AUTONOMOUS CAR USING DEEP LEARNING - FINAL REPORT

Rudra Dey
Student# 1010124866
rudra.dey@mail.utoronto.ca

Pravin Kalaivannan
Student# 1010141295
pravin.kalaivannan@mail.utoronto.ca

Aadavan Vasudevan
Student# 1010101514
aadavan.vasudevan@mail.utoronto.ca

Abishan Baheerathan
Student# 1010218756
abishan.baheerathan@mail.utoronto.ca

ABSTRACT

This template should be used for all your project related reports in APS360 course.
– Write an abstract for your project here. Please review the **First Course Tutorial** for a quick start —-Total Pages: 9

1 INTRODUCTION

2 ILLUSTRATION

3 BACKGROUND & RELATED WORK

4 DATA PROCESSING

5 ARCHITECTURE

As shown in Figure 1 , our project uses three separate deep learning models, velocity prediction, traffic light classification, and steering angle prediction. All models are built on a shared convolutional backbone of ResNet-18 which is pretrained on ImageNet. ResNet was selected for its strong low-level and mid-level feature extraction ability, particularly in structured domains such as road images.

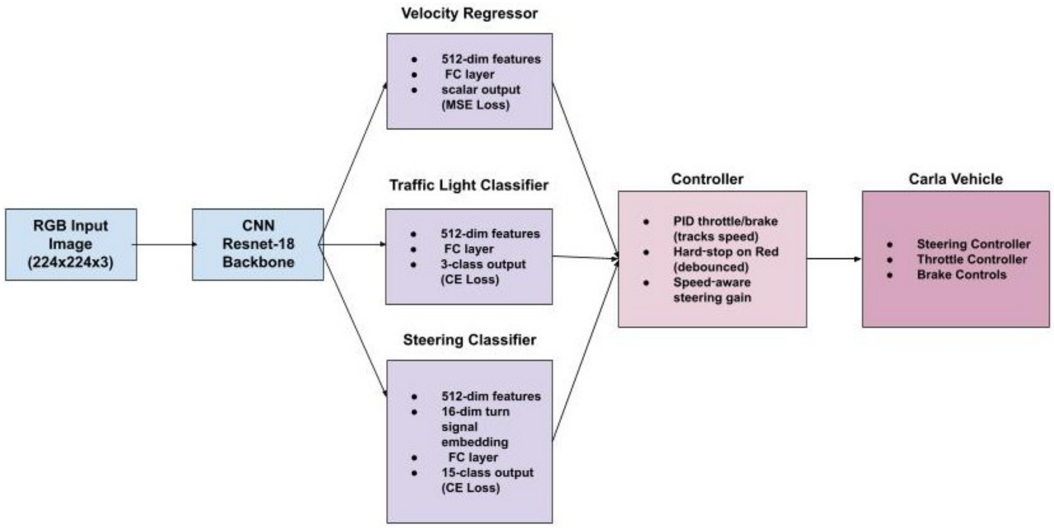

Figure 1: Final Architecture Low Level Diagram

5.1 TRANSFER LEARNING WITH RESNET-18

The ResNet-18 backbone processes each 224x224 dimensioned RGB frame to produce a 512- dimensional feature vector. For our application, only the convolutional layers of ResNet-18 are retained. In the early stages of training, these layers are frozen to preserve pretrained weights, with selective unfreezing applied during fine-tuning to adapt to the driving domain.

5.2 VELOCITY REGRESSOR

The velocity regressor receives the 512-dimensional feature vector and processes it through a two-layer fully connected network with a ReLU activation between layers. The first layer reduces dimensionality from 512 to 256 units, and the second produces a single scalar velocity value. Mean Squared Error (MSE) loss is applied during training to encourage accurate continuous speed prediction.

5.3 STEERING CLASSIFIER

For steering control, the 512-dimensional backbone output is concatenated with a 16-dimensional learned turn-signal embedding, producing a 528-dimensional input vector. This vector passes through a fully connected layer reducing it to 256 units, followed by a ReLU activation, and finally a fully connected layer outputting logits for 15 discrete steering angle bins. Cross-Entropy loss is used for classification

5.4 TRAFFIC LIGHT CLASSIFIER

The traffic light classifier head reduces the 512-dimensional input to 256 via a fully connected layer with ReLU activation, followed by a final fully connected layer outputting probabilities for three traffic light states (Red, Green and No-Light). Cross-Entropy loss drives this classification task.

