

Appendix

A. Cases of Omni IIE Bench

This section presents several GT images and their corresponding instructions for single-turn dialogues, including examples of edits at both low and high semantic scales as shown in Figures 7 and 8.

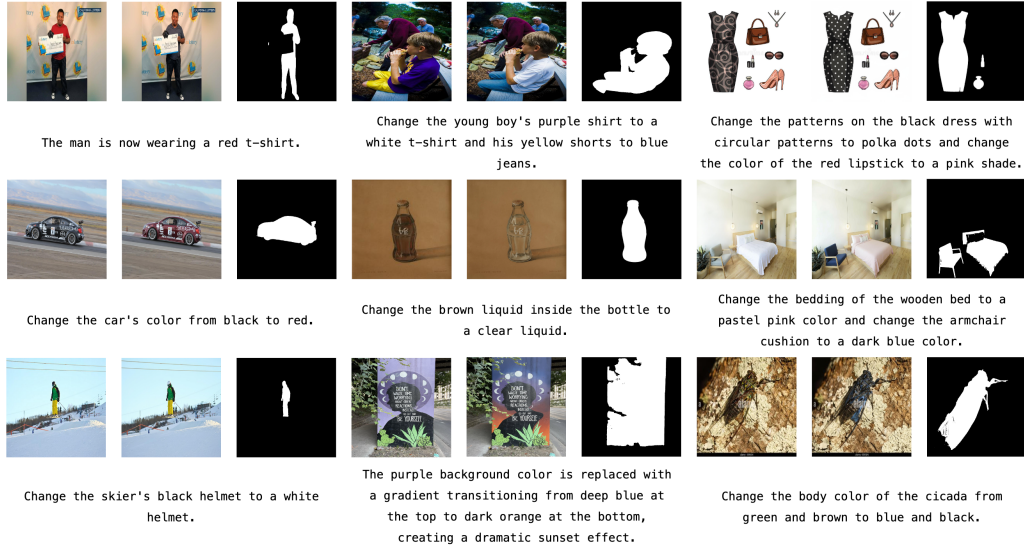


Figure 7. Cases of Single Turn Low



Figure 8. Cases of Single Turn High

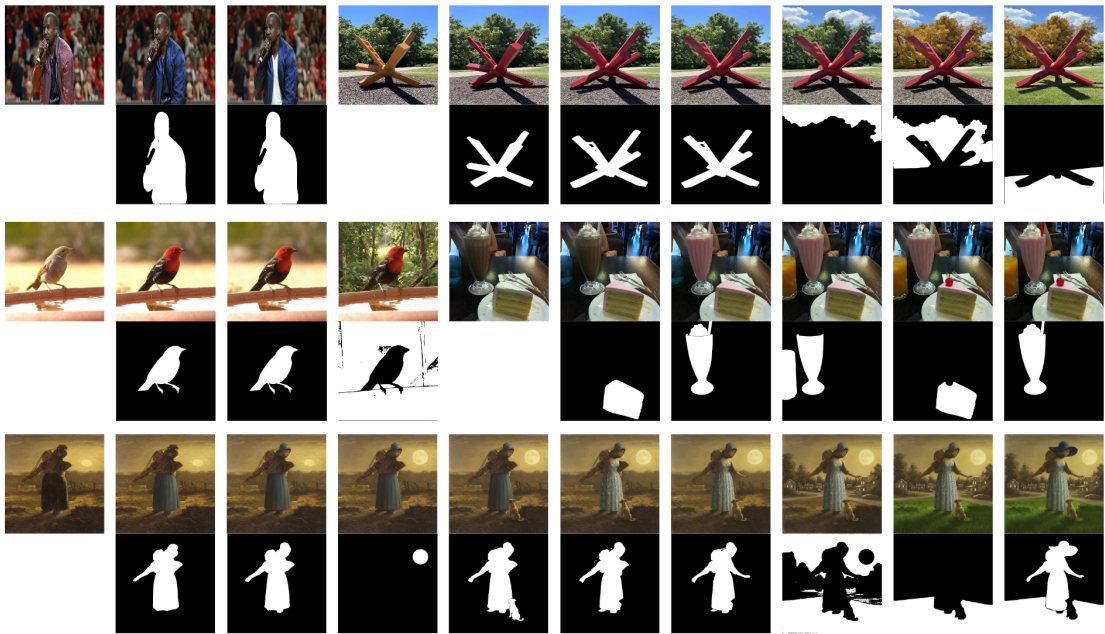


Figure 9. Cases of Multi Turn

B. Single-Turn Dialogue Metric Details

This section presents the metric details of different models at low and high semantic scales during single-turn dialogues. Overall, regardless of the model, edits at a high semantic scale perform worse than those at a low semantic scale. The specific details are shown in Tables 4 and 5.

