Multi-Food Items Detection and Localization in Images

Matthew Morgan

Supervised by Patrick Donnelly



Oregon State University

NCUR 2022

Introduction

- Patients with malnutrition are 50% more likely to be readmitted to the hospital within 100 days [1]

- In the United States, it is estimated that 31 percent of the available food supply went uneaten in 2010 [2]
  - 133 billion pounds of food waste
  - $161 billion at retail prices

- **Goal**: track nutritional intake and food waste in hospitals, schools, and correctional facilities

Foundational Aim

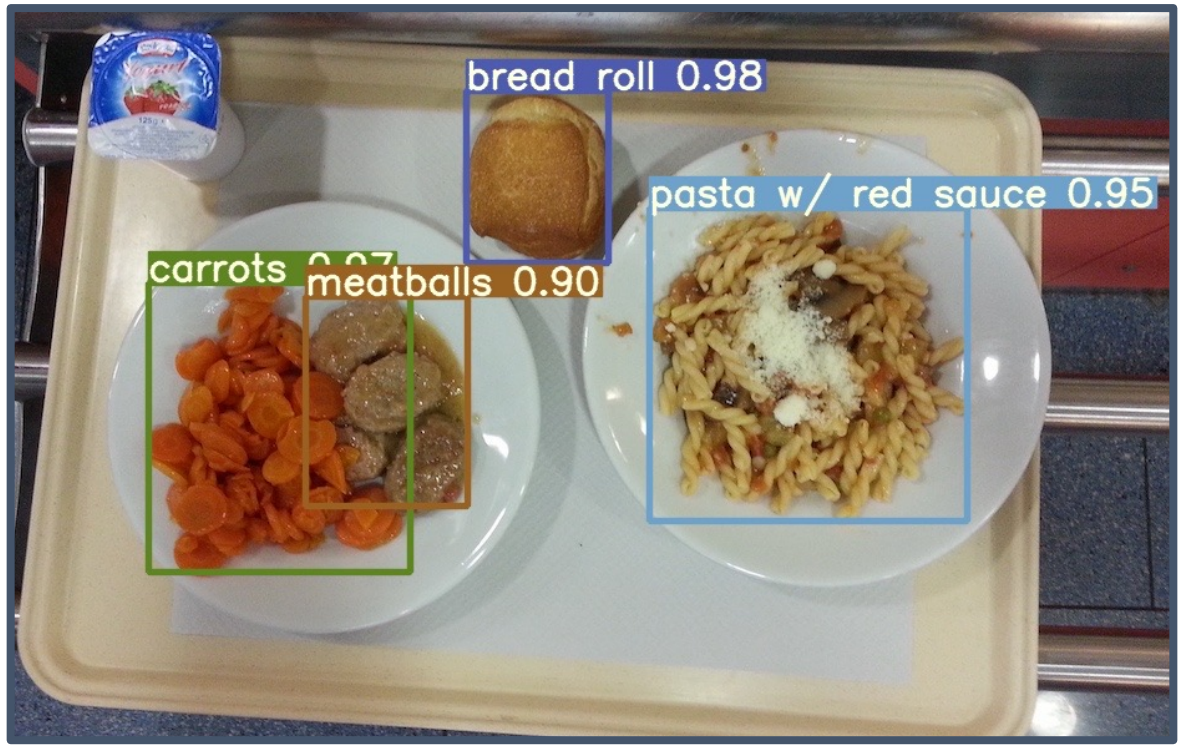


Figure 1: Example detection from our model

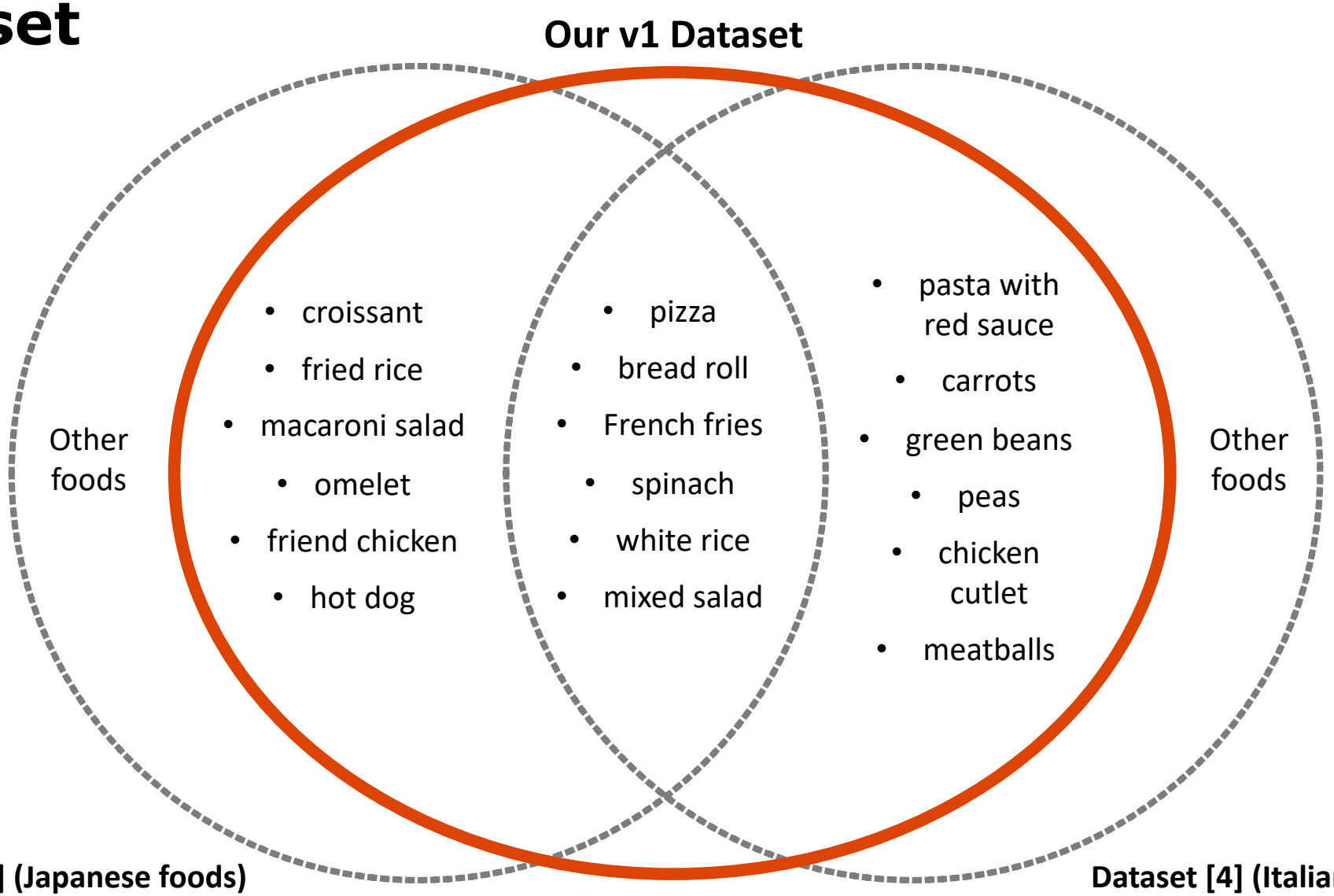


Dataset

Small American cafeteria food dataset of annotated meal images

- Select 12 foods from dataset [3] (100 primarily Japanese food categories)
- Select 12 from dataset [4] (73 primarily Italian food categories)
- 2891 images, 3839 targets, $n = 18$ classes
- 70 / 20 / 10% stratified split into train, test, and validation sets

Dataset


Our v1 Dataset
Other foods
croissant
fried rice
macaroni salad
omelet
friend chicken
hot dog
pizza
bread roll
French fries
spinach
white rice
mixed salad
pasta with red sauce
carrots
green beans
peas
chicken cutlet
meatballs
Other foods
Dataset [3] (Japanese foods)
Dataset [4] (Italian foods)

Dataset

Table 1: total target item count from dataset [3] and [4]

Food Name	Count from [3]	Count from [4]	Total Count
white rice	620	57	677
bread roll	107	479	586
mixed salad	342	46	388
pizza	134	89	223
French fries	153	64	217
spinach	102	110	212
pasta with red sauce	0	254	254
carrots	0	161	161
chicken cutlet	0	148	148
green beans	0	131	131
peas	0	36	36
meatballs	0	45	45
fried rice	169	0	169
fried chicken	154	0	154
croissant	120	0	120
macaroni salad	109	0	109
omelet	107	0	107
hot dog	102	0	102
TOTAL	**2219**	**1620**	**3839**

Experiment

Train the state-of-the-art YOLOR model [5]

- Train for 300 epochs
- Image size = 1280 x 1280
- Batch size = 8
- IoU training threshold = 0.2

