Appendix

A. Evaluation Criteria of Subjective Experiment

In our subjective experiment, evaluators are asked to express their preferences for the displayed AIGIs from three evaluation perspectives: quality, authenticity, and text-image correspondence. Evaluators rate AIGIs on a 0 to 5 scale with 0.01 increments. The detailed evaluation criteria for these evaluation perspectives is shown in Figure 1.

How does the quality of this image appear?

- Very poor (0~1 point)
- Poor (1~2 points)
- Average (2~3 points)
- Good (3~4 points)
- Excellent (4~5 points)

(a) Quality evaluation

Does this image look like it was generated by AI or is it a real image?

- It is AI-generated (0~1 point)
- It appears more likely to be AI-generated (1~2 points)
- It could be AI-generated or real (2~3 points)
- It looks more like a real image (3~4 points)
- It is a real image (4~5 points)

(b) Authenticity evaluation

Does the generated image match the text prompt?

- Completely inconsistent (0~1 point)
- There are significant differences from the text prompt (1~2 points)
- There are minor differences from the text prompt (2~3 points)
- There are almost no differences from the text prompt (3~4 points)
- Completely consistent (4~5 points)

(c) Correspondence evaluation

Figure 1. Illustration of evaluation criteria for three evaluation perspectives: quality, authenticity, and text-image correspondence.

B. Regression Methods Based on Text-Image Encoder

Unlike the evaluation of general image content, text prompts are also crucial for assessing AIGIs, especially in terms of text-image correspondence scores. Therefore, we further consider using the TIER[7] method and employing Bert-base[1] to extract information from text prompts. We will briefly explain how the TIER framework is incorporated into the proposed IQA methods as follows.

In the NR-AIGCIQA, FR-AIGCIQA, and PR-AIGCIQA methods, they first extract image features from images I_g using a image feature extraction module $F^I_{w_i}$, followed by a regression network R_θ that produces a predicted score. This process can be represented as:

$$\hat{s} = R_\theta(F^I_{w_i}(I_g)) \tag{1}$$

To incorporate the TIER framework into these methods, we just need to add a text encoder branch $F^T_{w_t}$ to extract text features from the text prompt T_p, then merge these text features with the image features using concatenation, and finally use a regression network to obtain the predicted score. The new method can be represented as:

$$\hat{s} = R_\theta(\text{Concat}(F^T_{w_t}(T_p), F^I_{w_i}(I_g))) \tag{2}$$

To clarify this operation, we provide a schematic diagram of how the TIER framework is incorporated into these IQA methods, as shown in Figure 2.

Method	Quality		Authenticity		Correspondence	
	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC
VGG16(NR)	0.7411	0.7439	0.6910	0.6919	0.7277	0.7222
VGG19(NR)	0.7401	0.7396	0.6808	0.6905	0.7279	0.7256
ResNet18(NR)	0.7480	0.7394	0.7072	0.7110	0.7573	0.7432
ResNet50(NR)	0.7424	0.7384	0.7149	0.7162	0.7329	0.7258
InceptionV4(NR)	0.7426	0.7509	0.7126	0.7157	0.7425	0.7355
ViT-L/16(NR)	0.8182	0.8216	0.7957	0.8007	**0.8197**	0.8456
VGG16(PR)	0.7531	0.7461	0.7024	0.7044	0.7449	0.7451
VGG19(PR)	0.7513	0.7520	0.6843	0.6902	0.7469	0.7495
ResNet18(PR)	0.7582	0.7609	0.7184	0.7246	0.7611	0.7828
ResNet50(PR)	0.7640	0.7667	0.7162	0.7169	0.7442	0.7573
InceptionV4(PR)	0.7637	0.7687	0.7285	0.7317	0.7614	0.7601
ViT-L/16(PR)	**0.8213**	**0.8265**	0.7917	0.7973	0.8170	0.8426
ResNet18(NR-TIER)	0.7369	0.7329	0.7082	0.7189	0.7514	0.7394
ResNet50(NR-TIER)	0.7587	0.7504	0.7205	0.7241	0.7385	0.7437
InceptionV4(NR-TIER)	0.7323	0.7358	0.6965	0.6955	0.7351	0.7292
ViT-L/16(NR-TIER)	0.8194	0.8220	**0.7970**	**0.8014**	0.8179	**0.8463**
ResNet18(PR-TIER)	0.7631	0.7666	0.7161	0.7229	0.7659	0.7653
ResNet50(PR-TIER)	0.7624	0.7614	0.7167	0.7206	0.7498	0.7478
InceptionV4(PR-TIER)	0.7621	0.7655	0.7163	0.7155	0.7547	0.7645
ViT-L/16(PR-TIER)	0.8205	0.8255	0.7934	0.7984	0.8144	0.8427

Table 1. Performance comparison of whether to use the TIER method on the PKU-AIGIQA-4K database.