Table 4. Performance comparison of IIE models on the Low group (metrics grouped by category). Text colors indicate top performances: 1st, 2nd, 3rd. Arrows (\uparrow / \downarrow) indicate whether higher or lower values are better. The Overall score is calculated by the formula $\frac{1}{4} \left[\frac{3 - (\sum \text{LPIPS})}{3} + \frac{\sum \text{CLIP}}{3} + \text{QA} + \text{SSIM} \right]$, where $\sum \text{LPIPS}$ and $\sum \text{CLIP}$ are the sums of their respective FG, BG, and ALL columns.

| Model | LPIPS (\downarrow) | | | CLIP (\uparrow) | | | QA (\uparrow) | Image Quality (\uparrow) | | Overall |
|------------------|------------------------|-------|-------|---------------------|-------|-------|-------------------|------------------------------|-------|---------|
| | FG | BG | ALL | FG | BG | ALL | | PSNR | SSIM | |
| Qwen-image-edit | 0.231 | 0.299 | 0.414 | 0.910 | 0.899 | 0.910 | 0.780 | 15.323 | 0.474 | 0.710 |
| Instruct_Pix2Pix | 0.284 | 0.435 | 0.555 | 0.861 | 0.797 | 0.796 | 0.400 | 12.476 | 0.443 | 0.559 |
| ICEdit | 0.250 | 0.311 | 0.407 | 0.888 | 0.896 | 0.889 | 0.550 | 16.107 | 0.519 | 0.662 |
| MGIE | 0.264 | 0.294 | 0.406 | 0.894 | 0.898 | 0.895 | 0.080 | 15.328 | 0.495 | 0.539 |
| HIVE | 0.283 | 0.376 | 0.505 | 0.873 | 0.836 | 0.836 | 0.300 | 14.257 | 0.444 | 0.551 |
| FLUX | 0.287 | 0.422 | 0.545 | 0.878 | 0.891 | 0.888 | 0.720 | 12.769 | 0.376 | 0.641 |
| HQEdit | 0.314 | 0.539 | 0.671 | 0.810 | 0.711 | 0.713 | 0.420 | 9.504 | 0.314 | 0.491 |
| Step1X | 0.218 | 0.244 | 0.353 | 0.912 | 0.923 | 0.920 | 0.650 | 16.663 | 0.547 | 0.711 |

Table 5. Performance comparison of IIE models on the High group (metrics grouped by category). Text colors indicate top performances: 1st, 2nd, 3rd. Arrows (\uparrow / \downarrow) indicate whether higher or lower values are better. Arrows (\uparrow / \downarrow) indicate whether higher or lower values are better. The Overall score is calculated by the formula $\frac{1}{4} \left[\frac{3 - (\sum \text{LPIPS})}{3} + \frac{\sum \text{CLIP}}{3} + \text{QA} + \text{SSIM} \right]$, where $\sum \text{LPIPS}$ and $\sum \text{CLIP}$ are the sums of their respective FG, BG, and ALL columns.

| Model | LPIPS (\downarrow) | | | CLIP (\uparrow) | | | QA (\uparrow) | Image Quality (\uparrow) | | Overall |
|------------------|------------------------|-------|-------|---------------------|-------|-------|-------------------|------------------------------|-------|---------|
| | FG | BG | ALL | FG | BG | ALL | | PSNR | SSIM | |
| Qwen-image-edit | 0.264 | 0.360 | 0.495 | 0.859 | 0.866 | 0.866 | 0.710 | 13.241 | 0.431 | 0.657 |
| Instruct_Pix2Pix | 0.283 | 0.448 | 0.586 | 0.816 | 0.749 | 0.747 | 0.230 | 12.356 | 0.431 | 0.498 |
| ICEdit | 0.254 | 0.311 | 0.448 | 0.840 | 0.849 | 0.842 | 0.360 | 14.830 | 0.492 | 0.589 |
| MGIE | 0.264 | 0.313 | 0.452 | 0.813 | 0.821 | 0.813 | 0.060 | 13.983 | 0.461 | 0.499 |
| HIVE | 0.278 | 0.415 | 0.552 | 0.819 | 0.783 | 0.778 | 0.220 | 13.122 | 0.407 | 0.501 |
| FLUX | 0.277 | 0.436 | 0.562 | 0.844 | 0.848 | 0.844 | 0.550 | 12.282 | 0.373 | 0.586 |
| HQEdit | 0.316 | 0.578 | 0.711 | 0.774 | 0.668 | 0.670 | 0.220 | 8.952 | 0.291 | 0.420 |
| Step1X | 0.247 | 0.278 | 0.411 | 0.857 | 0.878 | 0.872 | 0.510 | 14.818 | 0.515 | 0.646 |

