

# 1 ADDITIONAL QUANTITATIVE RESULTS

| Methods                                      | GenEval $\uparrow$ |          |              |         | DPG $\uparrow$ |          |           |         | T2I-CompBench $\uparrow$ |        |         |
|--|--------------------|----------|--------------|---------|----------------|----------|-----------|---------|--------------------------|--------|---------|
|  | Two Obj.           | Counting | Color Attri. | Overall | Entity         | Relation | Attribute | Overall | Color                    | Shape  | Texture |
| <b>Diffusion Models</b>                      |                    |          |              |         |                |          |           |         |                          |        |         |
| SDv1.5                                       | -                  | -        | -            | 0.40    | 74.23          | 73.49    | 75.39     | 63.18   | 0.3730                   | 0.3646 | 0.4219  |
| Lumina-Next                                  | 0.49               | 0.38     | 0.15         | 0.46    | 83.78          | 89.78    | 82.67     | 75.66   | 0.5088                   | 0.3386 | 0.4239  |
| SDv2.1                                       | 0.51               | 0.44     | 0.50         | 0.47    | -              | -        | -         | 68.09   | 0.5694                   | 0.4495 | 0.4982  |
| SDXL   | 0.74               | 0.39     | 0.23         | 0.55    | 82.43          | 86.76    | 80.91     | 74.65   | 0.6369                   | 0.5408 | 0.5637  |
| SD3-medium                                   | 0.74               | 0.63     | 0.36         | 0.62    | 91.01          | 80.70    | 88.83     | 84.08   | -                        | -      | -       |
| DALL-E3                                      | 0.87               | 0.47     | 0.45         | 0.67    | 89.61          | 90.58    | 88.39     | 83.50   | 0.8110                   | 0.6750 | 0.8070  |
| FLUX-dev                                     | 0.81               | 0.79     | 0.47         | 0.67    | 89.5           | 91.1     | 88.7      | 84.0    | -                        | -      | -       |
| FLUX-schnell                                 | 0.92               | 0.73     | 0.54         | 0.71    | 91.3           | 86.5     | 89.7      | 84.8    | -                        | -      | -       |
| <b>AutoRegressive-Diffusion Mixed Models</b> |                    |          |              |         |                |          |           |         |                          |        |         |
| Transfusion $\dagger$                        | -                  | -        | -            | 0.63    | -              | -        | -         | -       | -                        | -      | -       |
| Show-o $\dagger$                             | 0.52               | 0.49     | 0.28         | 0.53    | -              | -        | -         | 67.48   | -                        | -      | -       |
| <b>AutoRegressive Models</b>                 |                    |          |              |         |                |          |           |         |                          |        |         |
| LlamaGen                                     | 0.34               | 0.21     | 0.04         | 0.32    | -              | -        | -         | 65.16   | -                        | -      | -       |
| Chameleon                                    | -                  | -        | -            | 0.39    | -              | -        | -         | -       | -                        | -      | -       |
| Janus $\dagger$                              | 0.68               | 0.30     | 0.42         | 0.61    | 87.38          | 85.46    | 87.70     | 79.68   | -                        | -      | -       |
| Lumina-mGPT (Ours)                           | 0.77               | 0.27     | 0.32         | 0.56    | 86.60          | 91.29    | 84.61     | 79.68   | 0.6371                   | 0.4727 | 0.6034  |

Table 1: **Quantitative comparison on text-to-image benchmarks.**  $\dagger$  indicates concurrent or later work.

| Method             | Non-Spatial | 2D-Spatial |
|--------------------|-------------|------------|
| Lumina-mGPT (Ours) | 0.3176      | 0.2394     |

Table 2: Performance on T2I-Compbench across spatial and non-spatial dimensions. Metrics are reported independently as these were not provided by prior works.

## 2 REVISED QUALITATIVE COMPARISON (FIG. 3)

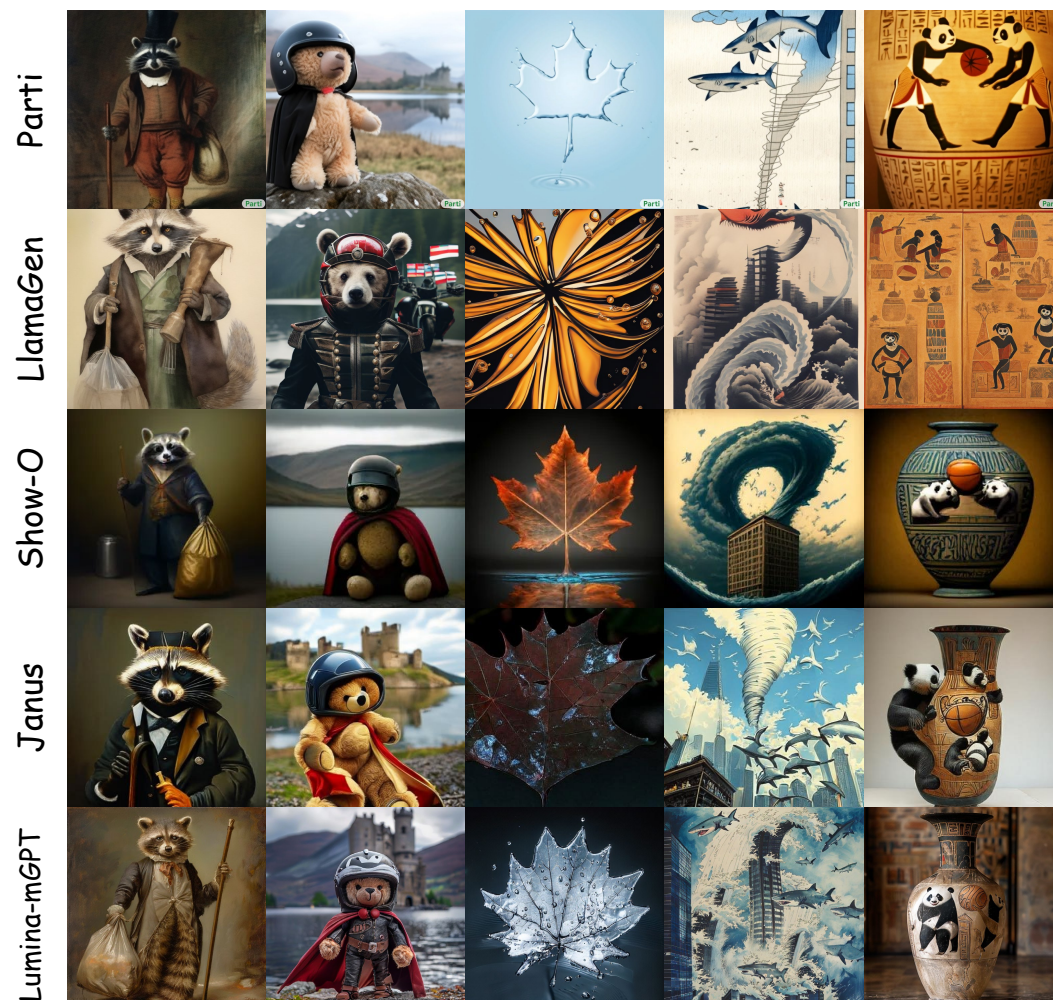


Figure 1: Qualitative Comparison