

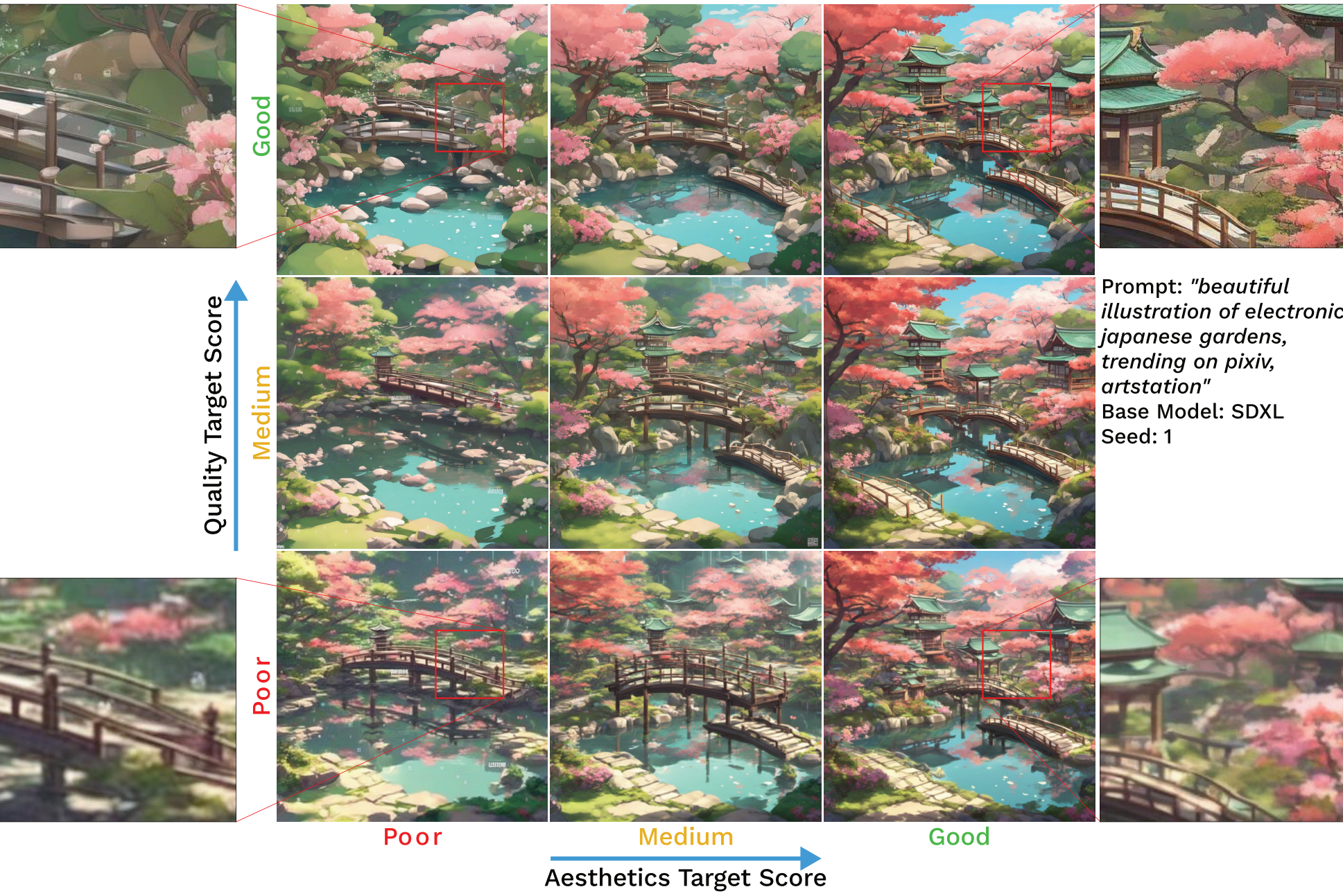
IQA-Adapter: Exploring Knowledge Transfer from Image Quality Assessment to Diffusion-based Generative Models