Method	Quality		Authenticity		Correspondence	
	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC
VGG16(NR)	0.8199	0.8119	0.7529	0.7413	0.7798	0.7872
VGG19(NR)	0.8174	0.8084	0.7498	0.7423	0.7839	0.7865
ResNet18(NR)	0.8128	0.8018	0.7815	0.7710	0.7879	0.7934
ResNet50(NR)	0.8138	0.8086	0.7717	0.7618	0.7811	0.7837
InceptionV4(NR)	0.8099	0.8067	0.7643	0.7419	0.7725	0.7746
ViT-L/16(NR)	**0.8649**	**0.8678**	0.8322	**0.8283**	**0.8287**	**0.8586**
ResNet18(NR-TIER)	0.8136	0.8102	0.7817	0.7682	0.7940	0.8007
ResNet50(NR-TIER)	0.8197	0.8178	0.7792	0.7686	0.7819	0.7865
InceptionV4(NR-TIER)	0.8317	0.8314	0.7783	0.7653	0.7901	0.8077
ViT-L/16(NR-TIER)	0.8625	0.8621	**0.8344**	0.8272	0.8269	0.8284

Table 2. Performance comparison of whether to use the TIER method on the T2IQA subset of PKU-AIGIQA-4K database.

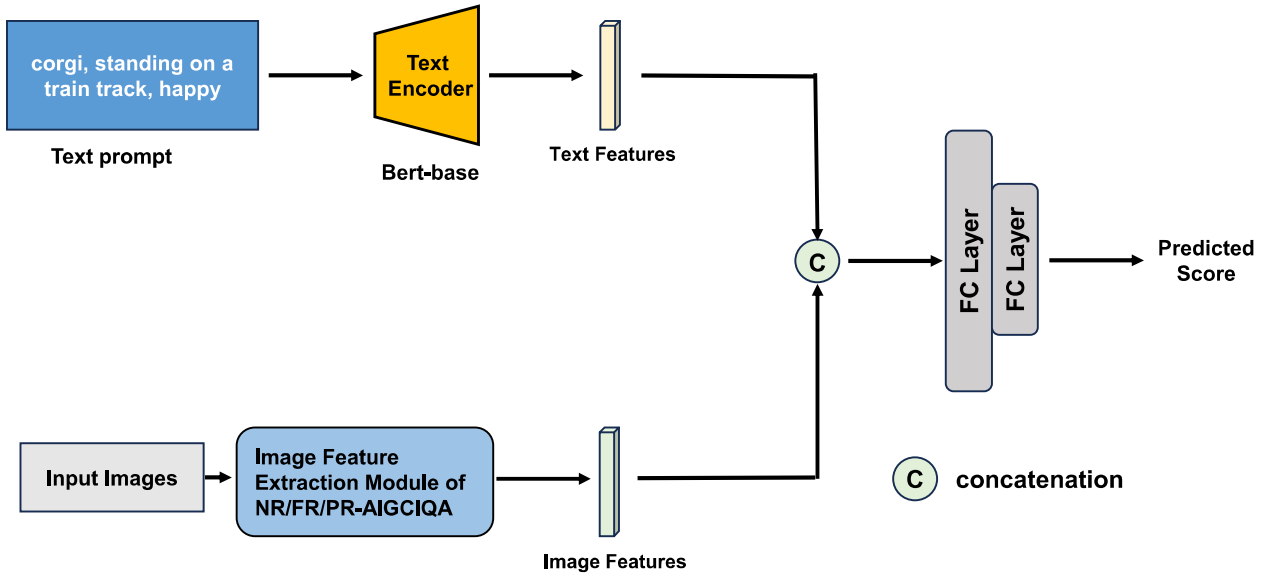

Figure 2. Illustration of the NR-TIER, FR-TIER, and PR-TIER methods.

