Nutrition Change Estimation Through Video

William Tao, Piotr Nabrzyski, Donny Weintz, Benjamin Nguyen

Video and Image Processing Laboratory (VIPER)

11/19/2025

PURDUE UNIVERSITY® | Elmore Family School of Electrical and Computer Engineering

VIP Vertically Integrated Projects

The Team

William Tao
Computer
Engineering

Piotr Nabrzyski
Computer
Engineering

Donny Weintz
Computer
Engineering

Benjamin Nguyen
First Year
Engineering

PURDUE UNIVERSITY® | Elmore Family School of Electrical and Computer Engineering

Problem Statement

Estimating nutrition information while cooking remains a largely manual process, creating a demand for software that can interpret food preparation directly from video.

PURDUE UNIVERSITY® | Elmore Family School of Electrical and Computer Engineering

Project Objectives

Detection

Detect foods in a recipe and provide bounding box visualizations
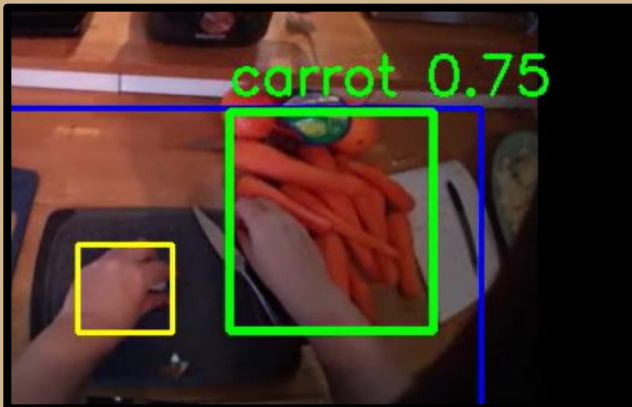
Nutrition Estimation

Automate nutrition change estimation via calorie count in each video frame

Recipe Prediction

Combine detection and nutrition estimation to reconstruct recipes

HD-Epic Dataset

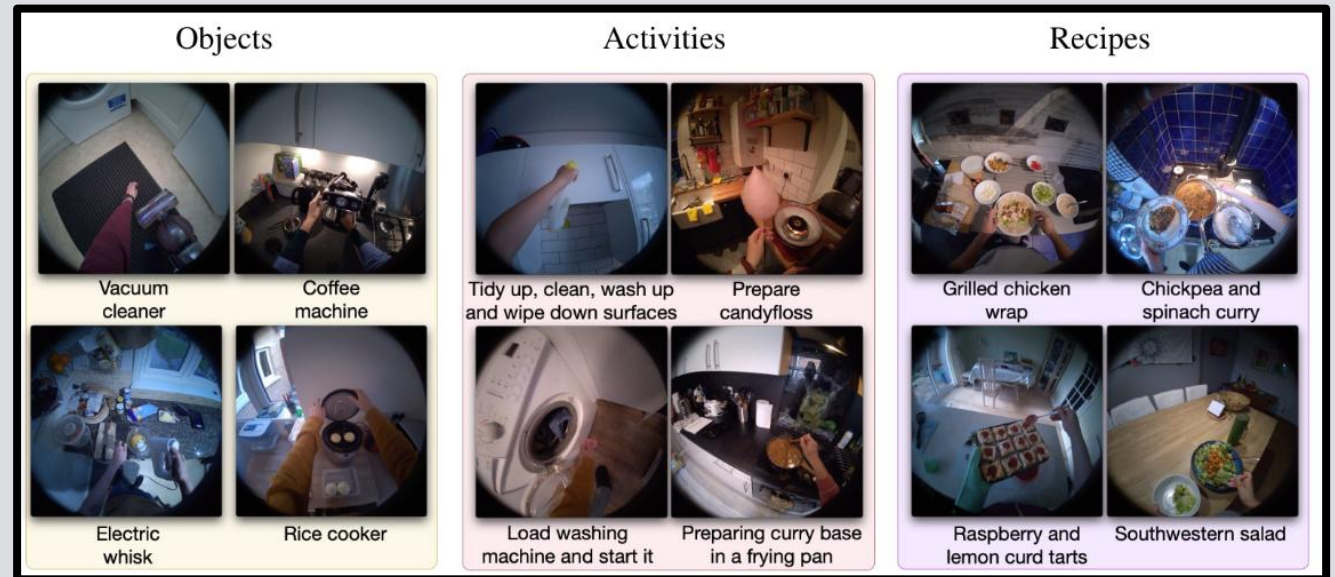## Purpose - Cookware Detection and Testing

