



# Bringing Textual Prompt to AI-Generated Image Quality Assessment

Bowen Qu, Haohui Li, Wei Gao\*

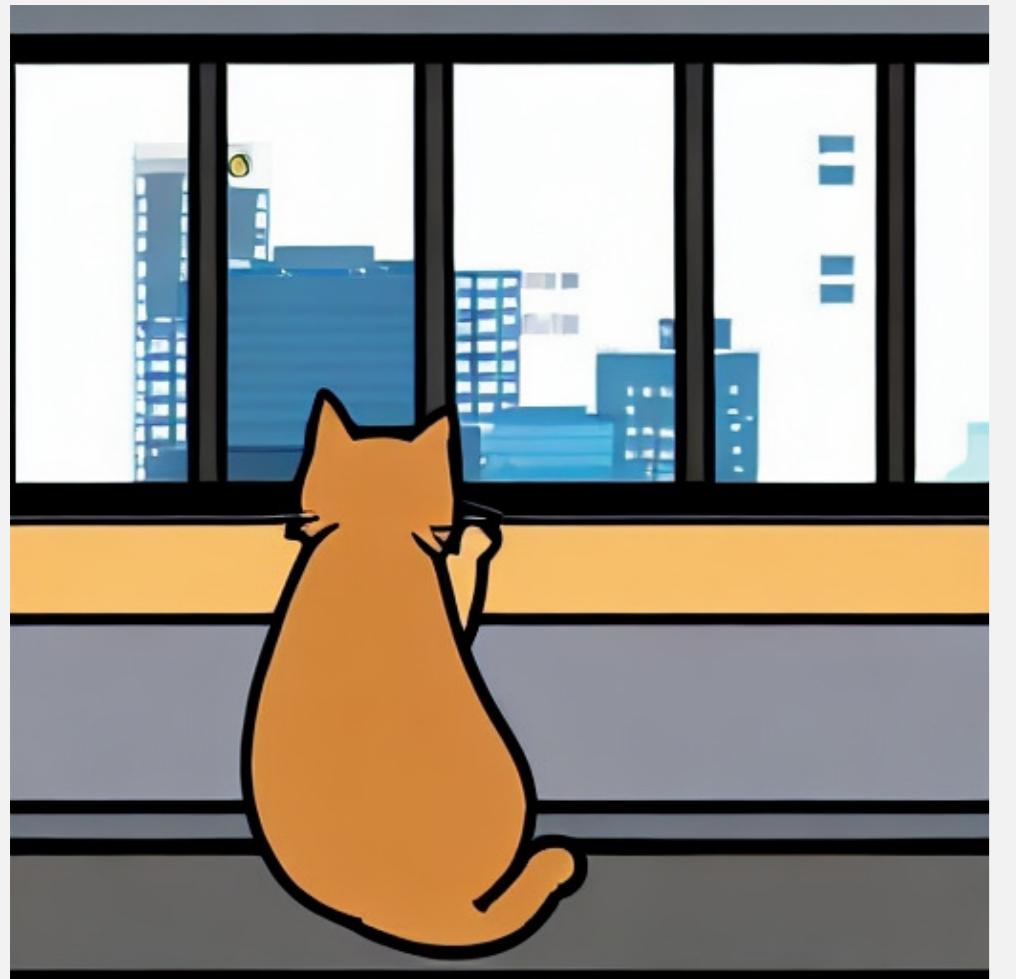
Peking University

\* Indicates corresponding author



Code will be available at:  
<https://github.com/Coobiw/IP-IQA>