Method	Quality		Authenticity		Correspondence	
	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC
VGG16(NR)	0.6734	0.6854	0.6449	0.6975	0.7130	0.7095
VGG19(NR)	0.6836	0.6855	0.6352	0.6845	0.7383	0.7348
ResNet18(NR)	0.6885	0.7112	0.6684	0.7108	0.7492	0.7317
ResNet50(NR)	0.6876	0.6875	0.6530	0.6918	0.7456	0.7385
InceptionV4(NR)	0.6988	0.7076	0.6733	0.7191	0.7509	0.7306
ViT-L/16(NR)	0.7505	0.7627	**0.7361**	**0.7740**	0.8147	0.8216
VGG16(FR)	0.6825	0.6918	0.6468	0.7005	0.7589	0.7740
VGG19(FR)	0.6832	0.7093	0.6505	0.7056	0.7594	0.7741
ResNet18(FR)	0.7063	0.7249	0.6724	0.7220	0.7737	0.7892
ResNet50(FR)	0.6885	0.6968	0.6567	0.6983	0.7662	0.7803
InceptionV4(FR)	0.7017	0.7246	0.6788	0.7298	0.7626	0.7627
ViT-L/16(FR)	0.7551	0.7657	0.7268	0.7689	**0.8294**	**0.8476**
ResNet18(NR-TIER)	0.6788	0.7104	0.6518	0.6968	0.7453	0.7399
ResNet50(NR-TIER)	0.6953	0.7003	0.6543	0.7055	0.7434	0.7360
InceptionV4(NR-TIER)	0.6862	0.7187	0.6274	0.6842	0.7440	0.7382
ViT-L/16(NR-TIER)	**0.7593**	**0.7667**	0.7321	0.7716	0.8135	0.8286
ResNet18(FR-TIER)	0.7082	0.7097	0.6681	0.7155	0.7825	0.7938
ResNet50(FR-TIER)	0.6869	0.7031	0.6710	0.7106	0.7752	0.7894
InceptionV4(FR-TIER)	0.6997	0.7245	0.6675	0.7081	0.7725	0.7840
ViT-L/16(FR-TIER)	0.7436	0.7524	0.7271	0.7635	0.8157	0.8372

Table 3. Performance comparison of whether to use the TIER method on the I2IQA subset of PKU-AIGIQA-4K database.

C. Additional Results

In this section, we will provide more experimental results which is not including in the main text.

Performance comparison of whether to use the TIER method. Here, we present a performance comparison between the method using text prompts and the one without text prompts. The experimental results are shown in Table 1, 2, and 3. The findings indicate that, in most cases, using text prompts improves the model's performance when evaluating AIGIs.

Method	AGIQA-1K		AGIQA-3K		AIGCIQA2023	
	Quality		Quality		Quality	
	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC
LinearityIQA[3]	0.8200	0.8578	0.8189	0.8309	0.7947	0.8057
HyperIQA[4]	0.8534	0.8912	0.8495	0.8923	0.8159	0.8212
MUSIQ[2]	0.8506	0.8850	0.8338	0.8698	0.8261	0.8382
MANIQA[6]	0.8804	**0.9084**	0.8916	0.9194	0.8412	0.8540
StairIQA[5]	0.8640	0.8899	0.8543	0.8943	0.8313	0.8376
ViT-B/16(NR)	0.8748	0.9063	0.8845	0.9146	0.8404	0.8532
NR-CLIPIQA	**0.8945**	0.9032	**0.9205**	**0.9368**	**0.8691**	**0.8803**

Table 4. Performance comparisons on several existing AIGIQA databases, including AGIQA-1K, AGIQA-3K, and AIGCIQA2023.

Additional Results on Several Existing AIGIQA Databases. Moreover, we conduct experiments on several existing AIGIQA databases, including AGIQA-1K, AGIQA-3K, and AIGCIQA2023. For dataset partitioning, we randomly split the dataset into training set and test set at a ratio of 4:1. The experimental results are presented in Table 4. The experimental results indicate that NR-CLIPIQA achieves the best performance across all datasets, further highlighting the superiority of the pre-trained CLIP model in evaluating AIGIs compared to other pre-trained models.

References

[1] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In *North American Chapter of the Association for Computational Linguistics*, 2019. 1

[2] Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and

Feng Yang. Musiq: Multi-scale image quality transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 5148–5157, 2021. 2

[3] Dingquan Li, Tingting Jiang, and Ming Jiang. Norm-in-norm loss with faster convergence and better performance for image quality assessment. In *Proceedings of the 28th ACM International conference on multimedia*, pages 789–797, 2020. 2

[4] Shaolin Su, Qingsen Yan, Yu Zhu, Cheng Zhang, Xin Ge, Jinqiu Sun, and Yanning Zhang. Blindly assess image quality in the wild guided by a self-adaptive hyper network. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3664–3673, 2020. 2

[5] Wei Sun, Huiyu Duan, Xiongkuo Min, Li Chen, and Guangtao Zhai. Blind quality assessment for in-the-wild images via hierarchical feature fusion strategy. In *2022 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, pages 01–06, 2022. 2

[6] Sidi Yang, Tianhe Wu, Shuwei Shi, Shanshan Lao, Yuan Gong, Mingdeng Cao, Jiahao Wang, and Yujiu Yang. Maniqa: Multi-dimension attention network for no-reference image quality assessment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1191–1200, 2022. 2

[7] Jiquan Yuan, Xinyan Cao, Jinming Che, Qinyuan Wang, Sen Liang, Wei Ren, Jinlong Lin, and Xixin Cao. Tier: Text and image encoder-based regression for aigc image quality assessment. *arXiv preprint arXiv:2401.03854*, 2024. 1