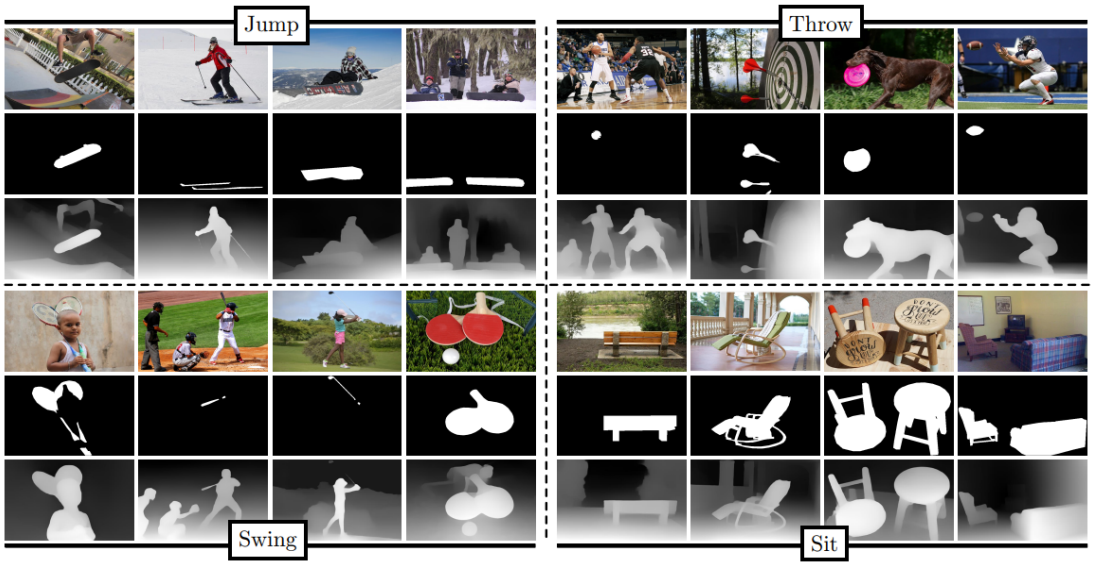
ADVANCED MACHINE LEARNING
CS5824/ECE5424 PROJECT PROPOSAL

ONE-SHOT AFFORDANCE DETECTION



Coordinator

Sandeep Chinnareddy/ Aruj Nayak

sandeepc98@vt.edu, *aruj@vt.edu*

Members

Trisha Bora, Prayati Dutta, Akhilesh Marathe

trishab@vt.edu, *dprayati@vt.edu*, *akhimarathe@vt.edu*

10.20.2022

DEFINITION

The ability of identifying the potential action possibilities of the objects present in an image can be defined as affordance detection. A support image estimating the action purpose and then transferring the same to all the objects in all the candidate images which would help in the detection of the common accordance can be defined as the *one-shot affordance detection network* needed to execute it.

In this project, we will be tackling the problem of Affordance Detection by using multiple deep learning models and leveraging state-of-the-art techniques like Attention & Reinforcement Learning. Our primary goal is to detect the potential action possibilities by processing images in the acquired dataset.

MOTIVATION

With the advent of AI in robots and the increasing levels of vehicle automation in modern self-driving cars, identifying the possibilities of movement is paramount. In autonomous vehicles, for example, estimating the movements a vehicle might make can improve the reaction times of cars in motion to apply corrective action [1]. In robotics, robots need to be able to identify the interaction possibilities of objects, when placed in scenarios with insufficient prior knowledge or environments with noisy perception [2]. Our aim in this project is to improve the underlying algorithm of affordance detection to hopefully improve both these applications in the future.

RELATED WORK

Visual Affordance Detection is a widely used strategy for perceiving action intentions and also for inferring visual affordance from the images or videos of human actions/object positions [3]. Most of the early studies focused on learning an association between object characteristics and their affordance for perceiving affordance. As the representational characteristics always tend to shift depending on the interactions between objects and humans, thus the affordances of objects cannot be directly associated with these characteristics.

As now it has been observed that the interaction between an object and human can change this object's affordance state, thus some research has been started by taking into account the human action as a cue for learning affordance. [4] performed affordance reasoning by using the OPRA dataset to understand human-object interactions and then predicted the affordance regions of static objects by linking human actions to object affordance.

