis of the hologram allowed to map the surface topography with a resolution close to 2 m, with such resolution depending on the particular NA of the exposure.

The lateral resolution of the image obtained by numerical reconstruction was assessed utilizing a wavelet image decomposition and image correlation. The best lateral resolution obtained with a high NA recording, 164 nm, represents an improvement of more than a factor two relative to previously published results.

8. Acknowledgment

We would like to express my sincere gratitude to Dr. Bibhas Ghoshal for his invaluable guidance and unwavering support throughout our academic journey and particularly in this project. His expertise and encouragement have been instrumental in shaping the success of this endeavor.

9. Refrences

1) Fernández, Nelson. “Two-stream convolutional networks for end-to-end learning of self-driving cars.” ArXiv abs/1811.05785 (2018)

2) [1710.03804] End-to-End Deep Learning for Steering Autonomous Vehicles Considering Temporal Dependencies (arxiv.org)

3) Shah, Syed Xuezhi, Xiang. (2021). Traditional and modern strategies for optical flow: an investigation. SN Applied Sciences. 3. 10.1007/s42452-021-04227-x.

4) Shen, Shihao Kerofsky, Louis Yogamani, Senthil. (2023). Optical Flow for Autonomous Driving: Applications, Challenges and Improvements.

5) Smolyakov, M. Frolov, A.I. Volkov, V.N. Stelmashchuk, I.V.. (2018). Self-Driving Car Steering Angle Prediction Based On Deep Neural Network An Example Of CarND Udacity Simulator. 1-5. 10.1109/ICAICT.2018.8747006.

6) Fathy, Mahmoud Ashraf, Nada Ismail, Omar Fouad, Sarah Shaheen, Lobna Hamdy, Alaa. (2020). Design and implementation of self-driving car. Procedia Computer Science. 175. 165-172. 10.1016/j.procs.2020.07.026.

7) Mariusz Bojarski, Philip Yeres, Anna Choromanska, Krzysztof Choromanski, Bernhard Firner, Lawrence Jackel, Urs Muller.Explaining How a Deep Neural Network Trained with End-to-End Learning Steers a Car

8) https://github.com/SullyChen/driving-datasets/tree/master