## Inherent Difference: AGI-QA vs Traditional IQA

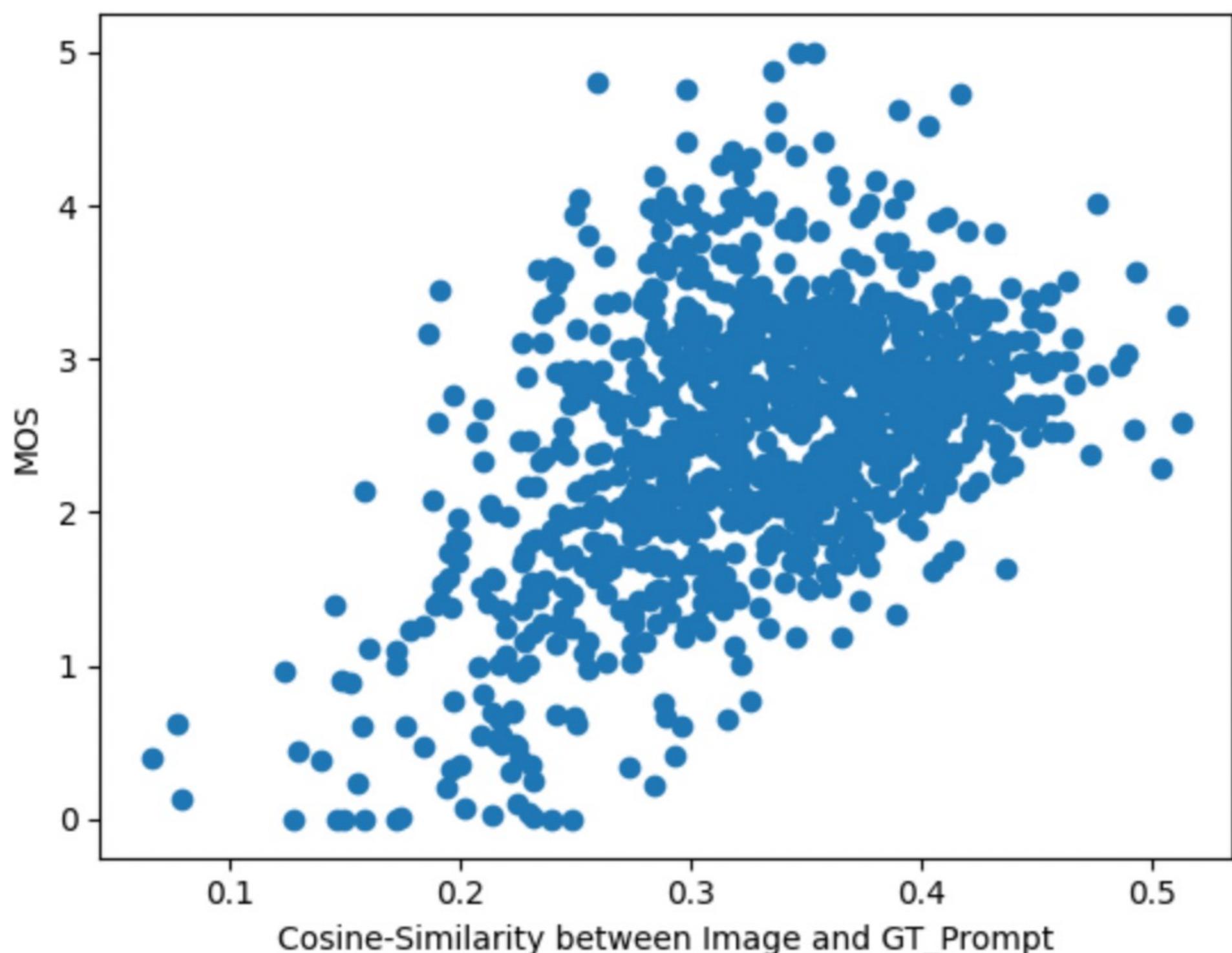


Prompt:  
cat, using laptop, city, anime style  
Quality: 😊  
Alignment: 😫  
Groundtruth quality: 2.0465 😫  
Prediction by ResNet50: 3.2716 😐