- The HD-EPIC dataset contains **2.3 TB** worth of recipe videos and annotations recorded with Project Aria glasses

- Captures the following activities:

  - Fetching ingredients

  - Prepping/weighing ingredients

  - Adding ingredients to recipe

Image from HD-EPIC Dataset Participant 1

Image from HD-EPIC Dataset Participant 3

Objects	Activities	Recipes
Vacuum cleaner	Tidy up, clean, wash up and wipe down surfaces	Grilled chicken wrap
Coffee machine	Prepare candyfloss	Chickpea and spinach curry
Electric whisk	Load washing machine and start it	Raspberry and lemon curd tarts
Rice cooker	Preparing curry base in a frying pan	Southwestern salad

Dataset Description from HD-Epic Website

PURDUE UNIVERSITY® | Elmore Family School of Electrical and Computer Engineering

VFN Dataset

Purpose - Training Food Ingredient Detection

- The VIPER-FoodNet (VFN) dataset has 82 categories from WWEIA (What We Eat in America) with 14,991 images, including:

  - Image

  - Coordinates of locating/bounding box

  - Category of food within image

Example image from
VFN Dataset

Example image with
locating/bounding box

Description from VFN Website

Nutrition Dataset

Purpose - Nutrition Estimation

- Food and Nutrition Database for Dietary Studies (FNDDS) Includes:

  - Unique 8-digit Food Codes

  - 4-digit Food Category Codes from WWEIA (What We Eat in America) Categories

  - Calorie Estimation of Unique Food Codes

FNDDS Nutrient Values

2021-2023 Food and Nutrient Database for Dietary Studies - At A Glance

| Food code	Main food description	WWEIA Category number	WWEIA Category description	Energy (kcal)
42401010	Coconut milk, used in cooking	1902	Plant-based milk	230
42401100	Yogurt, coconut milk	1904	Plant-based yogurt	108
42401200	Yogurt, almond milk	1904	Plant-based yogurt	128
42402010	Coconut cream, canned, sweetened	1902	Plant-based milk	357
42403010	Coconut water, unsweetened	7204	Fruit drinks	18
42404010	Coconut water, sweetened	7204	Fruit drinks	37
42500000	Trail mix, NFS	2804	Nuts and seeds	454
42500100	Trail mix with nuts	2804	Nuts and seeds	604
42501000	Trail mix with nuts and fruit	2804	Nuts and seeds	454
42501500	Trail mix with chocolate	2804	Nuts and seeds	503
42502100	Trail mix with pretzels, cereal, or granola	2804	Nuts and seeds	476

