

ImagenHub: Standardizing the evaluation of conditional image generation models

♠Max Ku, ♠Tianle Li, †Kai Zhang, ♣Yujie Lu, ♥Xingyu Fu, ◇Wenwen Zhuang, ♠Wenhu Chen

♠University of Waterloo, †Ohio State University, ♣University of California Santa Barbara,
♥University of Pennsylvania, ◇Central South University

tiger-ai-lab.github.io/ImagenHub

ImagenHub

ImagenHub Data

Text2Image hyperrealism fruits and vegetables market

Control-guided



.....

ImageEditing



ImagenHub Inference Library

```
import imagenhub  
  
model = imagenhub.load(name)  
img = model.generate(condition)
```

API



Inference



ImagenHub Evaluator



Raters



Semantic Consistency



Perceptual Quality

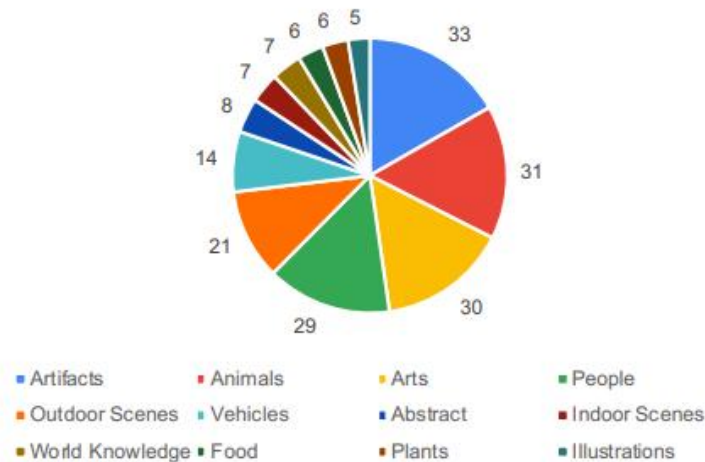
Motivation

Research community has to identify the models reliability

- Showing true progress of the field
- Standardizing Human evaluation
- To compare models on a fair platform

Inference Dataset

Task	Inference Dataset size
Text-guided Image Generation	197
Mask-guided Image Editing	179
Text-guided Image Editing	179
Subject-driven Image Generation	150
Multi-concept Image Composition	102
Subject-driven Image Editing	154
Control-guided Image Generation	150



(a) Text-to-Image Generation dataset

Human Evaluation

Semantic Consistency (SC) in range [0, 0.5, 1]

- how well the generated image is aligned with the condition X

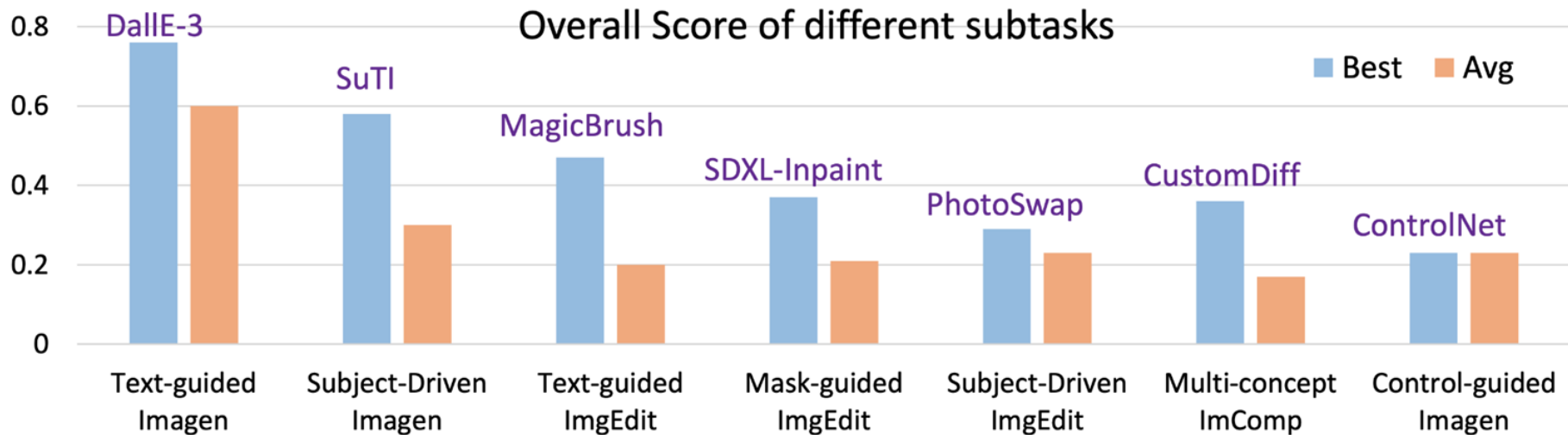
Perceptual Quality (PQ) in range [0, 0.5, 1]