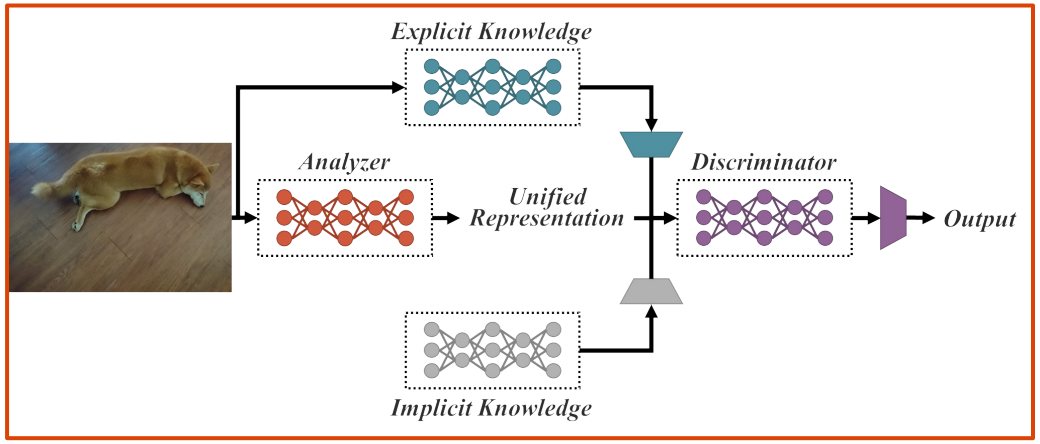


Figure 2: Diagram of YOLOR architecture [5]

Results

- 0.917 overall mAP@0.5
- 0.721 overall mAP@0.95

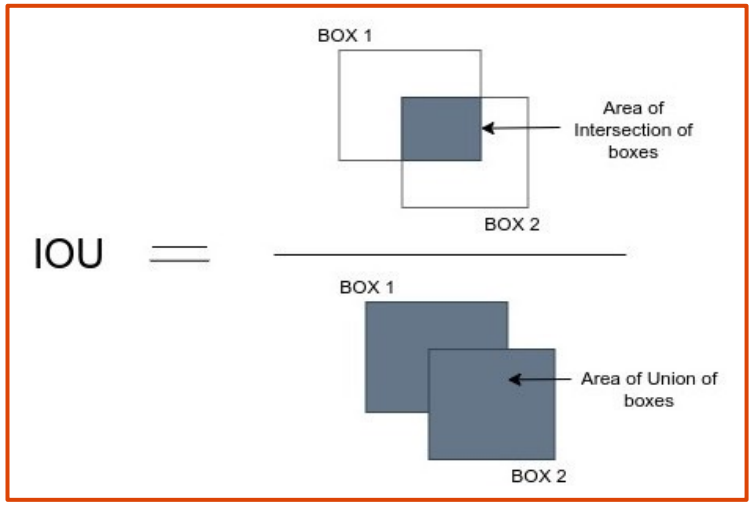


Figure 3: IoU explained [6]