Description from Agricultural Research Service
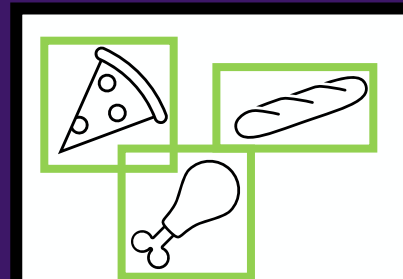
Proposed Solution
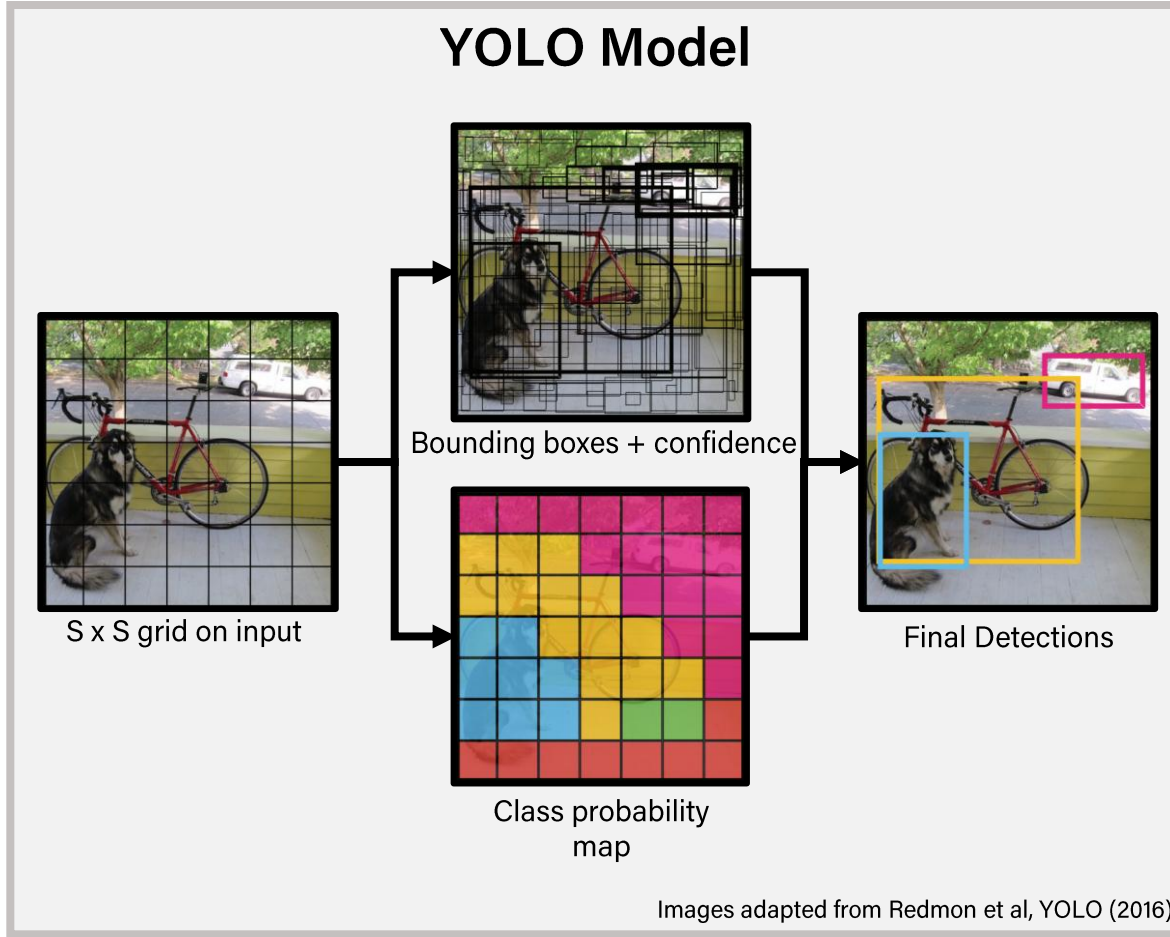
System Overview
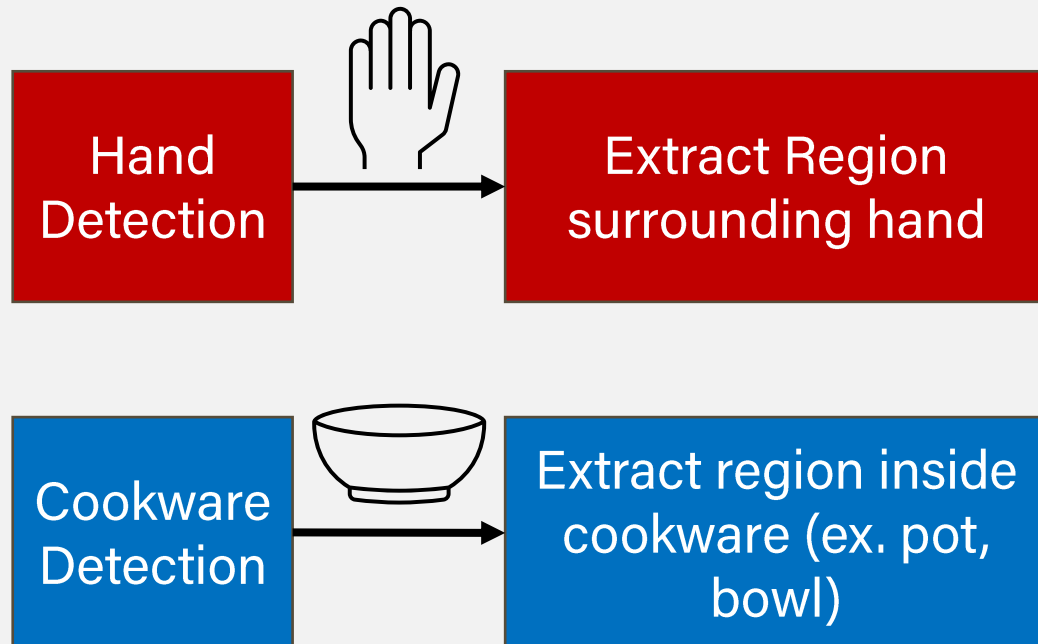
Detection Methodology

