

# UniFood: A Unified Food Dataset for Comprehensive Food and Nutrition Analysis

Pengkun Jiao<sup>1</sup>, Xinlan Wu<sup>1</sup>, Bin Zhu<sup>2</sup>,  
Jingjing Chen<sup>1</sup>, Chong-Wah Ngo<sup>2</sup>, Yugang Jiang<sup>1</sup> \*

## A Prompt Design

The prompt template is as follows:

“[input image] In this image, [known information1]. [known information2]... Please tell me the [missing information1], [missing information2]... Return N/A if you don’t know the answer. please offer numerical estimates for the quantity, calories, mass, fat, carbohydrate, and protein content of each ingredient without any explanatory reasoning. Use common sense for uncertain quantities to ensure the response remains solely numerical. Format as follows: [example].”

Here, “[input image]” is the food image, “[known information]” represents the annotation from the source dataset, “[missing information]” is the lacking information that the source annotation cannot provide for comprehensive food and nutrition analysis, and “[example]” provides an example of the response format.

Figures 1, 2, and 3 illustrate examples of prompts and responses from ChatGPT-4V on Food-101 [1], VireoFood-251 [2], and Recipe1M [4]. The instances are randomly sampled.

## B Effectiveness of Vocabulary Standardization

Table 1 presents the performance with and without vocabulary standardization. The results indicate that applying vocabulary standardization improves overall performance across multiple tasks.

## C Dataset Quality Analysis

### C.1 Quality analysis on Nutrition Anotation

As manually checking the nutritional reliability of the entire UniFood dataset is costly, we sampled a subset to test the nutritional reliability. We randomly selected 1000 samples from the human-checked subset (please refer to Sec. 3.6) and used ChatGPT-4V to regenerate nutritional information for these samples. We then calculated the mean absolute error (MAE) and percent Mean Absolute Error (pMAE) between the human annotations and the ChatGPT-4V annotations. The results are shown in Table 2. It is clear that ChatGPT-4V can provide accurate nutrition estimations given the food images and quantified ingredient list (see Figure 3 for details), with only slight differences from human annotations.

### C.2 Performance Comparison with ChatGPT-4V

We modified our annotation collection code to enable querying ChatGPT-4V for information related to a given image. However, conducting a comprehensive evaluation would be resource-intensive. To balance feasibility and insight, we randomly sampled 100 instances from the precise nutrition test set and queried ChatGPT-4V for both recipe and nutritional information. Titles and ingredients

#### Prompt:

In this image, the food is beet salad. Please tell me the ingredients, recipe and nutrition information of the food, return N/A if you don't know that answer. For nutrition, please offer numerical estimates for the quantity, calories, mass, fat, carbohydrate, and protein content of each ingredient without any explanatory reasoning. Use common sense for uncertain quantities to ensure the response remains solely numerical.

For example:

#### #Ingredients#

-tomato

-sugar

#### #Recipe#

1. cut tomato; 2. spread sugar on the tomato.

#### #Nutrition#

name; amount; mass; calories; fat; carbs; protein

Tomato; 4 slices; 98g; 22kcal; 0.42g; 3.84g; 0.7g



Image input

#### Response:

#### #Ingredients#

-Beets

-Arugula

-Red onion

-Pistachios

-Olive oil

-Salt

-Pepper

#### #Recipe#

1. Roast beets until tender, peel, and slice.

2. Arrange arugula on a plate.

3. Add beet slices and thinly sliced red onion.

4. Sprinkle with chopped pistachios.

5. Drizzle with olive oil and season with salt and pepper.

#### #Nutrition#

name; amount; mass; calories; fat; carbs; protein

Beets; 1 beet; 82g; 35kcal; 0.13g; 7.95g; 1.43g

Arugula; 1 cup; 20g; 5kcal; 0.1g; 0.7g; 0.5g

Red onion; 1/4 medium; 28g; 11kcal; 0.03g; 2.6g; 0.28g

Pistachios; 1 tbsp; 9g; 55kcal; 4.4g; 2.8g; 2g

Olive oil; 1 tbsp; 13.5g; 119kcal; 13.5g; 0g; 0g

Salt; to taste; 1g; 0kcal; 0g; 0g; 0g

Pepper; to taste; 1g; 3kcal; 0.04g; 0.74g; 0.12g

Figure 1: Prompt and Response Example on Food-101.

were excluded to minimize additional vocabulary normalization. The results are presented in Table 3.

### C.3 Case Study

We present some instances from the UniFood dataset and compare them with recipes from the internet, as illustrated in Figures 4, 5 and 6.