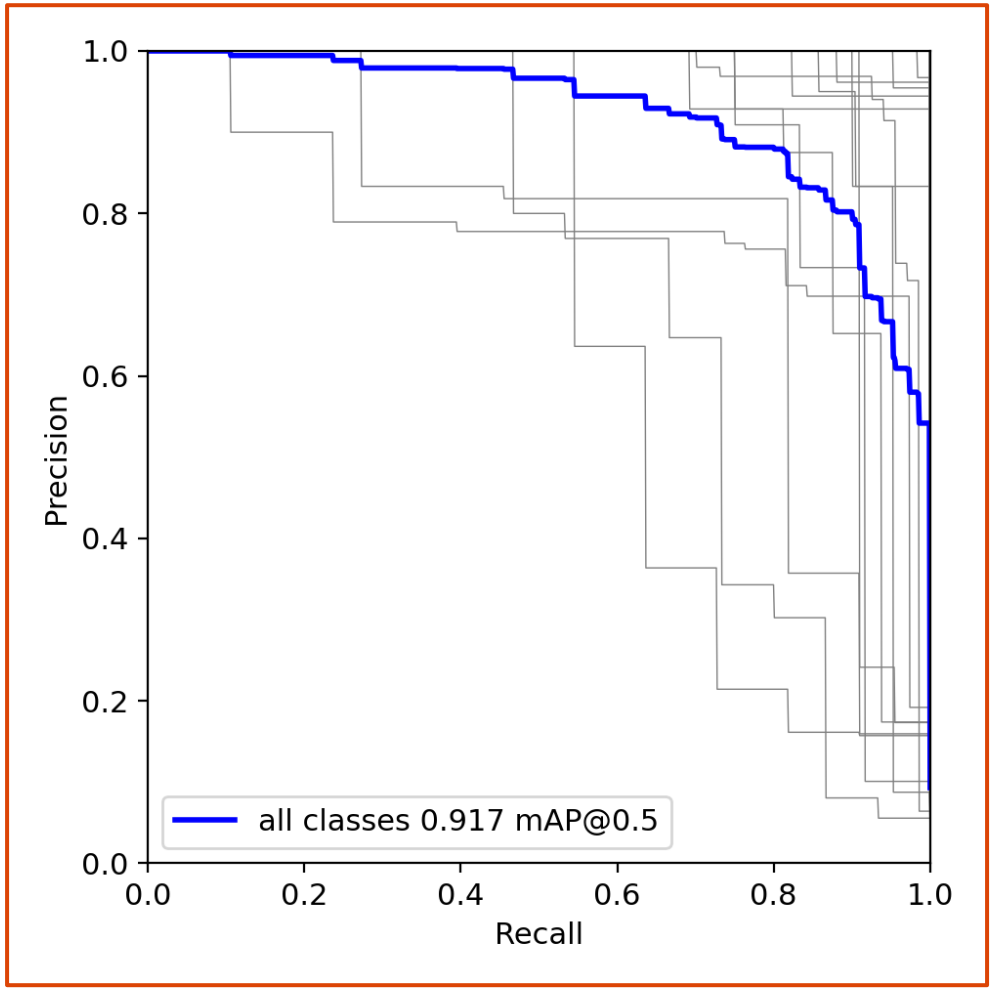


Figure 4: Overall mean average precision @ 0.5 IoU

Results

- data from [4]
- data from [3]
- data from both [3] and [4]