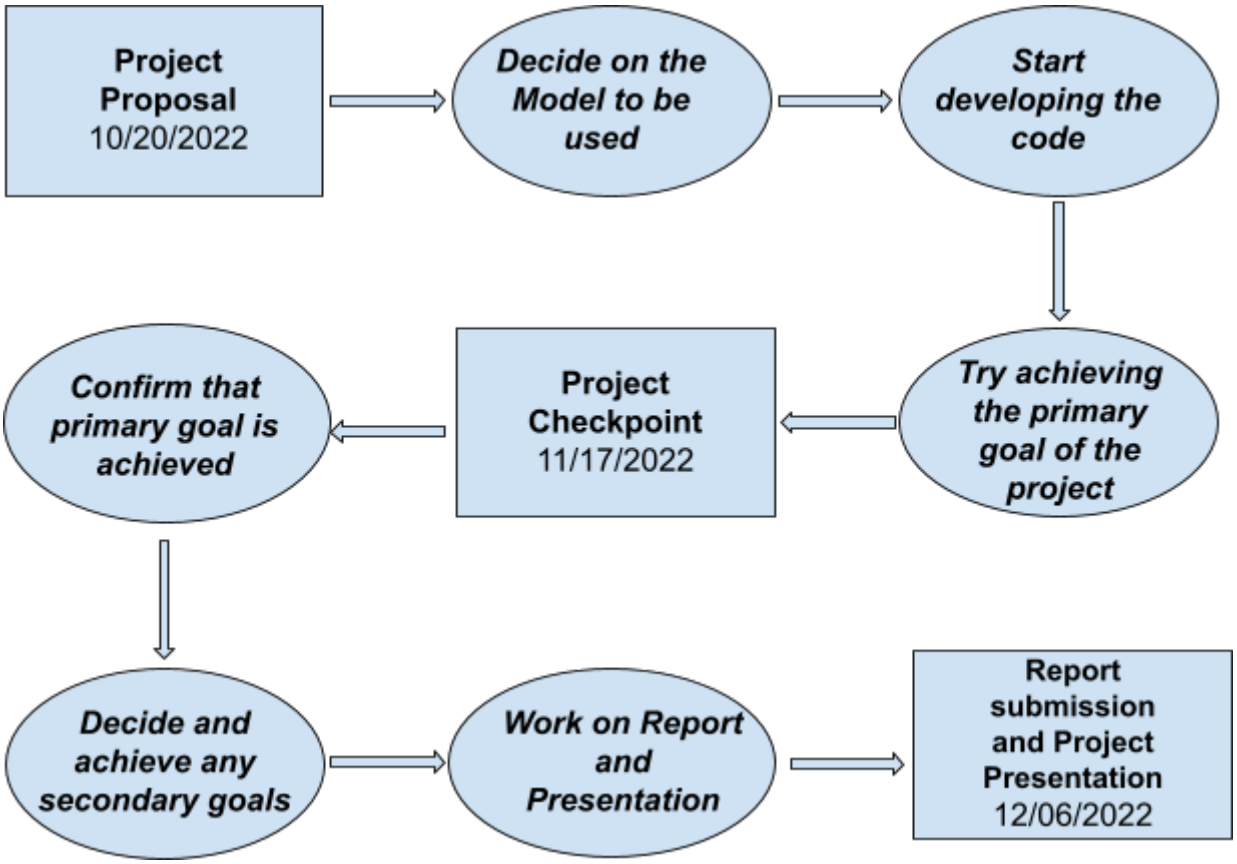
We will be working and intending to build our project upon the following paper- [5]

PLAN

1. Data Collection & Cleansing
2. Data Preprocessing
3. Exploratory Data Analysis
4. Feature Extraction
5. Model Building
6. Model training
7. Testing & Validation
8. Hyperparameter Tuning
9. Inferences and Results

TIMELINE

Week starting	10/18	10/25	11/01	11/08	11/15	11/22	11/29	12/06
Data Collection & Cleansing	✓							
Data Preprocessing		✓						
Exploratory Data Analysis			✓					
Feature Extraction			✓					
Model Building				✓				
Project checkpoint					✓			
Model training						✓		
Testing & Validation							✓	
Hyperparameter Tuning							✓	
Inferences and Results							✓	
Final Project Report & Presentation								✓



REFERENCES

[1] **https://deepdriving.cs.princeton.edu/**

[2] Christoph Pohl, Kevin Hitzler, Raphael Grimm, Antonio Zea, Uwe D. Hanebeck and Tamim. Asfour
Affordance-Based Grasping and Manipulation in Real World Applications. IROS, 2020

[3] Mohammed Hassanin, Salman Khan, and Murat Tahtali. Visual affordance and function understanding: A survey. arXiv, 2018.

[4] Kuan Fang, Te-LinWu, Daniel Yang, Silvio Savarese, and Joseph J. Lim. Demo2vec: Reasoning object affordances from online videos. In CVPR, 2018.

[5] Zhai Wei, Luo Hongchen, Zhang Jing, Cao Yang, and Tao Dacheng. One-Shot Object Affordance Detection in the Wild. arXiv, 2021.