Stack Overview

Technologies/Packages:

- Clip(ViT-B/32)
- CV2
- YOLO(medium)

Workflow:

1. User uploads images
2. YOLO detects objects, grabs their coordinates, and labels them (“chair”)
3. CV2 determines the dominant color within each object’s borders to give a color label (“blue”)
4. CLIP is used to track objects across photos
5. Calculates the distance an object has moved in pixels
6. Overlay containing information is placed on after image
7. Print image to screen

Weaknesses:

- Very Basic Color Detection
- Does not truly track objects
- Partially Obscured items = Removed
- Rare items will be harder to mark
- Lacks context (chair moved away from table)

Strengths:

- YOLO is fast (20s runtime outside browser - CPU)
- CLIP might be able to give context with finetuning
- Could potentially run locally
- Color Detection can be improved upon

Cost

Local:

- Little to no cost
- Runs slowly on cleaner side (have to wait for image processing)
- Larger app size
- Unsure if feasible

GPU server:

- Monthly cost depends on the needs of the feature
- Runs quickly on cleaners side (image processing is offloaded)
- App is less bloated
- Definitely feasible
- Potential Hosting Solution: Google Cloud Platform ~$0.35/hr (lower with commitment)