Hand Detection – MediaPipe Hands

- **MediaPipe Hands** was published by Google in 2020
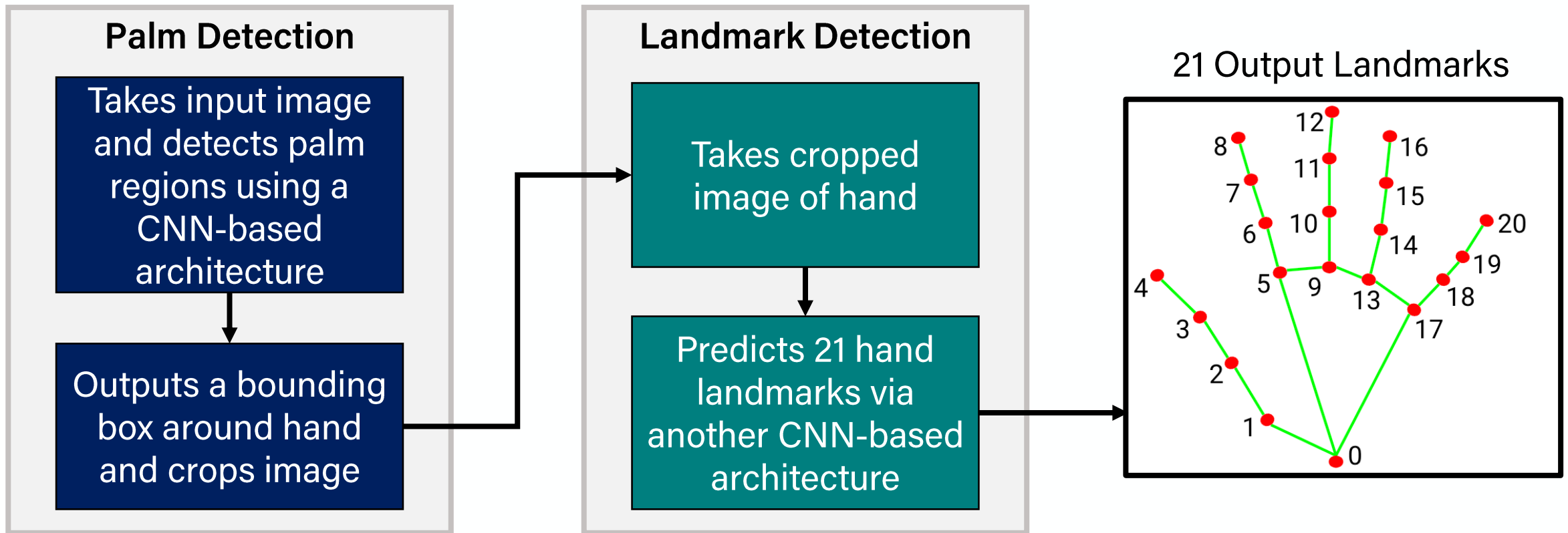
Detection Methodology

Ingredient and Cookware Detection – YOLO (You only look once)

YOLO Model

S x S grid on input

Bounding boxes + confidence

Class probability map

Final Detections

Images adapted from Redmon et al, YOLO (2016)

Our Implementation – Cookware Detection

- Train model with data from HD-Epic
  - **23** preparation tool classes
  - **2869** instances

Our Implementation – Ingredient Detection

- Train model with VIPER Food Net (VFN)
  - **58** ingredient classes
  - **16130** instances

Video Demo

Video Legend	
🟨	Hand Detection (Yellow)
🟦	Food Prediction Region (Cyan)
🟩	Ingredient Detection (Green)
🟪	Cookware Detection (Pink)

