Driving Hands - Hand Gesture Interaction for Autonomous Vehicles

Aadhya Kaul, Akshay Narkhede, Hiren Rupchandani, Nishad Gupta

Indiana University Bloomington

Introduction

- This project explores hand gesture interaction for autonomous vehicles.
- This system utilizes a dashboard-mounted camera to capture hand gestures in real-time.
- By starting with four essential gestures - up for acceleration, down for braking, left for turning left, and right for turning right - the project lays the groundwork for hand gestures in autonomous vehicles.

Research Questions

- How does altering the architecture of a neural network improve the classification of hand gestures?
- How do different data augmentation techniques affect the performance of hand gesture recognition models?
- How can a large hand gesture dataset be curated to ensure robust predictions in the presence of background noise?

Data

Figure 1. Sample images from our Dataset

Approaches

1. Increasing the dataset with custom images and assigning labels to the image data. Augmenting image data by shuffling and rotating the image data.
2. Implemented Resnet50 from scratch which has 49 convolutional layers and 14 identity blocks.
3. Implemented 3 layers of CNN with max pooling and Relu activation function after every layer.
4. During transfer learning, a custom head is used in MobileNet V3 Small/Large, V2 architecture.

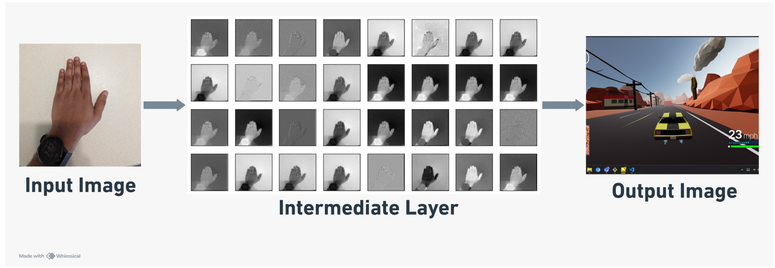


Figure 2. Visualizing Intermediate Layers in the Network

Model Architecture

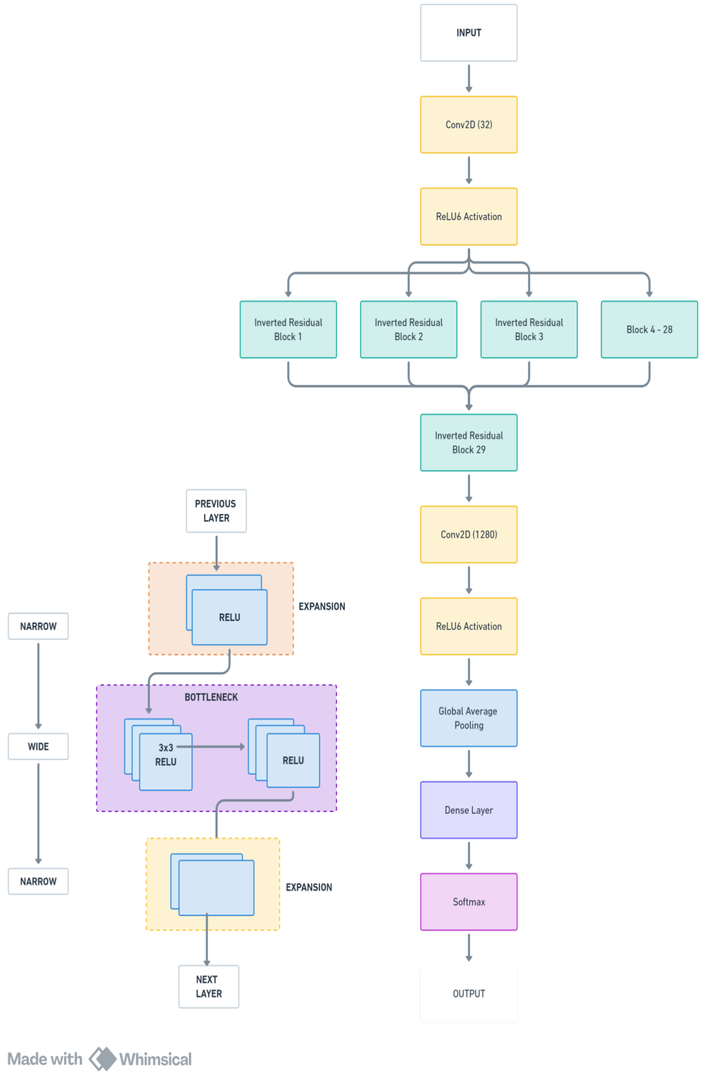
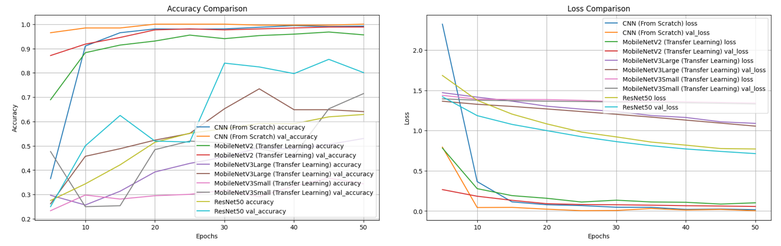


Figure 3. Model Architecture

Results



(a) Accuracy Comparison (b) Loss Comparison

Table 1. Model Performance Overview

Model Name	Size	Total Parameters	Training Loss	Training Accuracy	Validation Loss	Validation Accuracy
Convolutional Neural Network	169.37MB	44.4M	0.016	99.41%	0.004	100%
ResNet50	101.2MB	22.64M	0.66	84.86%	0.69	82.03%
MobileNetV2	11.12MB	2.91M	0.056	98.77%	0.086	96.82%
MobileNetV3 Small	4.3MB	2.9M	1.34	36.71%	1.32	71.48%
MobileNetV3 Large	10.9MB	5.4M	1.08	53.02%	1.05	73.43%

- A pre-trained MobileNetV2 was fine-tuned with a custom head. In contrast, ResNet50 is larger and slower even though it offers good accuracy.
- It excels in efficiency and size, making it ideal for mobile and edge computing, achieving 98% and 97% accuracy on training and validation sets, respectively.
- Custom CNNs offer flexibility but require careful design to balance efficiency and accuracy so we ended up choosing MobileNetV2.

Conclusion

We've found that a real time implementation of semi-assisted driving is possible with the help of MobileNetV2 model. The integration is seamless and the model is able to send the predicted hand signs to the car simulator in real-time

Future Work

The following things are being considered for our future work:

- Additional Gesture Detection
- We need more data for our project because the current dataset isn't big enough.
- Make our model more robust to noise
- Add bounding box for better detection of hand and gesture.

References

1. End to End Learning for Self-Driving Cars
extension://gphandlahdpffmccakmbngmbjnjiiahp/https://images.nvidia.com/content/tegra/automotive/images/2016/solutions/pdf/end-to-end-dl-using-px.pdf

2. SynthoGestures: A Novel Framework for Synthetic Dynamic Hand Gesture Generation for Driving Scenarios
https://arxiv.org/abs/2309.04421v1