5.5 PARAMETER BREAKDOWN

Table 1 summarizes the custom fully connected layers in each prediction head, excluding the ResNet-18 backbone parameters (11M).

Layer	Type	Input Size	Output Size	Parameters *(Formula)*	Description
Velocity_fc1	Fully Connected	512	256	131,328 *(512×256 + 256)*	Velocity regressor – feature reduction
Velocity_fc2	Fully Connected	256	1	257 *(256×1 + 1)*	Scalar speed output (MSE Loss)
Traffic_fc1	Fully Connected	512	256	131,328 *(512×256 + 256)*	Traffic light classifier – feature reduction
Traffic_fc2	Fully Connected	256	3	771 *(256×3 + 3)*	3-class output (CE Loss)
Steering_fc1	Fully Connected	528 *(512+16)*	256	135,168 *(528×256 + 256)*	Image + turn signal embedding fusion
Steering_fc2	Fully Connected	256	15	3,855 *(256×15 + 15)*	15 steering angle bins (CE Loss)

Table 1: Final Architecture Low Level Diagram

5.6 Quantitative Results

5.7 Qualitative Results

6 Baseline Model

Using a Ridge Regression Algorithm we created a baseline model that predicts steering angles for a self-driving car from grayscale camera images and turn signal inputs. This baseline model serves as a simple, interpretable baseline such that we can compare our more complex primary neural network model later. To collect data, we used a simulator called CARLA, to obtain camera images, turn signal inputs, and our ground truth label for the steering angles.

6.1 Ridge Regression with Image Features

In the model outlined in Figure 2, the grayscale images (of shape 160x120) were flattened into 1D feature vectors and normalized. The features themselves represent the visual input of the car's front-facing camera. Additionally, left and right turning signals were captured as an additional feature; combining this with our flattened grayscale images we were left with a numpy array of shape $N \times 19201$, where N represents the number of images, and 19201 are the number of feature column vectors. The Ridge Regression algorithm was selected because it penalized large coefficients (L_2 norm), allowing for a generalized model. A series of models were trained with different regularization strengths (α), and the performance was evaluated using Mean Squared Error (MSE) and R^2 Score on a held-out 20% test set.

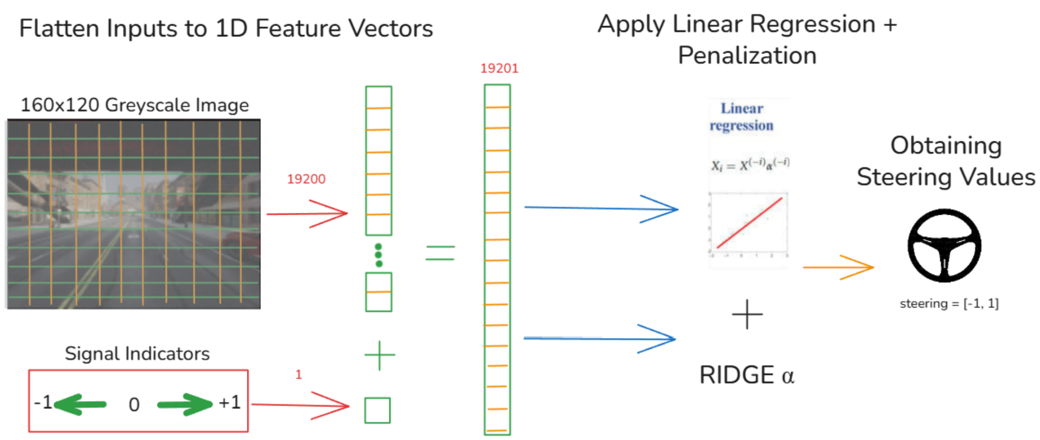

Figure 2: Basic Architecture of Steering Angle Ridge Regrsession Model

6.2 Minimal Tuning of α

The only hyperparameter needing to be tuned was the regularization strength for the Ridge Regression Algorithm. This was manually tuned, and thus no validation set was used. A small set of candidate values for α was chosen: $\alpha \epsilon \{1, 10, 100, 250, 1000, 8000\}$ With each value of α, the model was trained and tested with a dataset of 15 minutes of simulated driving. By obtaining the average MSE and the average R^2 score, our fourth model of $\alpha = 100$ came out to be the best.