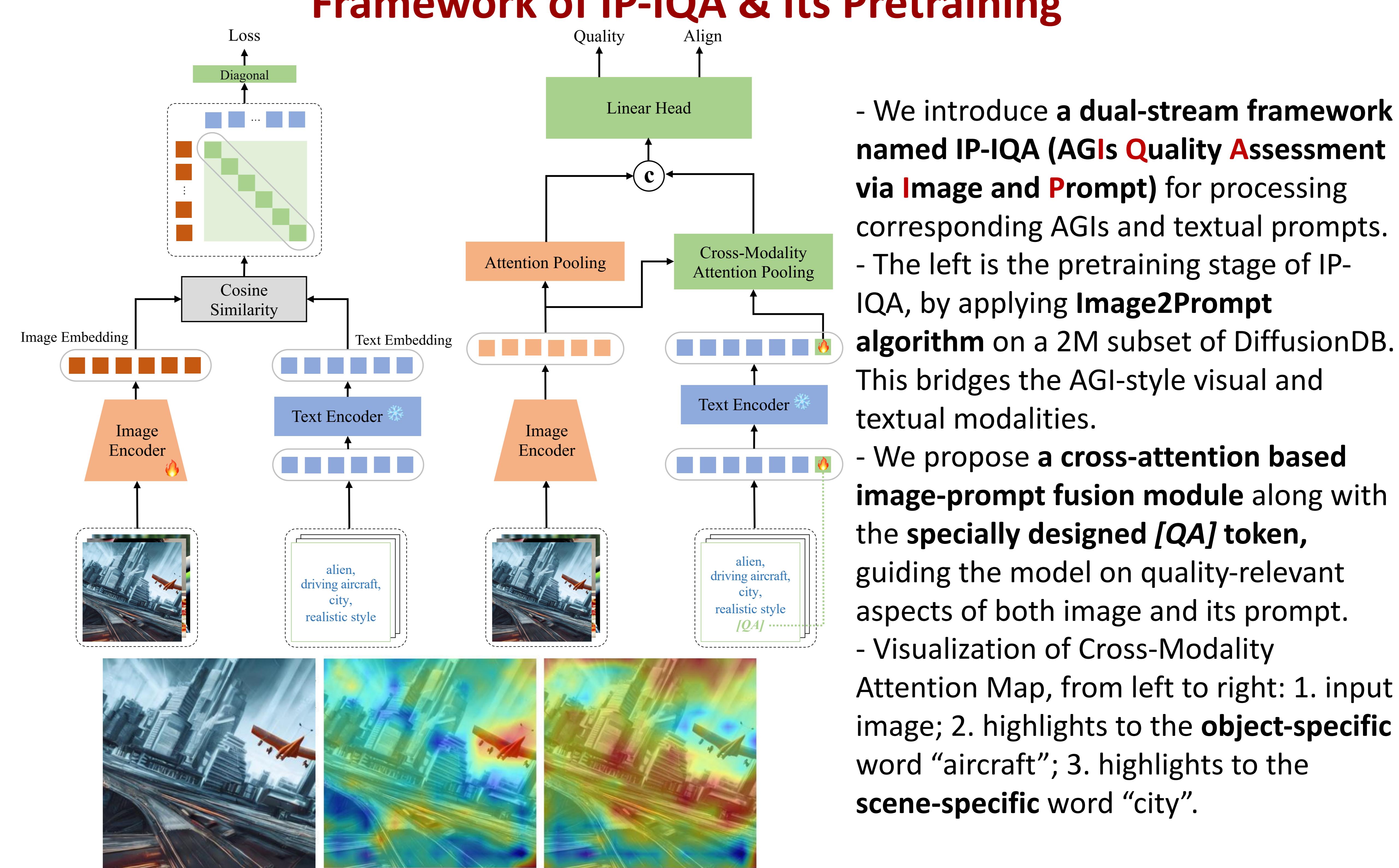


Prompt:  
snoppy, using laptop, city, realistic style  
Quality: 😊  
Alignment: 😫  
Groundtruth quality: 2.2140 😫  
Prediction by ResNet50: 3.4628 😐

- The ground truth score includes **considerations of image quality and the image-prompt correspondence**.
- Although these images exhibit high visual quality, **the IQA method fall short in terms of image-text correspondence**.
- This is because **AGIs have inherent multimodal natures from birth**. We need a multi-modal framework to get a comprehensive assessment.



- We used open-clip to calculate the cosine similarity between images and prompts.
- The results **indicate a positive correlation** between the MOS scores of images and their cosine similarity with prompts.



## Results on AGIQA-1k & AGIQA-3k

Table 1. The performance results of perceptual models.

Methods	AGIQA-1k			AGIQA-3k		
	SRCC	PLCC	KRCC	SRCC	PLCC	KRCC
Handcrafted-based	CEIQ	0.3069	0.2836	0.2097	0.3228	0.4166
	NIQE	-0.5490	-0.5048	-0.3824	0.5623	0.5171
	DSIQA	-0.3047	-0.0559	-0.2148	0.4955	0.5488
	SISBLIM	-0.1309	-0.3575	-0.0889	0.5479	0.6477
SVR-based	GMLF	0.5575	0.6356	0.4052	0.6987	0.8181
	HIGRADE	0.4056	0.4425	0.2860	0.6171	0.7056
DL-based	DBCNN	0.7491	0.8211	0.5618	0.8207	0.8759
	CNNIQA	0.5800	0.7139	0.4095	0.7478	0.8469
	CLIPQA	0.8227	0.8411	0.6399	0.8426	0.8053
	HyperNet	0.7803	0.8299	0.5943	0.8355	0.8903
	ResNet50	0.7136	0.7576	0.5254	0.6744	0.7365
	ResNet50*	0.7914	0.8404	0.6064	0.8306	0.8929
	Ours	<b>0.8401</b>	<b>0.8922</b>	<b>0.6635</b>	<b>0.8634</b>	<b>0.9116</b>

Table 2. The performance results of alignment models

Method	Alignment		
	SRCC	PLCC	KRCC
CLIPScore	0.5972	0.6839	0.4591
ImageReward	0.7298	0.7862	0.5390
HPS	0.6349	0.7000	0.4580
StairReward	0.7472	0.8529	0.5554
Ours	<b>0.7578</b>	<b>0.8544</b>	<b>0.5734</b>

Table 3. The ablation results on perceptual quality.

Image2prompt	Integral prompt	/Q4/ token	AGIQA-1k			AGIQA-3k		
			SRCC	PLCC	KRCC	SRCC	PLCC	KRCC
			0.8105	0.8595	0.6173	0.8398	0.8978	0.6561
✓			0.8180	0.8703	0.6501	0.8491	0.9031	0.6690
	✓		0.8317	0.8706	0.6532	0.8442	0.9008	0.6612
✓	✓		0.8383	0.8782	0.6603	0.8595	0.9084	0.6798
✓	✓	✓	<b>0.8401</b>	<b>0.8922</b>	<b>0.6635</b>	<b>0.8634</b>	<b>0.9116</b>	<b>0.6844</b>