### C.4 Nutrition Annotation Comparison with Recipe1M+

We conduct a comparison of shared samples between UniFood and Recipe1M+ [3], as illustrated in Figure 7 and Figure 8. The potentially inaccurate source annotation in Recipe1M could affect the nutritional analysis. As shown in Figure 7, the incorrect ingredient amount annotation in Recipe1M led to overcalculated nutrition values for "salt" and "pepper" in the Recipe1M+ dataset. However, in the case of UniFood, the ChatGPT4V has the capability to identify and correct errors, thereby producing more reliable annotations.

## D Quality Results

To offer a more comprehensive understanding of our approach and its performance, we present several example cases in Figure 9. s

\*<sup>1</sup> Fudan University, <sup>2</sup> Singapore Management University

	Food Classification (Acc)	Ingredient Recognition (IoU)	Ingredient Recognition (F1)	Recipe Generation (ScareBLEU)	Recipe Generation (Rouge-L)	Nutrition Estimation (pMAE)
w/o standardization	67.44	31.02	44.47	5.47	35.27	111.85
w standardization	68.30	34.03	47.04	5.55	35.69	99.16

Table 1: Effect of standardization on model performance across multiple tasks

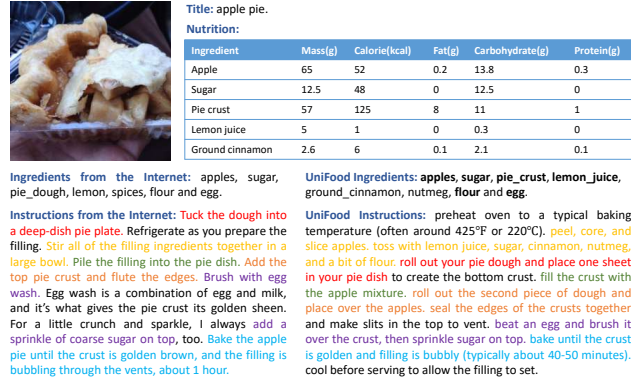


Figure 4: A Comparison between a UniFood Instance and Its Corresponding Recipe from <https://sallysbakingaddiction.com/apple-pie-recipe/>. The source dataset for this instance is Food-101.

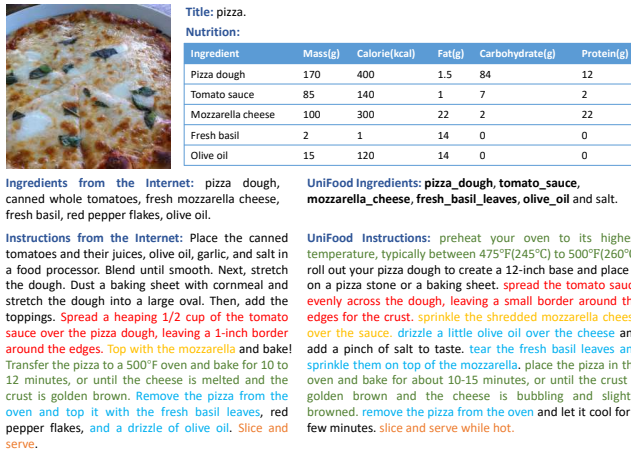


Figure 5: A Comparison between a UniFood Instance and Its Corresponding Recipe from <https://www.loveandlemons.com/margherita-pizza/>. The source dataset for this instance is Food-101.

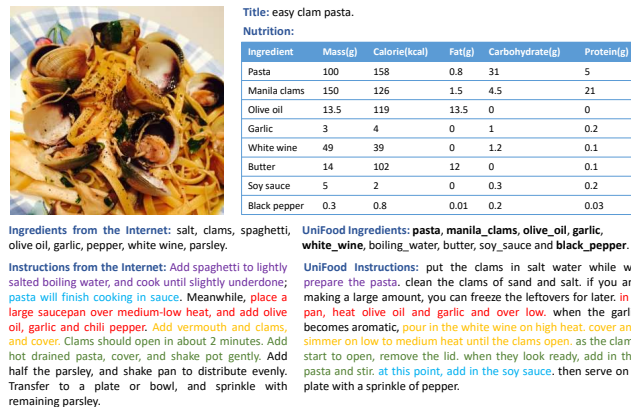


Figure 6: A Comparison between a UniFood Instance and Its Corresponding Recipe from <https://cooking.nytimes.com/recipes/11952-spaghetti-with-clams>. The source dataset for this instance is Recipe1M.