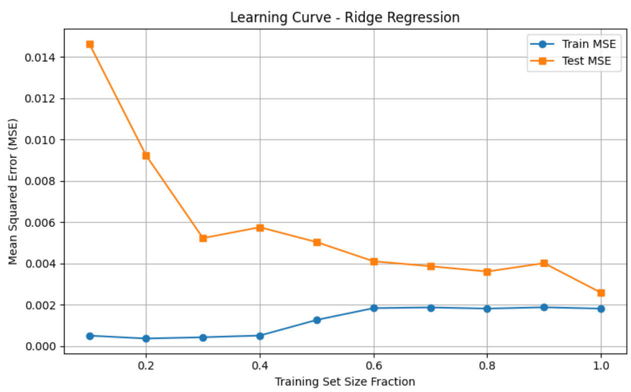

Figure 3: This is the learning curve for "Model 4" a Ridge Regression Model with $\alpha = 100$

6.3 Overfitting and Model Quality

As shown in Figure 3, the results suggest some overfitting due to the model's tendency to perform much better on training data than unseen data. This is reasonable since the training data frames, are all consecutive in nature, and the dataset used isn't extremely large either. In addition, as the training size increases the gap between Test and Train MSEs narrows, suggesting that with more data the model will generalize better. The final test MSE is (0.0025) which is quite low, meaning the model performs well overall after training on the data.

6.4 Qualitative Observations

When running the steering angle prediction model in the CARLA simulator, the model exhibited promising behaviour in terms of autonomous driving capabilities, however with large limitations. The car was able to drive freely, using the model's predictions from the front-facing grayscale camera images and turn signals provided by the user. In Figure 4 we can see the model taking a left turn in an intersection.

- **Turn Signal Response:** The car consistently turned in the correct direction when prompted, showing that the model incorporated turn signal input effectively.
- **Obstacle Avoidance:** The vehicle generally avoided obstacles but occasionally clipped walls, suggesting basic spatial awareness but limited precision.
- **Lane-Keeping:** The car had trouble staying within lane boundaries, often drifting, especially during turns or on complex roads.
- **Map Generalization:** Despite limitations, the car performed reasonably well across different map layouts, showing decent control over turns.

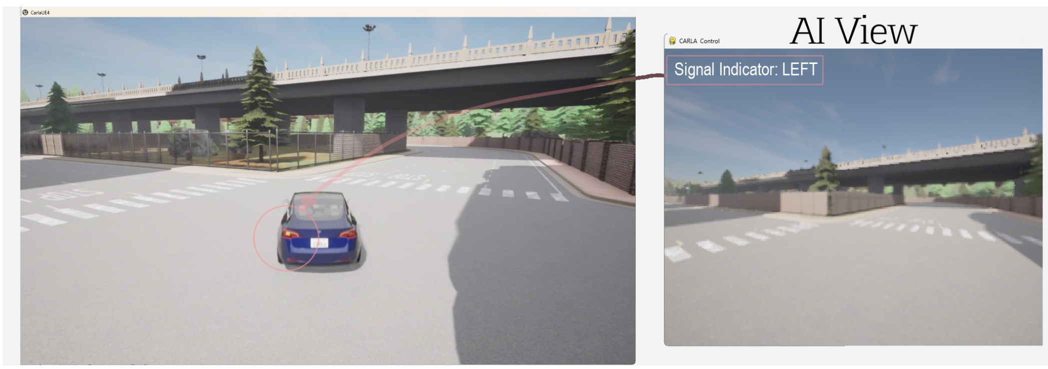

Figure 4: This is the demo for ”Model 4” in CARLA

7 Evaluate Model on New Data

8 Discussion of Results

9 Ethical Considerations

10 Project Difficulty / Quality

10.1 Citations within the text

Citations within the text should be based on the `natbib` package and include the authors' last names and year (with the "et al." construct for more than two authors). When the authors or the publication are included in the sentence, the citation should not be in parenthesis using `\citet{}` (as in "See Hinton et al. (2006) for more information."). Otherwise, the citation should be in parenthesis using `\citep{}` (as in "Deep learning shows promise to make progress towards AI (Bengio & LeCun, 2007).").

