

References

A Human Evaluations

A.1 Evaluations on Dish Generation

Evaluating image generation models has always been a challenging issue. Although various benchmarks have been designed to assess general-purpose image generation models, these benchmarks are often not suitable for evaluating food images. Therefore, we conduct manual scoring across six dimensions: Identification Accuracy (ID Acc.), Texture, Composition, Scene, Lighting, and the Rules and Completeness of the Subject. Experts assign one of three scores (1/2/3) to each dimension for every piece of data. When sending for human evaluation, considering that annotators may not be very familiar with extremely niche dishes, we will provide two real-life photos as references alongside the evaluation images. The evaluation criteria for each dimension are as follows.

- **Identification accuracy (ID Acc.)** Evaluate if the dish image matches the dish name. ID Acc. should not only evaluate whether the generated content matches the textual description of the dish (e.g., the "King-Sized Handgun Chicken Leg" should be a chicken leg rather than an actual handgun), but also assess whether the ingredients and cooking methods in the dish are faithful and accurate (e.g., "Fo Tiao Qiang" should include sea cucumber and abalone).
- **Texture** Evaluate whether the texture of the image elements exhibits clarity and fineness.
- **Composition** Evaluate whether the image has good composition, with the dish positioned in the center of the frame and a well-structured layout. The main dish should occupy more than two thirds of the image.
- **Scene** Evaluate whether the scene hierarchy is harmonious, including prominent subject emphasis and stylistic consistency between background and subject. If the overall image appears obviously fake or heavily edited (e.g., with a strong "Photoshopped" look), it will receive a low score.
- **Lighting** Evaluate whether the lighting is aesthetically pleasing, refined, soft, and natural. The ambient lighting in the image should not exhibit color casts, while the dish itself may have a warm color tone.
- **Subject** Evaluate whether the main contour of the dish complies with real-world physical principles: conforming to objectively existing forms, free from fractures, distortions, garbled text, etc.

A.2 Evaluations on Dish Editing

We conduct human evaluations of dish editing across three dimensions: effectiveness (Effect.), consistency (Consist.), and aesthetics (Aes.). Similarly to the evaluations for dish generation, we employ a 3-point scoring system. In particular, we consider consistency and aesthetics meaningful only when editing is effective. For example, if the task is to add steam to a dish but the editing model merely

replicates the source image without producing any steam effect, assigning a high score (e.g., 3) for consistency would be unreasonable, and evaluating aesthetics would also be meaningless. Therefore, only when effectiveness is not scored as 1 (lowest) do we assess the other two dimensions. Otherwise, the scores for consistency and aesthetics are marked as invalid.

- **Effectiveness (Effect.)** Evaluate whether the image has been edited according to the instructions.
- **Consistency (Consist.)** Evaluate whether the unedited areas of the image remain consistent before and after editing.
- **Aesthetics (Aes.)** Evaluate the aesthetic quality of the image.

B More Visual Results

We provide additional visualization results in Fig. 1, Fig. 2, Fig. 3 and Fig. 4, where Omni-Dish can generate highly aesthetic images with rich culinary details. Furthermore, in Fig. 5, we present a set of results labeled with "low-refinement" tags. Although these images are less aesthetically pleasing, they appear more realistic and closer to everyday life. All this evidence demonstrates that Omni-Dish achieves photorealistic and faithful text-to-dish generation.

C Introduction to Chinese Dishes

Considering that some readers may not be familiar with the Chinese dishes mentioned in the main text, we provide an introduction to these dishes here. As shown in the next page, on the left side of the table are real images from the internet, and on the right side is a textual description of each dish.

