

---

# ToyBox (Computer Vision)

---

**Anonymous Author(s)**

Affiliation

Address

email

## Abstract

This research project aims to develop and evaluate a computer vision model and compare the model to that of one trained on a single frame viewpoint. The project will utilize the ToyBox dataset to see how well various angles can improve training compared to a dataset of front viewpoints.

## 1 Project Design

This project plans to utilize the multi-view video from the ToyBox dataset to train a machine learning model to classify toys. The idea is that given the multiple perspectives from a “human” view, the model should result in a better trained model than that of a dataset containing only images. Our primary project is to have an enhanced model over that of an imaged trained model such that the model is able to more accurately generalize new samples given the varying perspectives of the objects. To do this we will create a model in Python to evaluate both the video format and a single frame from each video. We will then compare the models using various metrics such as accuracy, confusion matrix, and F1 score.

## 2 Dataset Details

The dataset for this project is a video dataset of children’s toys. Unlike most classification projects which contain an image, this dataset is a video that includes multiple viewpoints of each toy. While this data contains 12 categories broken down into 3 sets (household items, animals, and vehicles) our initial testing will focus primarily on the animals dataset. Both due to the complexity of the video dataset, and size of the dataset. The size of the animals dataset is 36.9 GB compressed for instance. The animal dataset has 4 types of animals: duck, cat, horse, giraffe. Additionally, the video is high resolution at 1080p and appears to have good focus which will enhance the training of our model. Time permitting we will expand our testing to the other datasets.

## 3 Software

We will use python for fundamental coding of our project and rely on various libraries. For basic data plotting and graph purposes we will make use of the Matplotlib library. Matplotlib is a great tool for data visualization and will enable us to display data in a report friendly way. For array functions and manipulation we will make use of the NumPy library as used in previous work during this class. OpenCV will be used for the integration of video formats from the dataset provided by ToyBox, this will save us significant time converting video to usable information. Git/Github will be used for version/project management and for coordinating our coding efforts. Lastly, TensorFlow will be used as a ML Framework.

## **4 Planned Analysis**

For our analysis we will compare our model trained on a video set of human viewpoint, but single frame photos to compare accuracy. We will also compare different amounts of training data.

## **5 Teammates**

This project will have two teammates: Jackson Shaw and Jonathan Pitts. While we have not determined a split on the workload, once we establish our fundamental functions and their requirements, we will begin to split them up.

## **6 Tentative Timeline**

Tentatively our project will have 4 phases. During our first phase "Basic Model Outline Creation" we will focus on the outline of our code and what functions need to be created for functionality of our project. From there we will complete the coding base to achieve a working model. During the next phase, "Complete Training for Video / Photo", we will complete any remaining coding and ensure the models are training and able to predict based on the dataset. From there we will evaluate the video from the single frame model to compare their performance. Lastly we will put all this information together in our final report to share our results.

Oct 1, 2024 - Oct 18, 2024 for "Basic Model Outline Creation"

Oct 18, 2024 - Oct 31, 2024 for "Complete Training for Video / Photo"

Oct 31, 2024 - Nov 15, 2024 for "Evaluate Models"

Nov 15, 2024 - Dec 1, 2024 for "Complete Final Report"