Nutrition5k: Towards Automatic Nutritional Understanding of Generic Food Supplementary Material

Figure 1. A single frame from one Nutrition5k dish, with fine-grained ingredient annotations (Table 1).

Ingredient	Mass
Pork	121.1g
Fried rice	109.0g
Brussels sprouts	66.9g
Arugula	48.7g
Mustard greens	29.8g
White rice	16.1g
Mangoes	14.9g
Cucumbers	14.9g
Tomatoes	14.9g
Cherry tomatoes	14.9g
Onions	8.6g
Zucchini	8.1g
Olive oil	3.0g
Lemon juice	2.2g
Lime	2.2g
Jalapenos	1.5g
Cilantro	1.5g
Rosemary	0.8g
Garlic	0.8g
Salt	0.7g
Parsley	0.4g
Pepper	0.1g

Table 1. Ingredient annotations associated with Figure 1.

A. Additional Dataset Examples

In this supplementary section, we provide more detailed and complete exemplars from the proposed *Nutrition5k* dataset. Figures 2 and 3 show frames from RGB videos, depth data, and some ground truth annotations associated with 12 unique dishes (of the roughly 5000 present in Nutrition5k). Omitted from these figures are the verbose annotations of individual ingredients present in each dish along with quantities. Instead, we provide Figure 1 and Table 1 to show the granularity of these ingredient annotations. We hope that these additional examples will help to convey the high quality and realistic nature of our data, and provide visualizations of the challenges presented in Section 3.6 of our paper submission.

B. Illustration Of Incremental Scanning

Figure 4 shows an example of the incremental scanning process used during data collection. One recipe (food item) is added at a time, followed by a complete device scan that captures RGB videos around the plate, depth data, and ingredient and mass annotations. Each incremental scan is represented as a unique dish within Nutrition5k, to provide more variety in our dataset. Note that our train and test sets, Nutri-Train and Nutri-Test, respect the relationship between incremental scans; all dishes belonging to the same incremental scan will exist in the same split.

Figure 2. Some examples from the proposed *Nutrition5k* dataset. Each row represents data associated with a single dish (of the roughly 5000 dishes). The first and second images are frames taken from the 30° and 60° videos, respectively. These frames represent only a single view that the rotating cameras see during their 360° capture of the plate. The third and fourth images are taken from the overhead Intel RealSense depth camera (depth is presented in RGB for better visualization). We also show some of the ground truth annotations associated with each dish.

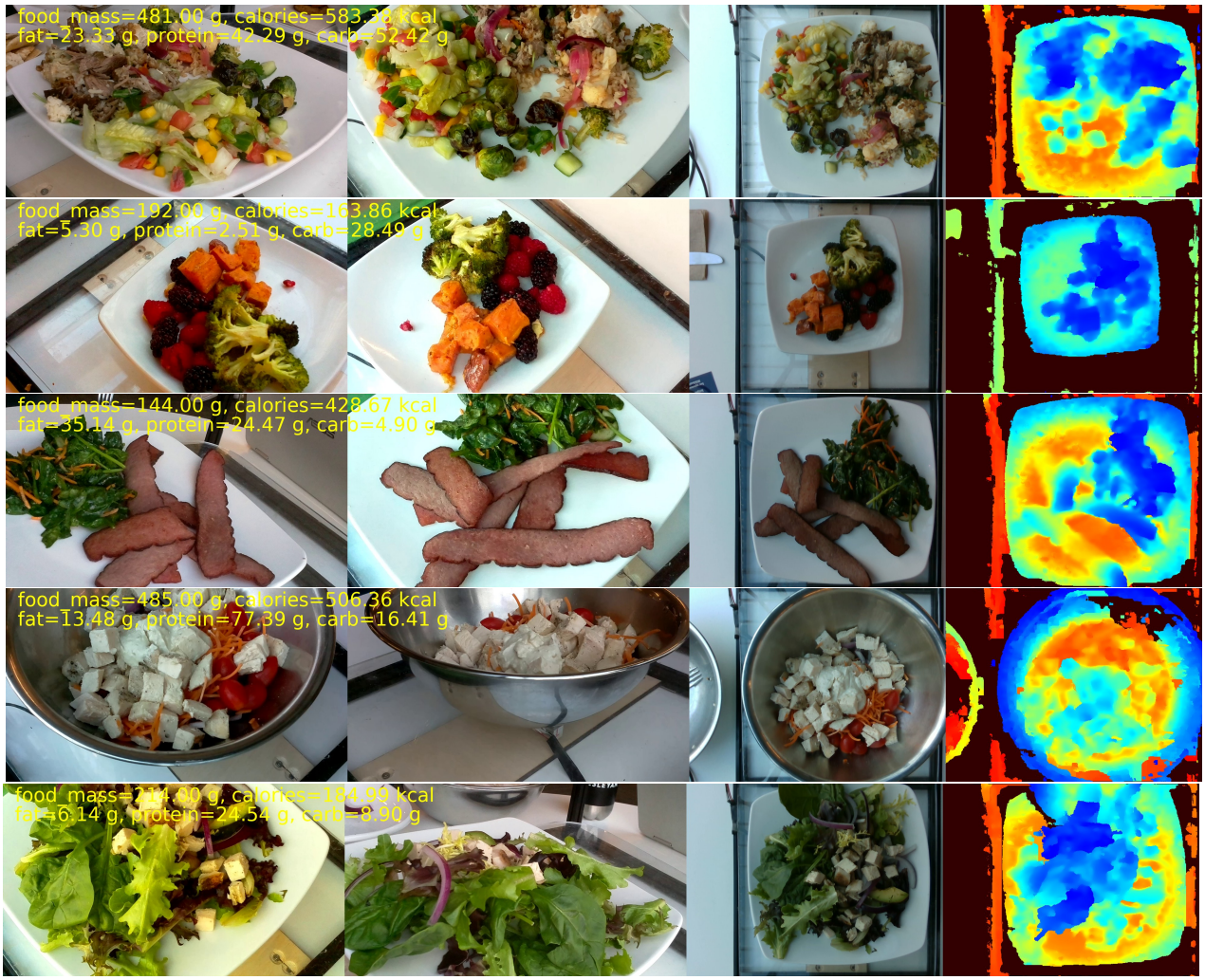

Figure 3. More examples from the proposed *Nutrition5k* dataset.

Figure 4. Illustration of the incremental scanning procedure for a single plate. Images were taken from *Nutrition5k*.