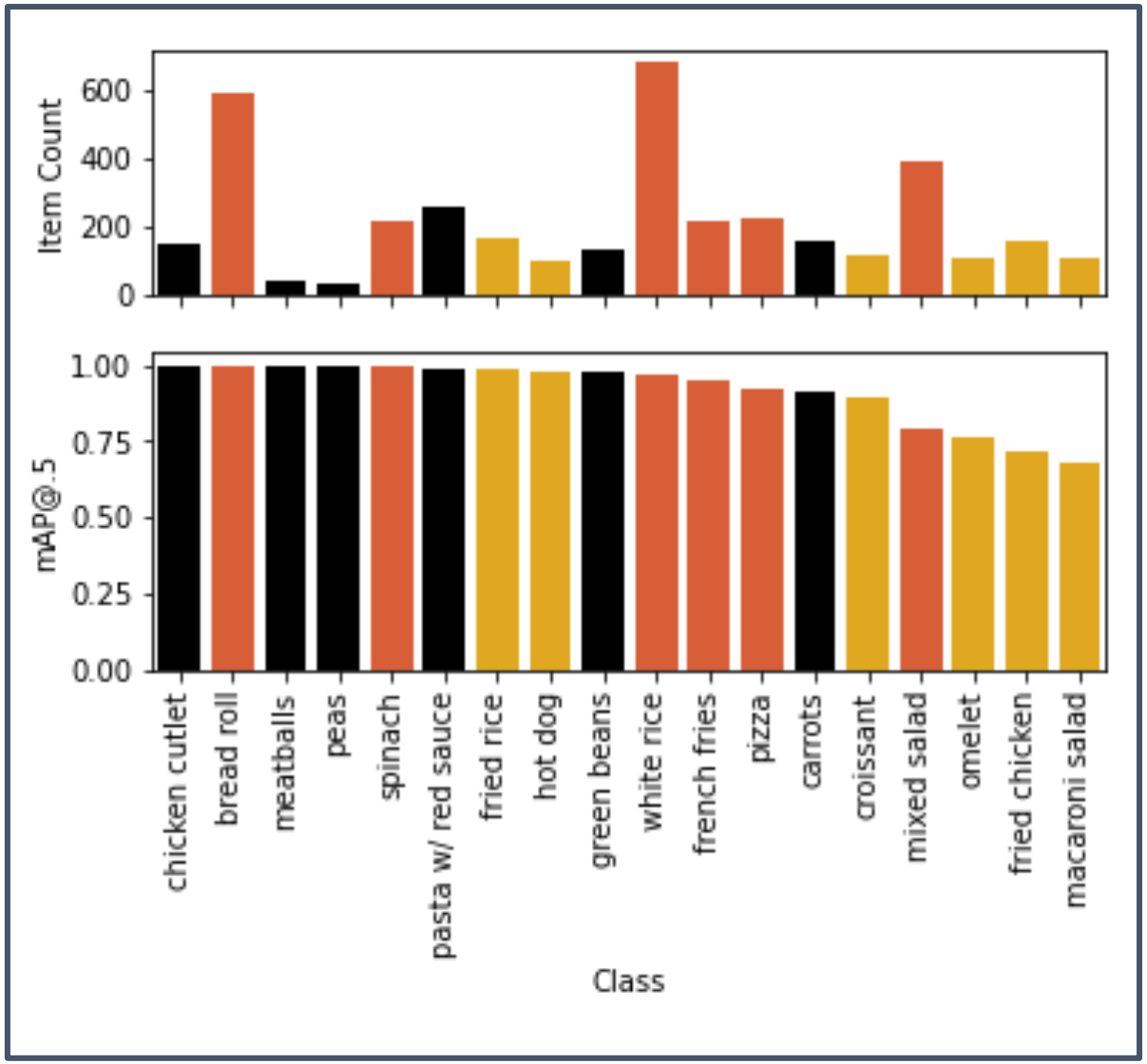


Figure 5: Per-class mAP@0.5 IoU

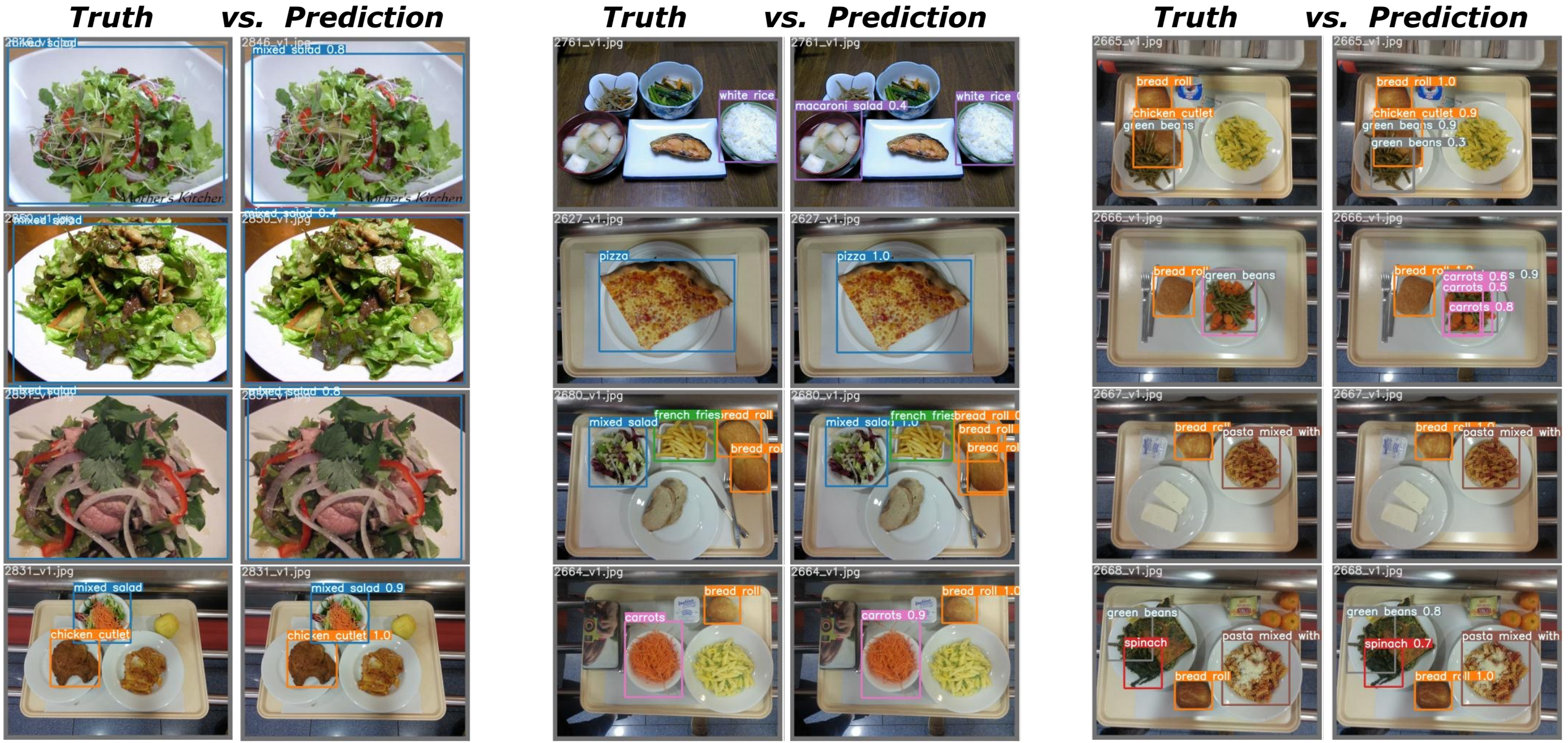


Figure 6: Comparison between truth and predictions on a selection of images from the test set