C. Examples of Image Descriptions and Editing Instructions

This section presents several image descriptions and editing instructions generated during the construction of Omni IIE Bench.

```
{
  "replaceable_subject": {
    "name": "a hooded figure in a brown robe"
  },
  "modifiable_attributes": {
    "subject": [
      "robe is brown",
      "figure is hooded",
      "hands are clasped",
      "posture is upright"
    ],
    "background": [
      "background is dark and shadowy",
      "lighting is dramatic and focused on the figure"
    ]
  }
}
```

Listing 1. Example description JSON in our benchmark.

```
{
  "replaceable_subject": {
    "name": "a brown bird with a long tail"
  },
  "modifiable_attributes": {
    "subject": [
      "feathers are brown",
      "tail is long and dark",
      "beak is curved and pointed",
      "legs are slender and gray"
    ],
    "background": [
      "floor is tiled and beige",
      "pillar is gray and cylindrical",
      "partial view of a black object (possibly a chair or table)"
    ]
  }
}
```

Listing 2. Example description JSON in our benchmark.

```
{
  "replaceable_subject": {
    "name": "a modern bathroom interior"
  },
  "modifiable_attributes": {
    "subject": [
      "vanity cabinet is dark wood",
      "sink is white ceramic",
      "toilet is white ceramic",
      "shower curtain is beige with a floral pattern",
      "tub is white ceramic",
      "mirror is framed in gold"
    ],
    "background": [
      "walls are painted beige",
      "floor is light wood",
      "lighting is warm and ambient"
    ]
  }
}
```

Listing 3. Example description JSON in our benchmark.

```
{
  "modifications": [
    {
      "level": "low",
      "modification_text": "Change the color of the robe from brown to black."
    },
    {
      "level": "high",
      "modification_text": "Replace the hooded figure in a brown robe with a
↔ cloaked wizard wearing a pointy hat."
    }
  ]
}
```

Listing 4. Example single-turn modification JSON in our benchmark.

```
{
  "modifications": [
    {
      "level": "low",
      "modification_text": "Change the feathers of the brown bird to be white."
    },
    {
      "level": "high",
      "modification_text": "Replace the brown bird with a gray rabbit."
    }
  ]
}
```

Listing 5. Example single-turn modification JSON in our benchmark.

```

{
  "modifications": [
    {
      "level": "1",
      "modification_text": "Change the color of the vanity cabinet to a lighter
↪ oak tone."
    },
    {
      "level": "2",
      "modification_text": "Replace the floral pattern of the shower curtain
↪ with a geometric pattern."
    },
    {
      "level": "3",
      "modification_text": "Change the lighting to a cooler, brighter tone."
    },
    {
      "level": "4",
      "modification_text": "Replace the gold frame of the mirror with a silver
↪ frame."
    },
    {
      "level": "5",
      "modification_text": "Change the color of the walls to a light gray."
    },
    {
      "level": "6",
      "modification_text": "Replace the white ceramic sink with a black granite
↪ sink."
    },
    {
      "level": "7",
      "modification_text": "Change the floor to a darker wood tone."
    },
    {
      "level": "8",
      "modification_text": "Replace the white ceramic toilet with a black
↪ ceramic toilet."
    },
    {
      "level": "9",
      "modification_text": "Change the white ceramic tub to a black granite
↪ tub."
    },
    {
      "level": "10",
      "modification_text": "Replace the light gray walls with a dark gray
↪ color."
    },
    {
      "level": "11",
      "modification_text": "Replace the vanity cabinet with a marble countertop
↪ and a black frame."
    },
    {
      "level": "12",
      "modification_text": "Replace the shower curtain with a clear glass shower
↪ door."
    }
  ]
}

```

Listing 6. Example multi-turn modification JSON in our benchmark.