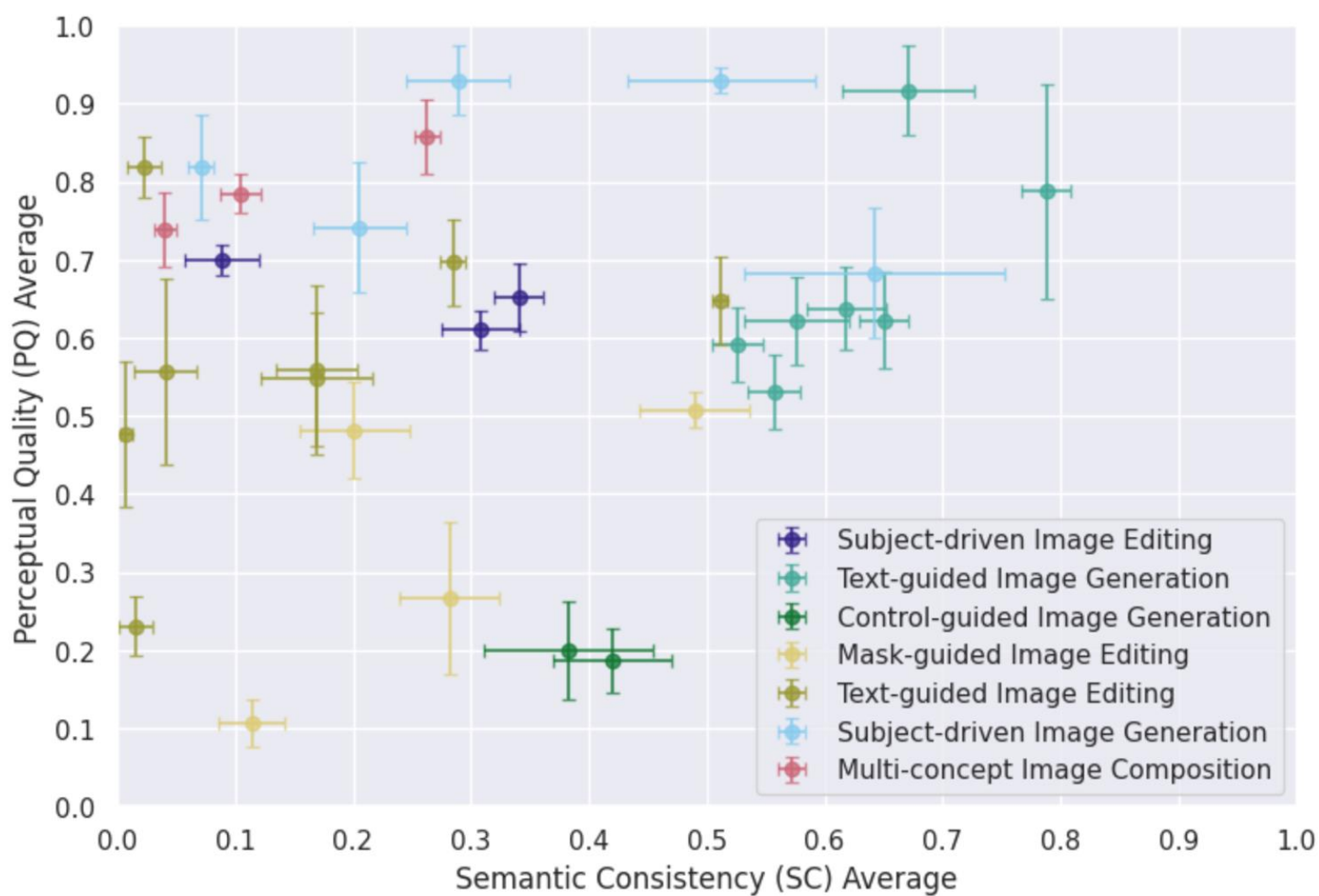
- how well is the image quality of the generated image

Overall (O)

