Using Machine Learning and Optical Character Recognition
to adaptively process text in pictures and scans

Deirdre Chong, Jorge Garcia,
Vanessa Nava-Camal, Zachary FitzHugh
Undergraduate Department of Computer Science
University of Nevada, Las Vegas

Jorge Ramón Fonseca Cacho
Department of Computer Science
University of Nevada, Las Vegas
Email: Jorge.FonsecaCacho@unlv.edu

I. Introduction

In this paper we will use Optical Character Recognition and Intelligent Character Recognition, with a

Machine Learning has seen a massive demand in computer applications over the past few years. From self-driving cars to voice assistants like Siri, Machine Learning makes it possible for the computer to solve such hard problems. Our research aims to apply Machine Learning in image processing and OCR. OCR stands for Optical Character Recognition. It is used to read text from images such as a scanned document or a picture. This technology is used to convert, virtually any kind of images containing written text (typed, handwritten, or printed) into machine-readable text data.

OCR has two major building blocks:

- Text Detection
- Text Recognition

Text detection or in general object detection has been an area of intensive research accelerated with deep learning. Today, object detection, and in our case, text detection, can be achieved through two approaches:

- Region-based Detection
- Single Shot Detection

Region-based methods work by doing multiple passes to process the input. They first try to find all the regions which have objects the application can recognize, then, pass the regions detected to a classifier engine, which as its name suggests, it classifies the regions into the types we specify.

A single shot method on the other hand, tries to find the regions and the type at the same time, or in a single shot. Since this is a single step process, it is usually much faster but the downside of this being that it struggles with bad quality inputs and smaller objects.

Index Terms—**Machine Learning, OCR, Computer Vision, Deep Learning, Text Processing, Optical Character Recognition, YOLOv3, Tesseract**

II. Objectives

The main objective of this research is developing an intelligent and adaptable character recognition is able to extract text and numbers from any type of card provided, these being credit or gift cards with high level of accuracy, regardless if the input is a scan or a photo. We will be experimenting with both methods mentioned, Region-based detection and Single shot detection, and see which provides a better performance given our constrains and input quality. We will gather enough data and samples to be create our own dataset and train our model to our specific cards to get a better accuracy than what broadly used datasets and pre-trained models provide.

III. Acknowledgement