Khaled Abud^{1,2}, Sergey Lavrushkin¹, Alexey Kirillov^{2,3}, Dmitriy Vatin^{1,2}

¹ MSU Institute for Artificial Intelligence,
² Lomonosov Moscow State University,
³ Yandex

Motivation

Recent advances in diffusion-based models have greatly improved text-to-image generation, producing highly realistic visuals from textual prompts. Despite these improvements, modern generative models still lack the understanding of image quality concept. At the same time, learning-based Image Quality and Aesthetics Assessment (IQA/IAA) systems have seen rapid growth. Integration of IQA/IAA models directly into the generative architectures is a logical next step towards aligning generated images with human preferences, moving beyond traditional text-image alignment. Motivated by recent successful transfers of generative priors to IQA, we explore the opposite direction: *incorporating IQA expertise into generative diffusion models via conditioning*.



Quality-aware image generation with **IQA-Adapter**
X and Y axes represent different inputs to the adapter for aesthetics and technical quality

Contributions

Qualitative adaptation method: We introduce **IQA-Adapter**, a novel conditioning tool for diffusion models that enables quality-aware generation guided by IQA/IAA scores.

Diverse IQA/IAA model integration: We experiment with a wide range of IQA/IAA models, demonstrating the adaptability of IQA-Adapter and generalization of its quality features.

Reference-based conditioning: We adopt IQA-Adapter for image-prompting scenario, demonstrating that IQA models can be used to transfer highly specific, content-agnostic qualitative information without semantics of the reference image.

Quality-conditioning with reference image

Aside from conditioning on quality scores, we modify IQA-Adapter to use images as qualitative references. We extract activations from deep layers of IQA models and use them as adapter input. Compared to classical feature extractors, IQA embeddings contain highly concentrated qualitative information without semantics of the reference image.