- Geometric mean of two aspects $O = \sqrt{SC \times PQ}$

Results

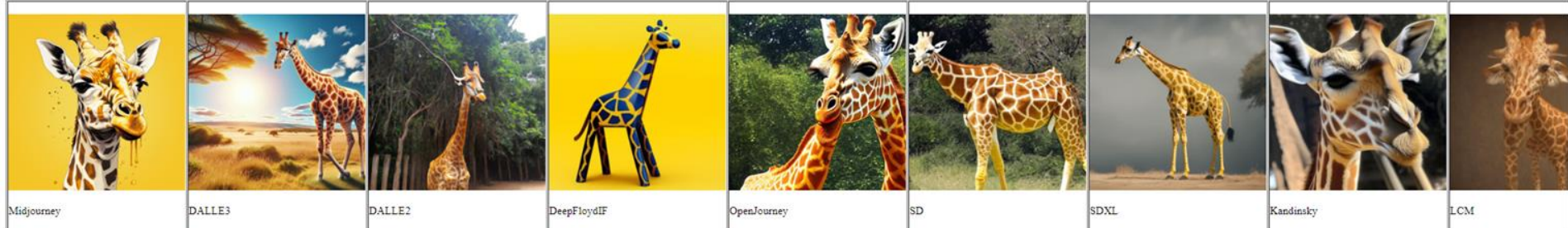




Task	PQ \uparrow				SC \uparrow
	corr(LPIPS)	corr(DINO)	corr(CLIP-I)	corr(DreamSim)	corr(CLIP)
Text-guided Image Generation	N/A	N/A	N/A	N/A	-0.0819
Mask-guided Image Editing	-0.0502	-0.0421	-0.0308	-0.0294	-0.0257
Text-guided Image Editing	0.0439	0.0449	0.0765	0.0432	0.0437
Subject-driven Image Generation	0.0905	0.1737	0.1204	0.0943	-0.0117
Subject-driven Image Editing	0.2417	0.1690	0.3770	0.3846	-0.0443
Multi-concept Image Composition	-0.1018	-0.1657	-0.0965	0.0140	-0.0107
Control-guided Image Generation	0.1864	0.1971	-0.0985	-0.0045	-0.2056

Imagen Museum

```
sample_6.jpg  
{  
  "prompt": "A yellow colored giraffe.",  
  "category": "Colors",  
  "source": "DrawBench"  
}
```



```
sample_7.jpg  
{  
  "prompt": "Rainbow coloured penguin.",  
  "category": "Conflicting",  
  "source": "DrawBench"  
}
```



```
sample_8.jpg  
{
```

Leaderboard

Model	SC_{Avg}	PQ_{Avg}	Overall	Fleiss \bar{O}	Kd \bar{O}
Text-guided Image Generation					
Dalle-3	0.79±0.02	0.79±0.14	0.76±0.08	0.19	0.34
Midjourney	0.67±0.06	0.92±0.06	0.73±0.07	0.34	0.51
DeepFloydIF	0.65±0.02	0.62±0.06	0.59±0.02	0.32	0.51
Stable Diffusion XL	0.62±0.03	0.64±0.05	0.59±0.03	0.37	0.61
Dalle-2	0.58±0.04	0.62±0.06	0.54±0.04	0.27	0.40
OpenJourney	0.53±0.02	0.59±0.05	0.50±0.02	0.30	0.47
Stable Diffusion 2.1	0.56±0.02	0.53±0.05	0.50±0.03	0.38	0.50
Mask-guided Image Editing					
SDXL-Inpainting	0.49±0.05	0.51±0.02	0.37±0.05	0.50	0.72
SD-Inpainting	0.28±0.04	0.27±0.10	0.17±0.07	0.31	0.49
GLIDE	0.20±0.05	0.48±0.06	0.16±0.05	0.33	0.56
BlendedDiffusion	0.12±0.03	0.11±0.03	0.05±0.02	0.36	0.44
Text-guided Image Editing					
MagicBrush	0.51±0.01	0.65±0.06	0.47±0.02	0.44	0.67
InstructPix2Pix	0.29±0.01	0.70±0.06	0.27±0.02	0.55	0.74
Prompt-to-prompt	0.17±0.05	0.55±0.09	0.15±0.06	0.36	0.53
CycleDiffusion	0.17±0.03	0.56±0.11	0.14±0.04	0.41	0.63
SDEdit	0.04±0.03	0.56±0.12	0.04±0.03	0.13	0.13
Text2Live	0.02±0.01	0.82±0.04	0.02±0.02	0.10	0.17
DiffEdit	0.02±0.01	0.23±0.04	0.01±0.01	0.24	0.24
Pix2PixZero	0.01±0.00	0.48±0.09	0.01±0.01	0.37	0.37

Subject-driven Image Generation					
SuTI	0.64±0.11	0.68±0.08	0.58±0.12	0.20	0.39
DreamBooth	0.51±0.08	0.93±0.02	0.55±0.11	0.37	0.60
BLIP-Diffusion	0.29±0.04	0.93±0.04	0.35±0.06	0.22	0.39
TextualInversion	0.21±0.04	0.74±0.08	0.21±0.05	0.35	0.52
DreamBooth-Lora	0.07±0.01	0.82±0.07	0.09±0.01	0.29	0.37
Subject-driven Image Editing					
PhotoSwap	0.34±0.02	0.65±0.04	0.36±0.02	0.35	0.46
DreamEdit	0.31±0.03	0.61±0.03	0.32±0.03	0.33	0.52
BLIP-Diffusion	0.09±0.03	0.70±0.02	0.09±0.03	0.41	0.47
Multi-concept Image Composition					
CustomDiffusion	0.26±0.01	0.86±0.05	0.29±0.01	0.73	0.88
DreamBooth	0.11±0.02	0.78±0.02	0.13±0.02	0.61	0.71
TextualInversion	0.04±0.01	0.74±0.05	0.05±0.01	0.62	0.77
Control-guided Image Generation					
ControlNet	0.42±0.05	0.19±0.04	0.23±0.04	0.37	0.57
UniControl	0.38±0.07	0.20±0.06	0.23±0.07	0.36	0.58