Evaluation

Food Detection with YOLO

Overall Metrics

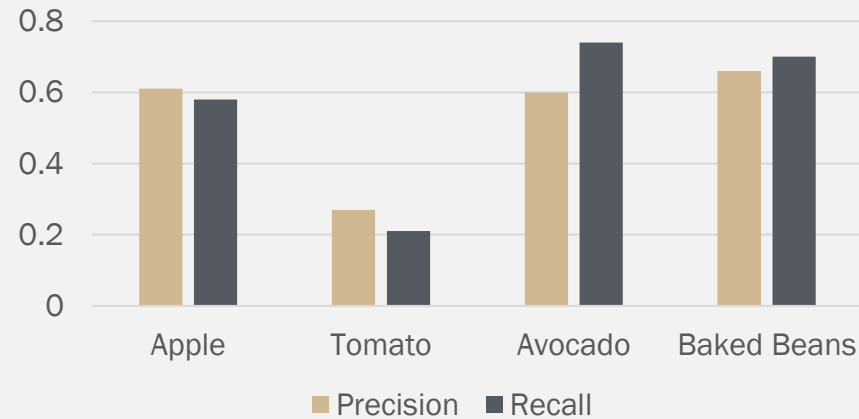| Metric	Value
Precision	0.58
Recall	0.49
mAP50	0.53

Cantaloupe   Onion

Ingredient Detection Comparison

Key Takeaways
- The model performs better on ingredient classes with distinct features
- If a food ingredient in new data isn't in the set of classes, it is often misclassified as a similar looking food
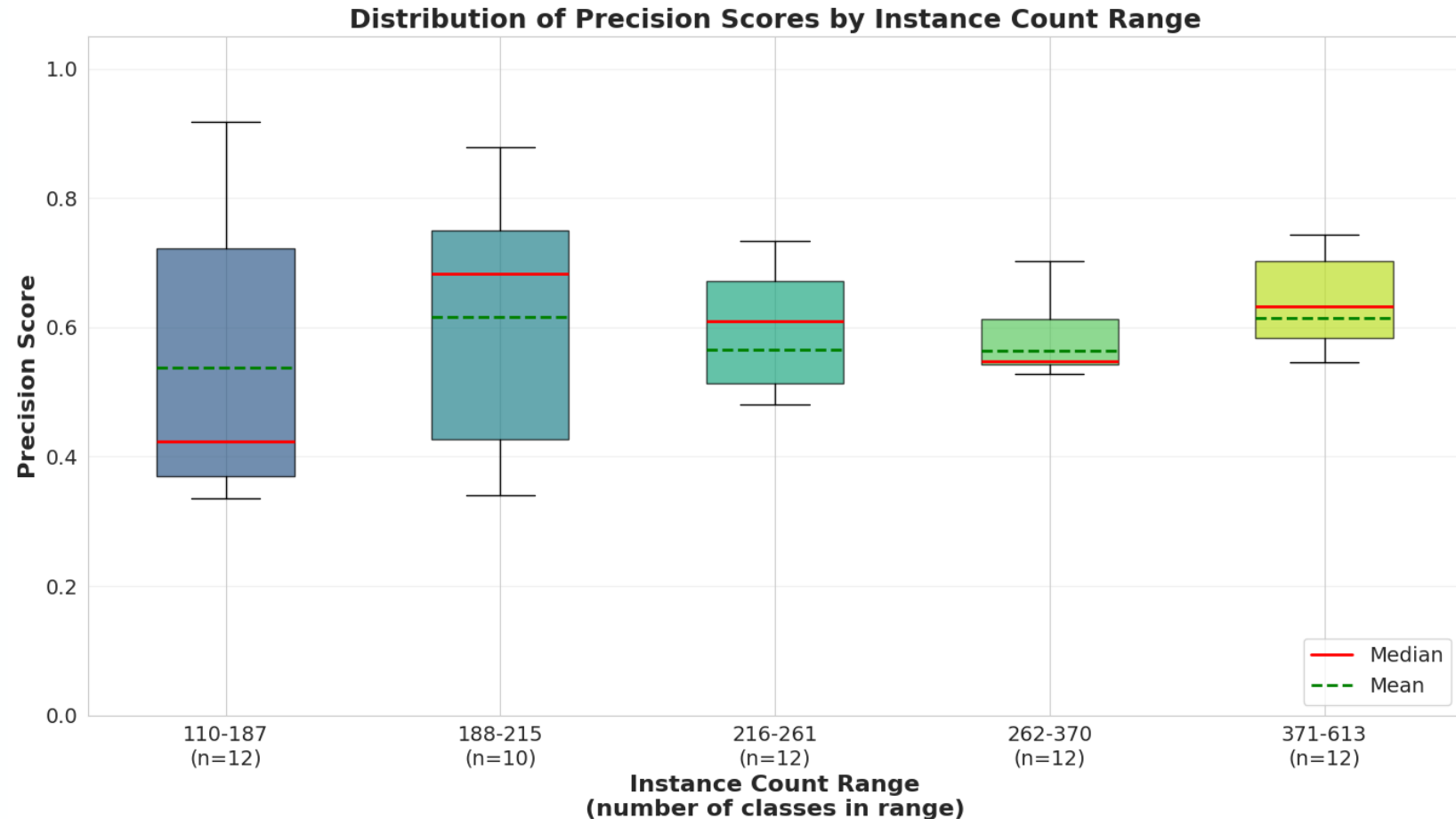
Evaluation