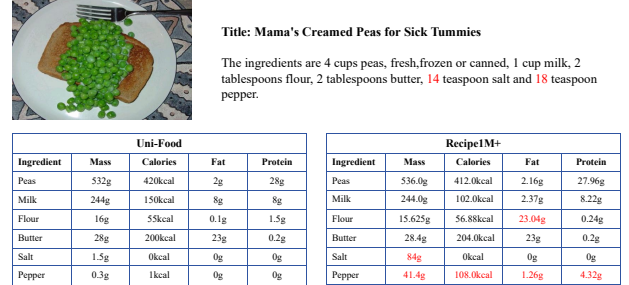


Figure 7: Comparison of Nutrition Annotation between UniFood and Recipe1M+

Table 2: Error between Human and ChatGPT-4V Annotations

Metric	Mass (g)	Calories (kcal)	Fat (g)	Protein Value (g)
MAE	13.22	54.42	2.67	2.39
pMAE	1.77%	2.7%	2.59%	6.91%

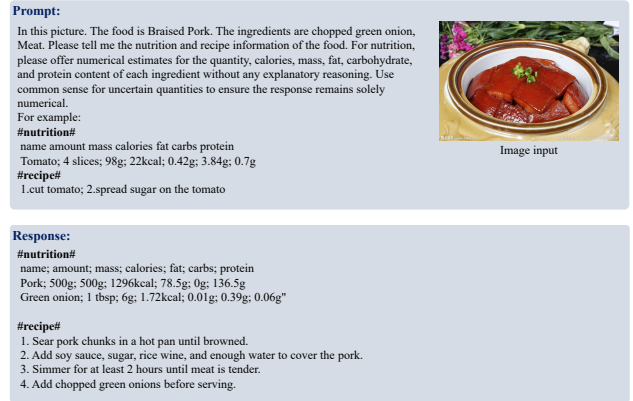


Figure 2: Prompt and Response Example on VireoFood-251.

**Table 3: Comparison of model performance across tasks**

	Unable to parse prediction	Recipe Generation (ScaREBLEU)	Recipe Generation (Rouge-L)	Nutrition Estimation
ChatGPT4-vision	3	3.82	30.44	62.39
UniFood	–	5.35	35.68	70.62



**Title: Chef Joey's Oatmeal, Fruit & Nut Bars**

The ingredients are 12 cup vegan margarine (or 1/2 cup light oil), 23 cup packed brown sugar, 1 12 cups oats (not instant), 12 cup spelt flour (white), 1 cup hazelnuts (toasted and chopped fine), 1 cup golden raisin and 12 cup tart jam preserves.

Uni-Food				
Ingredient	Mass	Calories	Fat	Protein
Vegan margarine	112g	814kcal	91g	0g
Brown sugar	145g	571kcal	0g	0g
Oats	120g	450kcal	7.5g	15g
Spelt flour	60g	200kcal	1.5g	8g
Hazelnuts	135g	861kcal	84g	20g
Golden raisins	165g	498kcal	0g	5g
Jam preserves	140g	383kcal	0g	0g

Recipe1M+				
Ingredient	Mass	Calories	Fat	Protein
Vegan margarine	2726.4g	19584kcal	2200.3g	3.84g
Brown sugar	3312.0g	12144kcal	0g	0g
Oats	234g	910.5kcal	16.14g	39.5g
Spelt flour	1500g	5460kcal	14.76g	154.9g
Hazelnuts	115.0g	722.0kcal	69.86g	17.19g
Golden raisins	165.0g	488.0kcal	0.89g	4.16g
Jam preserves	3840g	9216kcal	7.68g	26.88g

**Figure 8: Comparison of Nutrition Annotation between Uni-Food and Recipe1M+**

**Prompt:**

In this picture, The food is Kombu Tea Grilled Chicken Thigh The ingredients are 2 Chicken thighs, 2 tsp Kombu tea and 1 White pepper. Please offer numerical estimates for the quantity, calories, mass, fat, carbohydrate, and protein content of each ingredient without any explanatory reasoning. Use common sense for uncertain quantities to ensure the response remains solely numerical.

For example:

name amount mass calories fat carbs protein  
Couscous; 1/3 cup; 58g; 220kcal; 0.5g; 45g; 7g  
Orange extract; 1 teaspoon; 5g; 0kcal; 0g; 0g; 0g



Image input

**Response:**

name; amount; mass; calories; fat; carbs; protein  
Chicken thighs; 2 thighs; 520g; 660kcal; 42g; 0g; 60g  
Kombu tea; 2 tsp; 6g; 5kcal; 0g; 1g; 0g  
White pepper; 1 tsp; 3g; 8kcal; 0.1g; 2g; 0.3g

**Figure 3: Prompt and Response Example on Recipe1M.**

**References**

- [1] Lukas Bossard, Matthieu Guillaumin, and Luc Van Gool. 2014. Food-101—mining discriminative components with random forests. In *Computer vision—ECCV 2014: 13th European conference, zurich, Switzerland, September 6–12, 2014, proceedings, part VI 13*. Springer, 446–461.
- [2] Jingjing Chen, Bin Zhu, Chong-Wah Ngo, Tat-Seng Chua, and Yu-Gang Jiang. 2020. A study of multi-task and region-wise deep learning for food ingredient recognition. *IEEE Transactions on Image Processing* 30 (2020), 1514–1526.
- [3] Javier Marin, Aritro Biswas, Ferda Ofli, Nicholas Hynes, Amaia Salvador, Yusuf Aytar, Ingmar Weber, and Antonio Torralba. 2019. Recipe1M+: A Dataset for Learning Cross-Modal Embeddings for Cooking Recipes and Food Images. *IEEE Trans. Pattern Anal. Mach. Intell.* (2019).
- [4] Amaia Salvador, Nicholas Hynes, Yusuf Aytar, Javier Marin, Ferda Ofli, Ingmar Weber, and Antonio Torralba. 2017. Learning cross-modal embeddings for cooking recipes and food images. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3020–3028.