The corresponding references are to be listed in alphabetical order of authors, in the References section. As to the format of the references themselves, any style is acceptable as long as it is used consistently.

To cite a new paper, first, you need to add that paper's BibTeX information to `APS360_ref.bib` file and then you can use the `\citep{}` command to cite that in your main document.

Table 2: Sample table title

PART	DESCRIPTION
Dendrite	Input terminal
Axon	Output terminal
Soma	Cell body (contains cell nucleus)

10.2 Footnotes

Indicate footnotes with a number[1] in the text. Place the footnotes at the bottom of the page on which they appear. Precede the footnote with a horizontal rule of 2 inches (12 picas).[2]

10.3 Figures

All artwork must be neat, clean, and legible. Lines should be dark enough for purposes of reproduction; art work should not be hand-drawn. The figure number and caption always appear after the figure. Place one line space before the figure caption, and one line space after the figure. The figure caption is lower case (except for first word and proper nouns); figures are numbered consecutively.

Make sure the figure caption does not get separated from the figure. Leave sufficient space to avoid splitting the figure and figure caption.

You may use color figures. However, it is best for the figure captions and the paper body to make sense if the paper is printed either in black/white or in color.

Figure 5: Sample figure caption. Image: ZDNet

10.4 Tables

All tables must be centered, neat, clean and legible. Do not use hand-drawn tables. The table number and title always appear before the table. See Table 2.

Place one line space before the table title, one line space after the table title, and one line space after the table. The table title must be lower case (except for first word and proper nouns); tables are numbered consecutively.

[1] Sample of the first footnote

[2] Sample of the second footnote

11 DEFAULT NOTATION

In an attempt to encourage standardized notation, we have included the notation file from the textbook, *Deep Learning* Goodfellow et al. (2016) available at `https://github.com/goodfeli/dlbook_notation/`. Use of this style is not required and can be disabled by commenting out `math_commands.tex`.

Numbers and Arrays

a	A scalar (integer or real)
$\boldsymbol{a}$	A vector
$\boldsymbol{A}$	A matrix
$\mathsf{\mathbf{A}}$	A tensor
$\boldsymbol{I}_n$	Identity matrix with n rows and n columns
$\boldsymbol{I}$	Identity matrix with dimensionality implied by context
$\boldsymbol{e}^{(i)}$	Standard basis vector $[0,\dots,0,1,0,\dots,0]$ with a 1 at position i
$\text{diag}(\boldsymbol{a})$	A square, diagonal matrix with diagonal entries given by $\boldsymbol{a}$
a	A scalar random variable
$\mathbf{a}$	A vector-valued random variable
$\mathbf{A}$	A matrix-valued random variable

Sets and Graphs

$\mathbb{A}$	A set
$\mathbb{R}$	The set of real numbers
$\{0,1\}$	The set containing 0 and 1
$\{0,1,\dots,n\}$	The set of all integers between 0 and n
$[a,b]$	The real interval including a and b
$(a,b]$	The real interval excluding a but including b
$\mathbb{A}\backslash\mathbb{B}$	Set subtraction, i.e., the set containing the elements of $\mathbb{A}$ that are not in $\mathbb{B}$
$\mathcal{G}$	A graph
$Pa_{\mathcal{G}}(\mathrm{x}_i)$	The parents of x_i in $\mathcal{G}$

Indexing

a_i	Element i of vector $\boldsymbol{a}$, with indexing starting at 1
a_{-i}	All elements of vector $\boldsymbol{a}$ except for element i
$A_{i,j}$	Element i,j of matrix $\boldsymbol{A}$
$\boldsymbol{A}_{i,:}$	Row i of matrix $\boldsymbol{A}$
$\boldsymbol{A}_{:,i}$	Column i of matrix $\boldsymbol{A}$
$\mathsf{A}_{i,j,k}$	Element (i,j,k) of a 3-D tensor $\mathsf{\mathbf{A}}$
$\mathsf{\mathbf{A}}_{:,:,i}$	2-D slice of a 3-D tensor
a_i	Element i of the random vector $\mathbf{a}$