Precision versus Instances

Key Takeaways

- No clear upward trend

- Slight increase in consistency
  - "Tighter" spread (IQR)

- Goes back to classes with distinct features dominating performance rather than number of instances

Distribution of Precision Scores by Instance Count Range

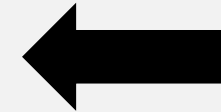# *Evaluation*

Calorie Mapping

- The FNDDS dataset labels calorie values *__per 100 grams__* for a given ingredient

  - We can't predict the quantity of ingredients

  - Some mappings may also not being present or require canonicalization

We don't know how many grams of butter is being detected

Records 4 chicken breasts as 307 total calories

Reflection and Future Work

- **Reflection:**

Strengths

• Fairly consistent general object detection
    • YOLO for cookware and ingredients
• Stable hand tracking
    • MediaPipe Hands for hand overlaps

Weaknesses

• Insufficient training data
    • Domain translation
    • Mismatches in classes
• Inability to estimate quantity of ingredients
    • Incorrect calorie mappings

- **Future work:**

Quantity of Ingredient Added	Improving Ingredient Lexicon	Recipe Mapping
Use distance and relative size of ingredient to estimate quantity for better calorie mapping	Canonicalize ingredient names E.g., meat_loaf = ground beef parmesan = cheese	Map ingredients and preparation tools to RecipeNLG dataset to find best match for recipe in video

PURDUE UNIVERSITY® | Elmore Family School of Electrical and Computer Engineering

References

[1] T. Perrett *et al.*, "HD-EPIC: A Highly-Detailed Egocentric Video Dataset," *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, TN, USA, 2025, pp. 23901-23913, doi: 10.1109/CVPR52734.2025.02226.
Available: https://hd-epic.github.io/index#starthere

[2] F. Zhang, V. Bazarevsky, A. Vakunov, A. Tkachenka, G. Sung, C. Chang, and M. Grundmann, "MediaPipe Hands: On-device real-time hand tracking," arXiv preprint arXiv:2006.10214, 2020. [Online].
Available: https://arxiv.org/pdf/2006.10214

[3] M. Bień, M. Gilski, M. Maciejewska, W. Taisner, D. Wiśniewski, and A. Lawrynowicz, "RecipeNLG: A Cooking Recipes Dataset for Semi-Structured Text Generation," in Proc. 13th Int. Conf. Natural Language Generation (INLG), Dublin, Ireland, Dec. 2020, pp. 22–28. [Online].
Available: https://www.aclweb.org/anthology/2020.inlg-1.4

[4] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 779-788, doi: 10.1109/CVPR.2016.91.

[5] R. Mao, J. He, Z. Shao, S K. Yarlagadda, Z. Fengqing, "VIPER-FoodNet (VFN) Dataset," Purdue University, West Lafayette, IN, USA. [Online].
Available: https://lorenz.ecn.purdue.edu/~vfn/

Elmore Family School of Electrical
and Computer Engineering

Thank You

Any Questions?

Total Calories: 544 kcal

Detected Ingredients:

shrimp: 146 kcal
steak: 229 kcal
meat_loaf: 0 kcal
strawberries: 54.33 kcal
carrot: 115.17 kcal

Video Legend

- Hand Detection (Yellow)
- Food Prediction Region (Cyan)
- Ingredient Detection (Green)
- Cookware Detection (Pink)

carrot