Title of the Project: Video Temporal Action Localization (TAL) and Spatial Temporal Action Localization (STAL)

Group Members: 1. Jyothi Vishnu Vardhan Kolla 2. Karan Shah

Problem we are trying to solve:
TAL and STAL challenges aim at localizing the actions of interest in the video frames. For TAL, the output of the model is the temporal segments of the actions of interest. For STAL, the output of the model is the bounding boxes localizing the actions of the interest. In this task, we focus on the products and actions that associate with potential shopping behaviors, i.e., products that is taken into the basket, and products that is taken out of the basket.

What is novel aspect of your work?
Our project can automate the process of grocery shopping by first keeping track of the items the customer picked and then use Image matching methods to match the picked product with an image in the inventory and complete the final billing.

Data requirements and how we acquired it?
We need the video of a person who goes into a grocery store in form of Image per frame and then a file that contains annotations which includes the type of action and the coordinates of bounding boxes.

We are using real-world physical retail data collected in grocery store environments, named Grocery Vision 2023. This dataset contains individual shopping actions of anonymized customers collected with a GoPro camera mounted on a standard US shopping cart.

Computational requirements:
we might require a decent GPU for the process of training and experimenting with different networks and we are planning to discovery research cluster provided by Northeastern.

Expected Results:
To evaluate model's performance, frame-mAP and tube-mAP will be used.