Merck Computer Vision

Deep Graph Team: Yuke Zhang, Yan Chen, Sarthak Mangla
Human Parsing Team: Divay Gupta
Mentor: Ti-chiun Chang TA: Anish Thangavelu

Introduction

Deep Graph Team

Goal: To convert simulated drug formulation video into text data detailing procedures and methodology.

Human Parsing Team

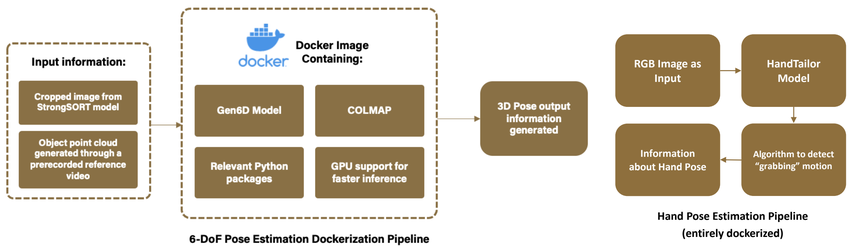
Goal: To ensure the safety and accuracy of laboratory procedures in pharmaceutical labs by verifying PPE (Personal Protective Equipment) compliance.

Acknowledgements

We'd like to sincerely thank our mentor, Ti-chiun Chang and our TA, Anish Thangavelu!

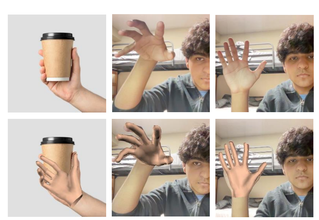
Additionally, we greatly appreciate the support from Terri Bui, Merck, and The Data Mine!

3D Attributes Extraction Flow

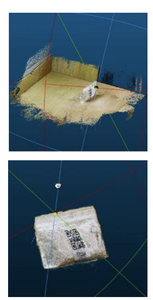


6-DoF Pose Estimation Dockerization Pipeline

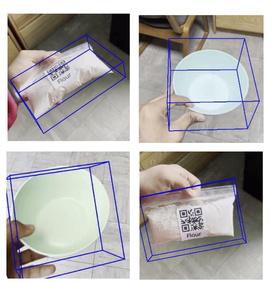
Hand Pose Estimation Pipeline (entirely dockerized)



Input-Output Pairs for Hand Pose Detection



Sample Point Cloud Generations



Sample Output Images for Rigid Objects

Future Work

Deep Graph Team

Deep Graph Action Modeling

Remove the restriction of position awareness by implementing a permutation invariant neural network structure

Utilize energy-based modeling to enhance the scalability of the model so that more object categories could be included

Compare the performance of this MLP framework to a GNN network

Improve accuracy of the algorithm to detect grabbing hand pose and make it rotation-agnostic

Human Parsing Team

PPE Compliance

Improve PPE classification accuracy by implementing a different model architecture or by using a larger or a better, more representative dataset.

Add more classes to the PPE classifications, in order to capture more kinds of PPE.

Choose or implement an alternative method of detecting PPE on the human (such as by using pose estimation to grab different body segments, and then classifying them).

Build models that work better on fisheye images, when it comes to handling edge cases like highly distorted body part ratios.

Deep Graph Action Modeling

Goal: Classify actions between objects within the context of the entire image

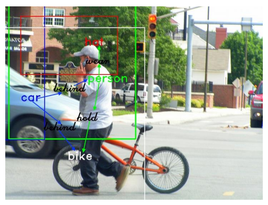
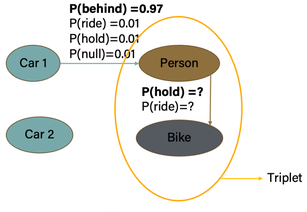


Figure 1: An example of a Generated Scene Graph



Object Set: P B C1 C2

Action Set: {ride, hold, behind, null}

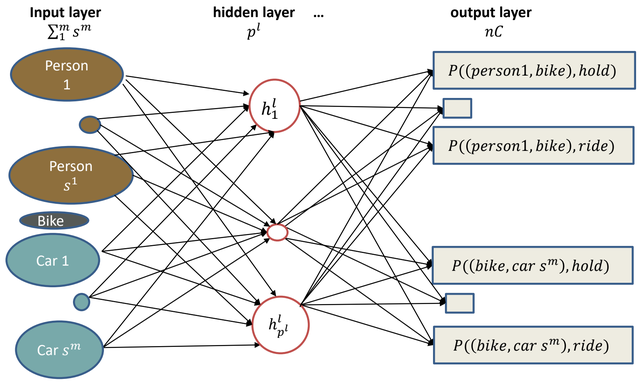


Figure 2: Illustration of position-aware NN model for action modeling.
m: number of object classes
s^i: number of instances in object category i
C: number of action classes
n: number of triplets
p^l : size of l^{th}hidden layer

Table: F1 score for the training dataset

Action	N Sample	F1 before	F1 after
ride	56	0.00	0.72
carry	59	0.00	0.74
hold	128	0.00	0.70
wear	718	0.02	0.75
behind	71	0.00	0.56
null	62388	0.25	0.99

Table: F1 score for the test dataset

Action	N Sample	F1 before	F1 after
ride	17	0.00	0.26
carry	17	0.00	0.41
hold	28	0.00	0.26
wear	166	0.02	0.64
behind	19	0.00	0.25
null	15713	0.25	0.99

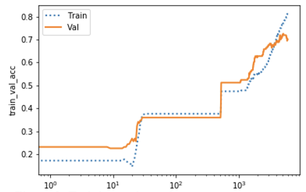


Figure 3: Training and test datasets accuracy trajectory Train: ~80%, Test: ~70%

Human Parsing Flow