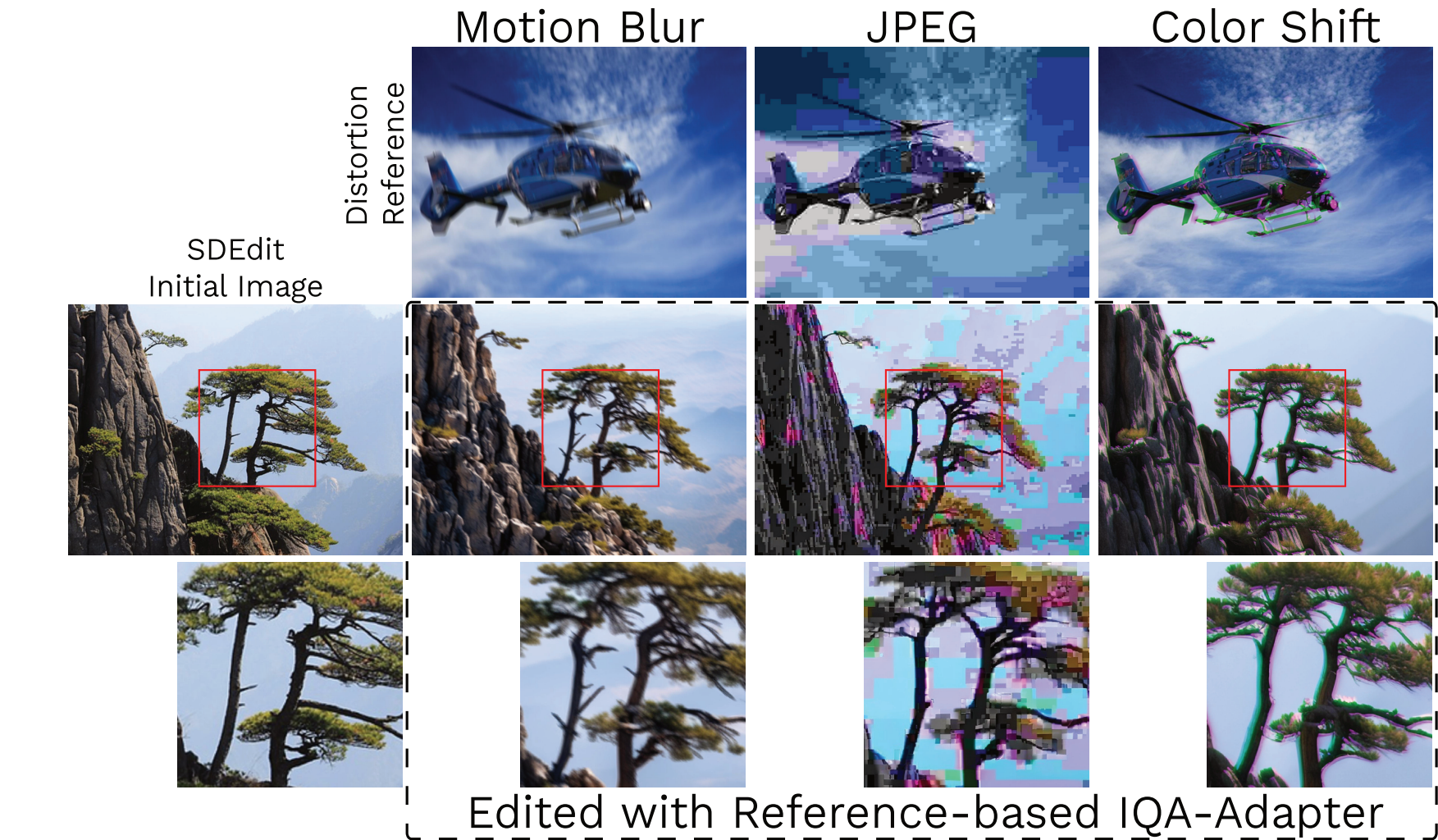


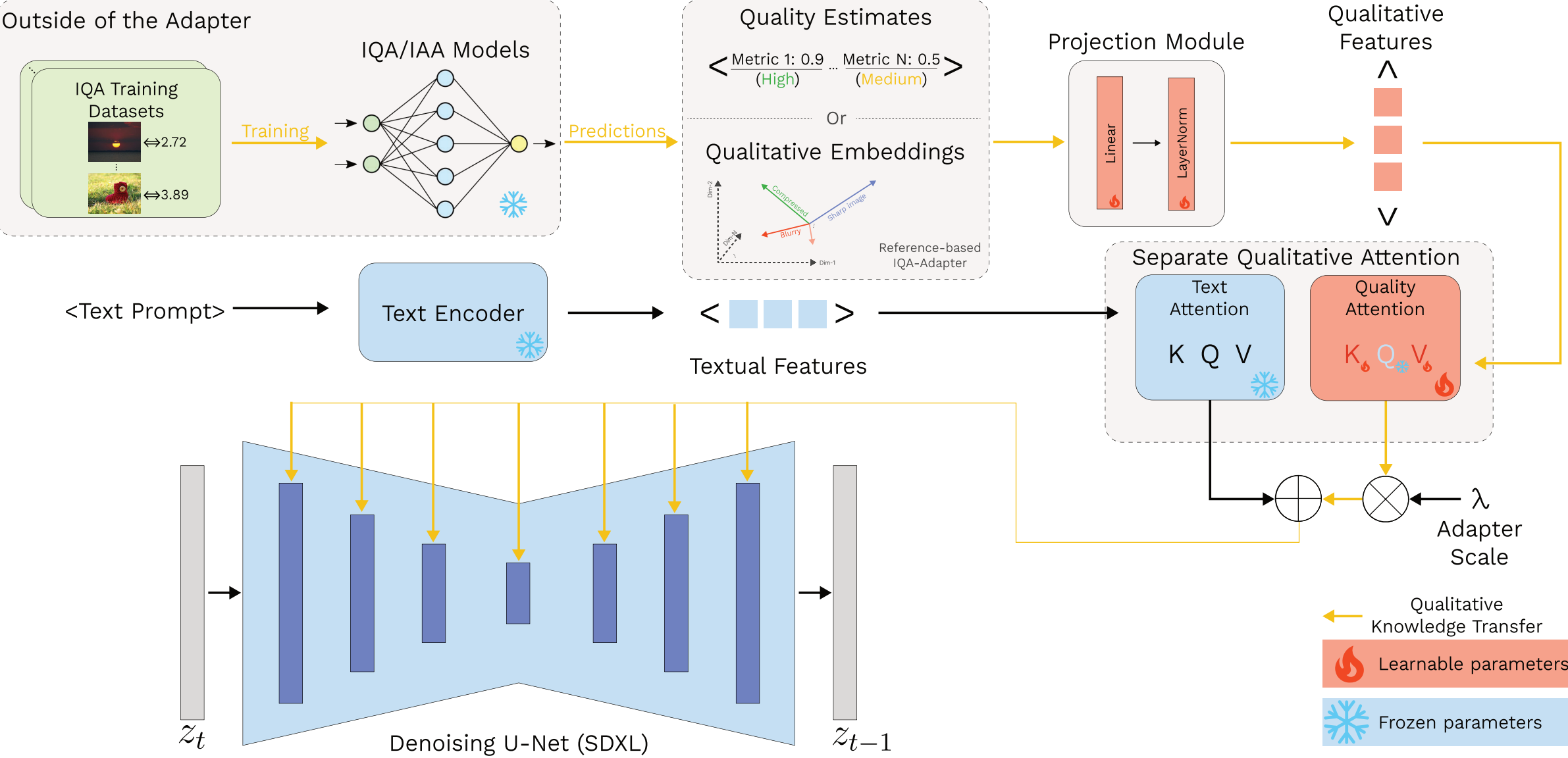
Image-to-Image generation with Reference-based IQA-Adapter conditioned on different distortions