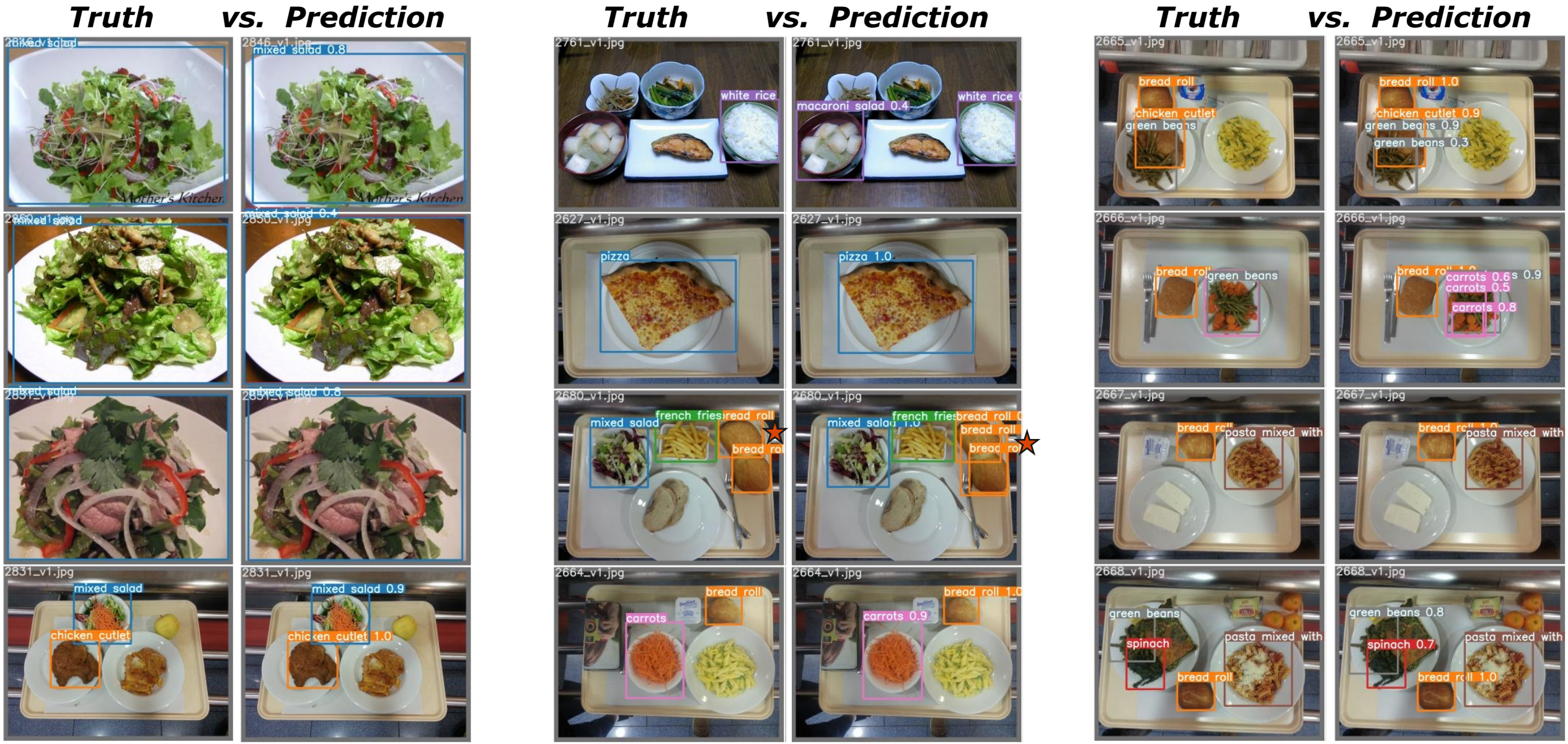


Figure 6: Comparison between truth and predictions on a selection of images from the test set