Ground Truth Dish Image	Introduction
	驴打滚; Lü Da Gun. Lü Da Gun is a traditional Beijing snack made with glutinous rice flour, sweet red bean paste, roasted soybean flour, and sugar. The steamed rice dough is rolled out, spread with red bean paste, then rolled up and coated with fragrant soybean flour. It has a chewy, sweet texture with a rich bean aroma.
	宫保鸡丁; Kun Pao Chicken. Kung Pao Chicken is a classic Chinese dish made with diced chicken breast, peanuts, dried chili peppers, Sichuan peppercorns, scallions, ginger, and garlic. The sauce typically includes soy sauce, vinegar, sugar, and cooking wine, giving it a sweet, sour, and slightly spicy flavor with a distinctive numbing sensation from the peppercorns.
	夫妻肺片; Fuqi Feipian. Fuqi Feipian. is a classic Sichuan dish featuring sliced braised beef offal (tongue, heart, tripe) tossed in a spicy sauce made with chili oil, Sichuan peppercorn powder, soy sauce, garlic, cilantro, and crushed peanuts. Named after a Chengdu couple who popularized it, the dish is numbing, spicy, and savory with tender meat and a bold flavor.
	铁观音; Tieguanyin Tea. Tieguanyin Tea is a renowned oolong tea from Anxi, Fujian. The leaves are semi-ball-shaped with a sandy green color, brewing into a golden-yellow liquor with a natural orchid and fruity aroma. Rich in polyphenols, amino acids, and caffeine, it offers a mellow, sweet aftertaste and is believed to aid digestion and refresh the mind.
	佛跳墙; Fo Tiao Qiang. Fo Tiao Qiang. is a luxurious Fujianese dish known for its premium ingredients, such as sea cucumber, abalone, shark fin, dried scallops, shiitake mushrooms, tendon, quail eggs, ham, and a rich broth. The ingredients are layered and slow-cooked in sealed jars with Shaoxing wine for hours, resulting in an incredibly rich and flavorful soup.
	牛肉面; Beef Noodles. Beef Noodles is a classic Northwestern Chinese noodle dish. Key ingredients include high-gluten wheat noodles stretched by hand, braised beef, a savory broth simmered from beef bones, and toppings like daikon radish, cilantro, and garlic chives. The noodles are springy, the beef tender, and chili oil is often added for extra heat.
	猪肉炖粉条; Pork Stew with Vermicelli. Pork Stew with Vermicelli is a homestyle dish from Northeast China. Its main ingredients include pork belly, sweet potato vermicelli, napa cabbage, scallions, ginger, and garlic. The pork belly is first stir-fried to render its fat, then braised with seasonings until tender. Finally, the vermicelli and cabbage are added to soak up the rich broth.






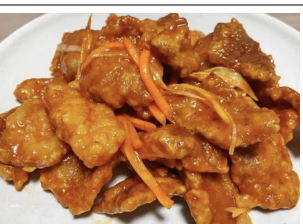

Ground Truth Dish Image	Introduction
	梅菜扣肉; Braised Pork with Preserved Mustard Greens. Braised Pork with Preserved Mustard Greens is a classic Hakka dish made with pork belly, preserved mustard greens, soy sauce, sugar, and cooking wine. The pork belly is first parboiled, then deep-fried until golden, sliced, and layered with soaked preserved mustard greens.
	九转大肠; Braised Pork Intestines "Nine-Turns Style" Braised Pork Intestines "Nine-Turns Style" is a traditional Shandong dish made with pork intestines. The intestines are thoroughly cleaned, blanched, deep-fried, and then braised with soy sauce, sugar and vinegar. The dish is known for its glossy red color and tender yet chewy texture. Its name "Nine-Turns" refers to the intricate cooking process.
	蚂蚁上树; Ants Climbing Trees. Ants Climbing Trees is a classic Sichuan dish made with vermicelli and minced meat. The dish gets its name from the minced meat clinging to the noodles, resembling ants on a tree. It is stir-fried with spicy bean paste, garlic, ginger, scallions, and soy sauce, offering a numbing-spicy, savory flavor with slippery, flavorful noodles and umami-rich meat—a perfect pairing with rice.
	京酱肉丝; Jingjiang Pork Shreds Jingjiang Pork Shreds is a popular street food named for its resemblance to a handgun. It uses a whole chicken leg, marinated, coated in batter or breadcrumbs, and deep-fried until crispy outside while remaining juicy inside. Seasonings often include chili powder, cumin, and five-spice, and it may be served with sweet chili or garlic sauce.
	霸王手枪腿; King-Sized Handgun Chicken Leg. King-Sized Handgun Chicken Leg is a popular street food named for its resemblance to a handgun. It uses a whole chicken leg, marinated, coated in batter or breadcrumbs, and deep-fried until crispy outside while remaining juicy inside. Seasonings often include chili powder, cumin, and five-spice.
	锅包肉; Pot-Fried Pork Pot-Fried Pork is a classic Northeastern Chinese dish, famous for its crispy exterior and sweet-and-sour flavor. The main ingredient is pork tenderloin, which is sliced, marinated, coated in batter, deep-fried, and then topped with a sauce made from sugar, vinegar, ketchup, and other seasonings. The dish features a crunchy outside with tender meat inside and a vibrant red appearance.
	玫瑰鲜花饼; Rose Shortcake Rose Shortcake is a traditional Yunnan delicacy made primarily with edible rose petals. It features a flaky, crispy crust and a fragrant filling of fresh rose petals blended with honey and sugar, offering a sweet yet delicate floral taste. The petals are handpicked at dawn to preserve their natural aroma, reflecting meticulous craftsmanship.



Figure 1: More visual results of Omni-Dish.



Figure 2: More visual results of Omni-Dish.



Figure 3: More visual results of Omni-Dish.



Figure 4: More visual results of Omni-Dish.



Figure 5: With low-refinement tags, we can generate low-refinement but more realistic and down-to-earth dish images.