## GT



The ingredients include: bacon, cheese, potato, dressing, parsley.

The mass of this dish is 1107.80 grams, the energy content is 1609.00 kcal, the fat content is 77.90 grams, the protein content is 45.60 grams, and the carbohydrate content is 177.20 grams.

The cooking instruction are as follow: heat oven to 350f. mix first 3 ingredients in large bowl. add potatoes; toss to coat. spoon into 13x9-inch baking dish sprayed with cooking spray; cover. bake 55 min. or until potatoes are tender, uncovering after 40 min. sprinkle with parsley.



The ingredients include: vinegar, oil, flour, sugar, baking\_soda, salt, vanilla, milk, baking\_powder.

The mass of this dish is 944.20 grams, the energy content is 2753.00 kcal, the fat content is 113.10 grams, the protein content is 27.20 grams, and the carbohydrate content is 404.50 grams.

The cooking instruction are as follow: preheat oven to 350 degrees f (175 degrees c). grease and flour a 9x13 inch baking pan. in a large bowl, mix together the oil, sugar and 2 teaspoons vanilla until well blended. combine the flour, 1/2 cup cocoa, baking soda and salt; stir into the sugar mixture. fold in the zucchini and walnuts. spread evenly into the prepared pan. bake for 25 to 30 minutes in the preheated oven, until brownies spring back when gently touched. to make the frosting, melt together the 6 tablespoons of cocoa and margarine; set aside to cool. in a medium bowl, blend together the confectioners' sugar, milk and 1/2 teaspoon vanilla. stir in the cocoa mixture. spread over cooled brownies before cutting into squares.



The ingredients include: sesame, sauce, oil, beans.

The mass of this dish is 502.00 grams, the energy content is 390.50 kcal, the fat content is 26.10 grams, the protein content is 12.50 grams, and the carbohydrate content is 34.90 grams.

The cooking instruction are as follow: warm a large skillet or wok over medium heat. when the skillet is hot, pour in canola and sesame oils, then place whole green beans into the skillet. stir the beans to coat with oil. cook until the beans are bright green and slightly browned in spots, about 10 minutes. remove from heat, stir in soy sauce, cover, and let sit about 5 minutes. transfer to a serving platter, and sprinkle with toasted sesame seeds.

## UniFood Fintuned Model

The ingredients include: cheese, potato, bacon, pepper, onion.

The mass of this dish is 1017.00 grams, the energy content is 1110.00 kcal, the fat content is 20.00 grams, the protein content is 28.00 grams, and the carbohydrate content is 180.00 grams.

The cooking instruction are as follow: cook potatoes in boiling water until tender; drain. meanwhile, cook bacon in large skillet on medium-high heat until crisp. drain bacon on paper towels; crumble. add potatoes to bacon in skillet; toss to coat. add onions and cheese; toss to coat. cook and stir until cheese is melted and onions are tender.

The ingredients include: egg, clove, oil, flour, sugar, cinnamon, salt, nutmeg, vanilla, milk, baking\_powder.

The mass of this dish is 1117.00 grams, the energy content is 2410.00 kcal, the fat content is 62.00 grams, the protein content is 28.00 grams, and the carbohydrate content is 421.00 grams.

The cooking instruction are as follow: preheat oven to 350 degrees fahrenheit. grease a 9x9-inch baking pan. in a medium bowl, mix together brownie mix, water, oil, and eggs. spread batter into prepared pan. bake for 25 minutes. remove from oven and let cool. in a medium bowl, mix together cream cheese, sugar, and vanilla. spread over cooled brownies. refrigerate until set. cut into squares and serve.

The ingredients include: garlic, oil, sugar, ginger, beans, pepper, onion, sauce.

The mass of this dish is 108.00 grams, the energy content is 100.00 kcal, the fat content is 0.00 grams, the protein content is 2.00 grams, and the carbohydrate content is 24.00 grams.

The cooking instruction are as follow: wash and trim green beans. cut beans into 2-inch pieces. heat oil in a pan over medium heat. add beans and stir-fry until tender. season with salt and pepper to taste. serve hot.

Figure 9: Quality Results. The image illustrates both successful and failed cases across tasks including ingredient recognition, nutrition estimation, and recipe generation.