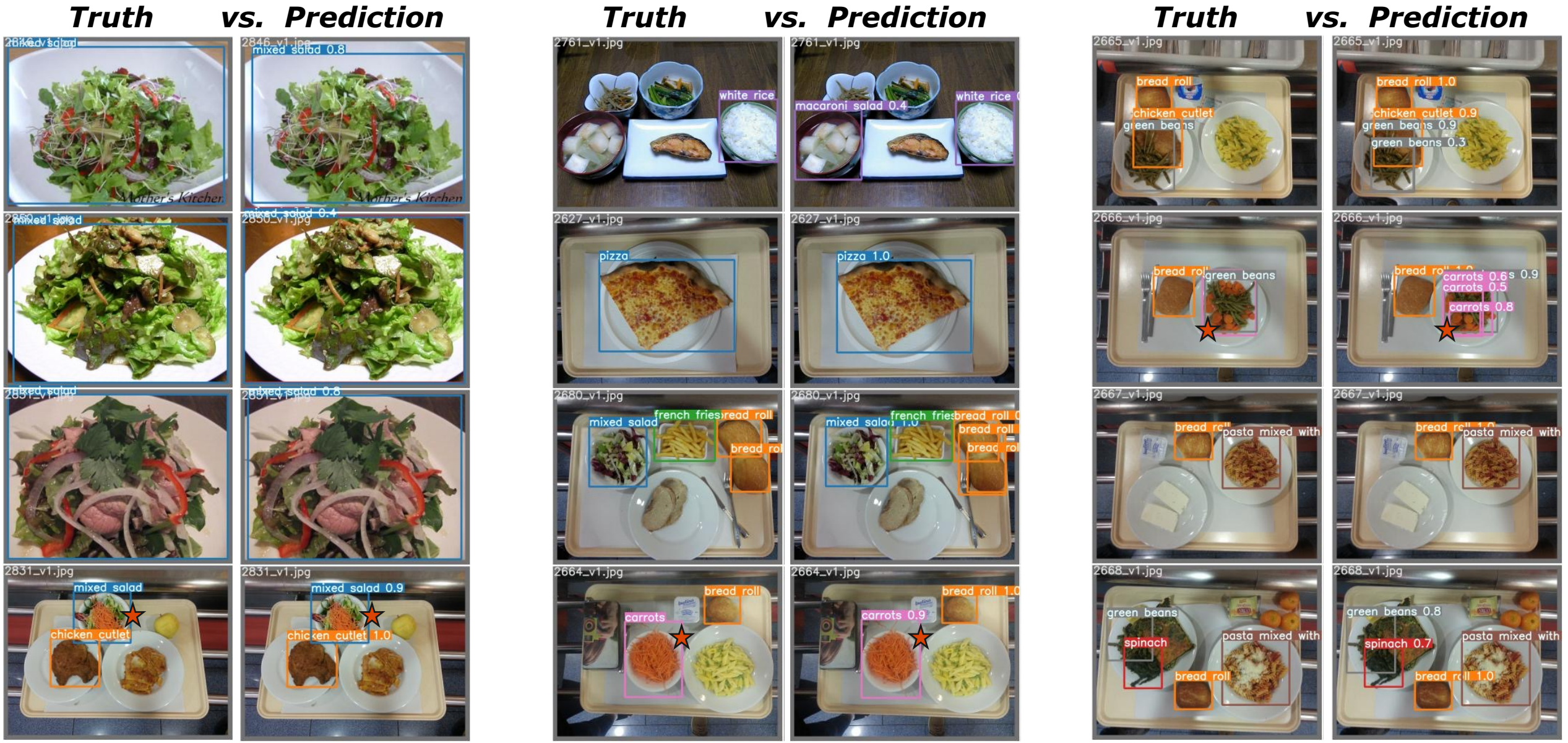


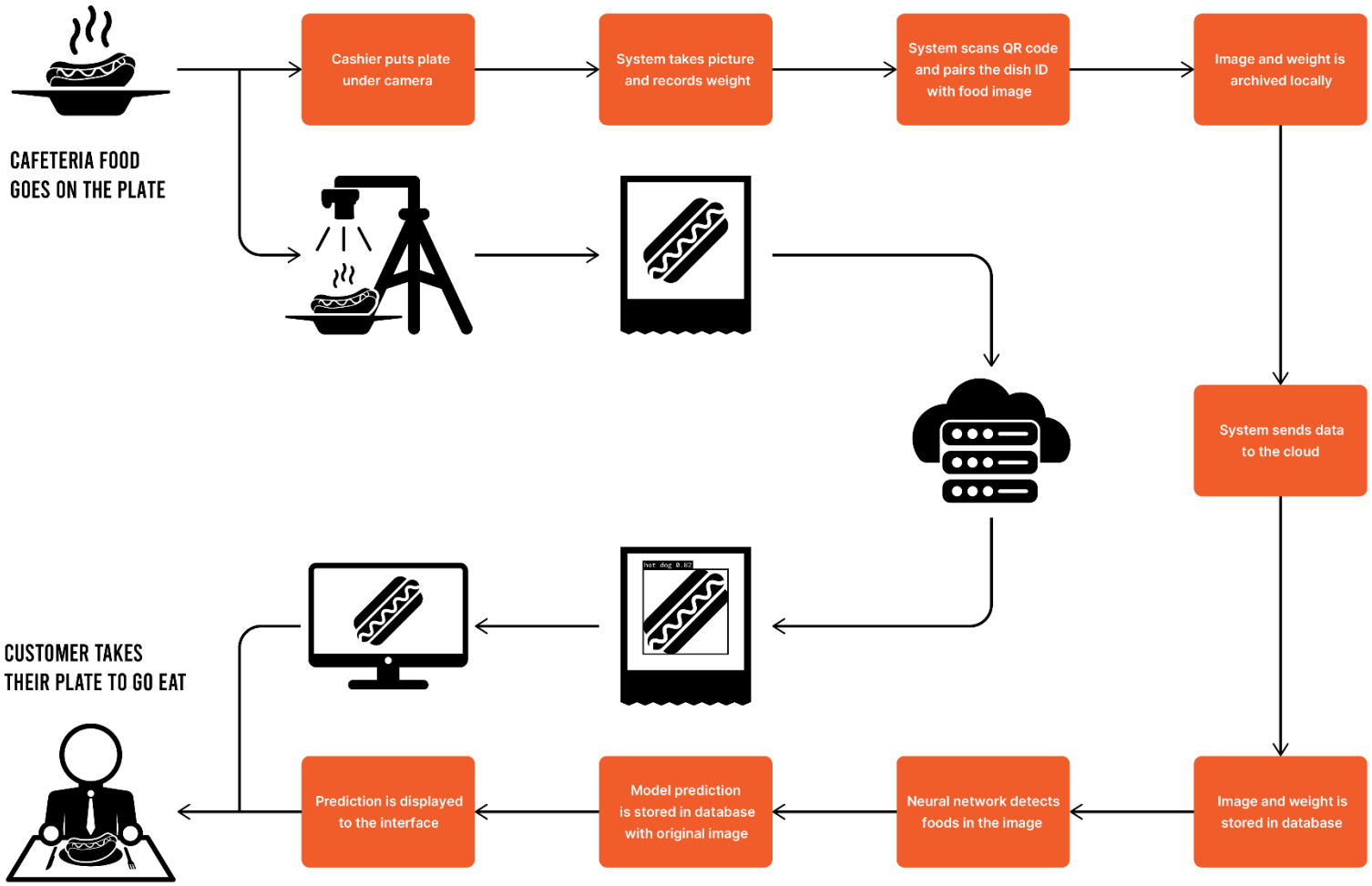
Figure 6: Comparison between truth and predictions on a selection of images from the test set

Conclusion & Future Work

- Aggregating images from disparate datasets is an effective method for developing a food detection model with high accuracy

- We plan to produce a more thorough annotated American food image dataset starting with a small data collection study in a university cafeteria in late April 2022

Conclusion & Future Work



Works Cited

[1] K. R. Fingar *et al.*, “All-Cause Readmissions Following Hospital Stays for Patients With Malnutrition, 2013,” *Agency for Healthcare Research and Quality*. https://hcup-us.ahrq.gov/reports/statbriefs/sb218-Malnutrition-Readmissions-2013.jsp (accessed Dec. 03, 2021).

[2] J. C. Buzby, H. Farah-Wells, and J. Hyman, “The estimated amount, value, and calories of postharvest food losses at the retail and consumer levels in the United States,” *USDA-ERS Econ. Inf. Bull.*, no. 121, 2014.

[3] Y. Matsuda, H. Hoashi, and K. Yanai, “Recognition of Multiple-Food Images by Detecting Candidate Regions,” 2012.

[4] G. Ciocca, P. Napoletano, and R. Schettini, “Food recognition: a new dataset, experiments and results,” *IEEE J. Biomed. Health Inform.*, vol. 21, no. 3, pp. 588–598, 2017, doi: 10.1109/JBHI.2016.2636441.

[5] C.-Y. Wang, I.-H. Yeh, and H.-Y. M. Liao, “You Only Learn One Representation: Unified Network for Multiple Tasks,” *ArXiv Prepr. ArXiv210504206*, 2021.

[6] S. Vineeth S, “IOU (Intersection over Union). What is IOU?,” Medium. https://medium.com/analytics-vidhya/iou-intersection-over-union-705a39e7acef (accessed Mar. 25, 2022).