

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58

59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116

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.

117	Ground Truth Dish Image	Introduction	175
118		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.	176
119		驴打滚; Lü Da Gun.	177
120			178
121			179
122			180
123			181
124			182
125			183
126		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.	184
127		宫保鸡丁; Kun Pao Chicken.	185
128			186
129			187
130			188
131			189
132			190
133			191
134		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.	192
135		夫妻肺片; Fuqi Feipian.	193
136			194
137			195
138			196
139			197
140			198
141			199
142		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.	200
143		铁观音; Tieguanyin Tea.	201
144			202
145			203
146			204
147			205
148			206
149			207
150		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.	208
151		佛跳墙; Fo Tiao Qiang.	209
152			210
153			211
154			212
155			213
156			214
157			215
158		牛肉面; Beef Noodles.	216
159		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.	217
160			218
161			219
162			220
163			221
164			222
165			223
166		猪肉炖粉条; Pork Stew with Vermicelli.	224
167		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.	225
168			226
169			227
170			228
171			229
172			230
173			231

233



249

250

251

252

253

254

255

256

257

258

259

260

261

262

263

264

265

266

267

268

269

270

271

272

273

274

275

276

277

278

279

280

281

282

283

284

285

286

287

288

289

290

291

292

293

294

295

296

297

298

299

300

301

302

303

304

305

306

307

308

309

310

311

312

313

314

315

316

317

318

319

320

321

322

323

324



325

326

327

328

329

330

331

332

333

334

335

336

337

338

339

340

341

342

343

344

345

346

347

348



Figure 1: More visual results of Omni-Dish.

349

350

351

352

353

354

355

356

357

358

359

360

361

362

363

364

365

366

367

368

369

370

371

372

373

374

375

376

377

378

379

380

381

382

383

384

385

386

387

388

389

390

391

392

393

394

395

396

397

398

399

400

401

402

403

404

405

406

407

408

409

410

411

412

413

414

415

416

417

418

419

420

421

422

423

424

425

426

427

428

429

430

431

432

433

434

435

436

437

438

439

440

441

442

443

444

445

446

447

448

449

450

451

452

453

454

455

456

457

458

459

460

461

462

463

464



Figure 2: More visual results of Omni-Dish.

465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481



Figure 3: More visual results of Omni-Dish.

517
518
519
520
521
522

523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580

581

582

583

584

585

586

587

588

589

590

591

592

593

594

595

596

597

598

599

600

601

602

603

604

605

606

607

608

609

610

611

612

613

614

615

616

617

618

619

620

621

622

623

624

625

626

627

628

629

630

631

632

633

634

635

636

637

638

639

640

641

642

643

644

645

646

647

648

649

650

651

652

653

654

655

656

657

658

659

660

661

662

663

664

665

666

667

668

669

670

671

672

673

674

675

676

677

678

679

680

681

682

683

684

685

686

687

688

689

690

691

692

693

694

695



Figure 4: More visual results of Omni-Dish.

697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807749
750
751
752
753
754

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

808
809
810
811
812