

Domain Background:

Optical Character Recognition (OCR) has been around since early 1900's. It is a use of computer software to scan graphical objects and recognize them as characters.

OCR has been widely used in postal services in which mailing address is scanned and translated into computer-encoded text such that with minimal human aid, mail / parcels can be sorted by scanning systems. This improves delivery efficiency.

Problem statement:

Vehicle license plate recognition is useful for law enforcement purposes. Being able to use machines to read plates is essential for officers to pull vehicle-related information such as overdue tolls and offence history. The question is how to make our model learn to read correctly. In this capstone project, I will be implementing a machine learning program that attempts to recognize vehicle license plates containing numbers and alphabets.

Datasets and Inputs:

There is a [website](#) that provides vehicle license plate template (1994 spec) that allows users to enter their alphabets and numbers. The 1994 specifications consisted of the word 'Ontario' at the top with a 'crown' right in the middle of the plate and ended with a slogan – "Your To Discover" in the footer. For this project, I am using the template for Ontario in which I live. To generate data, I have written a script to pull a plate every four seconds from the site. I was able to generate about sixty thousand (60,000) plates with characters and numbers combined ranging from AAA 000 to ZZZ 999. The script can be found [here](#). Each plate is saved using the characters and numbers from itself. For example, plate ABC 123 is saved as 'ABC 123.jpg'

Below are samples of license plates,



Solution statement:

In this project, the data will be split into a training set and a test set. The workflow is to train our model to read 50,000 images using convolutional neural networks (CNN). Convolutional Neural Network is popular and commonly used in image classification. The test set that contains at least a thousand (1000) plates will be used to gauge the accuracy of our model prediction.

Benchmark Model:

Since the template given is the same in the training set and test set, it is the characters/numbers that differ among the plates. I assume the trained model can achieve about ninety percent(90%) accuracy from the test data.