Calculus

$\dfrac{dy}{dx}$	Derivative of y with respect to x
$\dfrac{\partial y}{\partial x}$	Partial derivative of y with respect to x
$\nabla_{\boldsymbol{x}} y$	Gradient of y with respect to $\boldsymbol{x}$
$\nabla_{\boldsymbol{X}} y$	Matrix derivatives of y with respect to $\boldsymbol{X}$
$\nabla_{\mathsf{X}} y$	Tensor containing derivatives of y with respect to X
$\dfrac{\partial f}{\partial \boldsymbol{x}}$	Jacobian matrix $\boldsymbol{J} \in \mathbb{R}^{m \times n}$ of $f : \mathbb{R}^n \to \mathbb{R}^m$
$\nabla_{\boldsymbol{x}}^2 f(\boldsymbol{x})$ or $\boldsymbol{H}(f)(\boldsymbol{x})$	The Hessian matrix of f at input point $\boldsymbol{x}$
$\int f(\boldsymbol{x}) d\boldsymbol{x}$	Definite integral over the entire domain of $\boldsymbol{x}$
$\int_{\mathbb{S}} f(\boldsymbol{x}) d\boldsymbol{x}$	Definite integral with respect to $\boldsymbol{x}$ over the set $\mathbb{S}$

Probability and Information Theory

$P(\mathrm{a})$	A probability distribution over a discrete variable
$p(\mathrm{a})$	A probability distribution over a continuous variable, or over a variable whose type has not been specified
$\mathrm{a} \sim P$	Random variable a has distribution P
$\mathbb{E}_{\mathrm{x} \sim P}[f(x)]$ or $\mathbb{E} f(x)$	Expectation of $f(x)$ with respect to $P(\mathrm{x})$
$\mathrm{Var}(f(x))$	Variance of $f(x)$ under $P(\mathrm{x})$
$\mathrm{Cov}(f(x), g(x))$	Covariance of $f(x)$ and $g(x)$ under $P(\mathrm{x})$
$H(\mathrm{x})$	Shannon entropy of the random variable x
$D_{\mathrm{KL}}(P \Vert Q)$	Kullback-Leibler divergence of P and Q
$\mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$	Gaussian distribution over $\boldsymbol{x}$ with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$

Functions

$f : \mathbb{A} \to \mathbb{B}$	The function f with domain $\mathbb{A}$ and range $\mathbb{B}$
$f \circ g$	Composition of the functions f and g
$f(\boldsymbol{x}; \boldsymbol{\theta})$	A function of $\boldsymbol{x}$ parametrized by $\boldsymbol{\theta}$. (Sometimes we write $f(\boldsymbol{x})$ and omit the argument $\boldsymbol{\theta}$ to lighten notation)
$\log x$	Natural logarithm of x
$\sigma(x)$	Logistic sigmoid, $\dfrac{1}{1 + \exp(-x)}$
$\zeta(x)$	Softplus, $\log(1 + \exp(x))$
$\|\|\boldsymbol{x}\|\|_p$	L^p norm of $\boldsymbol{x}$
$\|\|\boldsymbol{x}\|\|$	L^2 norm of $\boldsymbol{x}$
x^+	Positive part of x, i.e., $\max(0, x)$
$\mathbf{1}_{\mathrm{condition}}$	is 1 if the condition is true, 0 otherwise

12 Final instructions

Do not change any aspects of the formatting parameters in the style files. In particular, do not modify the width or length of the rectangle the text should fit into, and do not change font sizes (except perhaps in the REFERENCES section; see below). Please note that pages should be numbered.

Author Contributions

If you'd like to, you may include a section for author contributions as is done in many journals. This is optional and at the discretion of the authors.

Acknowledgments

Use unnumbered third level headings for the acknowledgments. All acknowledgments, including those to funding agencies, go at the end of the paper.

References

Yoshua Bengio and Yann LeCun. Scaling learning algorithms towards AI. In *Large Scale Kernel Machines*. MIT Press, 2007.

Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. *Deep learning*, volume 1. MIT Press, 2016.

Geoffrey E. Hinton, Simon Osindero, and Yee Whye Teh. A fast learning algorithm for deep belief nets. *Neural Computation*, 18:1527–1554, 2006.