CS63 Spring 2025
Final Project Checkpoint

Victor Sumano Arango
Nicholas D'Andrea

1 Project Goal

Before the recent breakthroughs in hardware and software, object detection relied primarily on traditional feature-based algorithms. One of the earliest methods, sliding windows, involved scanning an image with a fixed-size window that moved across every possible position, checking each window to determine if an object was present. In 2005, the Histogram of Oriented Gradients (HOG) method introduced a more sophisticated approach by analyzing edge orientations within an image, capturing its shape and appearance. Three years later, in 2008, Deformable Part Models (DPM) advanced the field further by breaking objects into parts, each represented by a bounding box, allowing for a more flexible and robust detection process.

Following the AI boom and the subsequent surge in interest in Convolutional Neural Networks, R-CNNs became the object detection standard. In 2016, YOLO (You Only Look Once) revolutionized object detection due to its speed and efficiency compared to R-CNNs.

For our final AI project, we aim to study the YOLO architecture and explore its methods for object detection. Our goal is to build a simple model from scratch, applying our knowledge of CNNs and Genetic Algorithms to fine-tune the model and solve Google's reCaptcha v2 challenges. If time permits, we would like to fully create a system using OCR technology to locally deploy our model to solve reCaptcha v2 challenges.

2 AI Methods Used

We plan to use the following AI methods:

1. Convolutional Neural Networks
    (a) We plan to customize our architecture and backpropagation algorithm
2. We plan to develop a Genetic Algorithm for hyperparameter tuning
3. We plan to use Transfer Learning for multi-model cross comparisons

3 Staged Development Plan

1. Complete, train, and evalute simple YOLO network on novel images

2. Train simple YOLO network using pre-trained weights from 1. and evaluate on Google reCaptcha v2 images

3. Develop Genetic Algorithm to fine tune simple YOLO network and find optimal hypterparameters

4. Use Transfer Learning to conduct model cross comparisons with YOLOv11 and our simple YOLO network

5. Locally deploy our simple YOLO network, use OCR, and evaluate on real-time Google reCaptcha v2 challenges

4 Measure of Success

Although our goal is to build a model that can accurately detect objects most of the time, we are mindful of our time constraints. So, to be successful, we would like to complete up to 2. in our proposed staged development plan.

5 Plans for Analyzing Results

Some of the images we plan to use aren't fully annotated so we may need to annotate by hand or through YOLO World. Due to the lack of annotated data, we will perform a 80/20 split on our annotated data for training/testing.

We plan to use the following performance metrics:

1. Confusion Matrix

2. Accuracy, precision, recall, and F1-score

3. Anchor box confidence scores

4. Training and validation loss and accuracy