Visit our project page:
<https://x1716.github.io/IQA-Adapter/>
Code is available
in our GitHub repository:
<https://github.com/X1716/IQA-Adapter>



Method overview



IQA-Adapter architecture

To condition the generative model on image quality, we leverage an adapter-based approach. The core idea is to project new data into additional tokens, which are then integrated into the model via cross-attention mechanisms, enabling the base model to receive detailed information from the new sources without altering its core weights. We separate cross-attention operators for textual and qualitative information to make features learned from IQA/IAA models content-agnostic:

$$\text{CrossAttn}(Z, c_t, c_q) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V + \lambda \times \text{Softmax}\left(\frac{QK'^T}{\sqrt{d}}\right)V'$$

Text Attention Quality Attention

Additionally, we implement qualitative negative guidance to isolate the concepts of “good” and “bad” quality and boost the effects of the adapter:

$$\hat{\epsilon}_{\theta}(z_t|c_t, q) = \epsilon_{\theta}(z_t|c_t^{\text{neg}}, q^{\text{neg}}) + g \cdot (\epsilon_{\theta}(z_t|c_t, q) - \epsilon_{\theta}(z_t|c_t^{\text{neg}}, q^{\text{neg}}))$$

q Target quality q^{neg} Opposite quality ϵ_{θ} Diffusion model c_t Text prompt

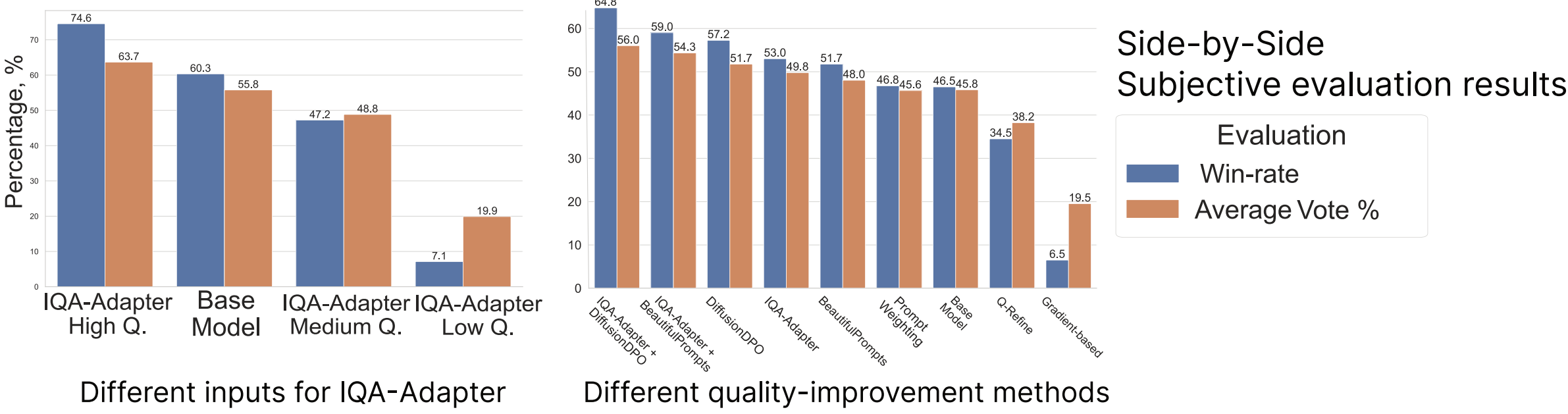
Results

Alignment with qualitative condition

When conditioned on high target quality, IQA-Adapter enhances generated images with improved clarity, fine-grained details and vibrant colors. Conversely, low-quality condition forces generative model to reproduce various artifacts, e.g. compression and blur. To test the conditioning capabilities of the IQA-Adapter, we conduct a side-by-side user study (1000+ users, 22k+ responses) comparing generations with different target quality.

High-quality conditioning

To objectively evaluate positive impact of the IQA-Adapter conditioned on high quality, we measure the gain of IQA/IAA scores relative to base model. We test IQA-Adapters trained with a wide range of IQA/IAA models: they demonstrate significant generalization across evaluation metrics. We also compare IQA-Adapters with other generation quality improvement methods and conduct a subjective study. IQA-Adapter can be easily combined with other techniques, and these combinations yield the best results.



IQA/IAA models used during IQA-Adapter training	IQA																				IAA					VLM-based					Average	Quality increase relative to base model, %													
	ARNIA (KADID)	TOPIQ (KONIQ)	MANIQ (KONIQ)	ARNIA (KONIQ)	MUSIQ (KONIQ)	HYPER-IQA	CNNQA	DBCNN	MANIQ (PPAL)	TOPIQ (SPAQ)	MUSIQ (FLIVE)	TOPIQ (FLIVE)	TRES (FLIVE)	ARNIA (FLIVE)	MUSIQ (AVA)	NIMA (AVA)	TOPIQ (AVA)	LAION-AES	HPSv2	CLIP-IQA+	LQIE MIX	LQIE																							
	TOPIQ, LAION-AES + Neg. G.																				5.1	16.3	32.8	14.8	6.0	20.4	3.0	15.9	12.1	24.2			4.2	5.6	5.8	8.9	4.3	4.4	4.1	0.3	-0.2	15.8	8.5	26.7	10.9
	TOPIQ, LAION-AES																				1.3	13.7	26.3	11.7	5.3	17.9	3.3	14.1	9.5	20.8			3.7	4.6	4.8	6.9	3.4	3.2	3.3	0.5	-0.1	13.0	7.2	22.8	9.0
	DBCNN + Neg. G.																				-4.6	11.9	27.8	9.3	5.2	17.4	3.2	12.7	9.0	21.5			3.7	4.1	4.1	5.1	4.4	3.2	3.2	-0.6	-0.3	13.2	6.0	20.9	8.2
	ARNIQA (3 versions), LQIE-MIX + Neg. G.																				11.2	11.6	23.3	9.0	4.9	13.9	2.8	10.4	6.7	18.0			4.0	4.2	4.0	6.1	2.8	2.6	1.8	-0.8	-0.2	11.8	8.1	22.4	8.2
	TOPIQ (KONIQ)																				-0.3	11.8	24.5	8.8	4.8	15.9	3.3	11.3	9.5	20.2			3.6	4.2	4.3	6.2	2.2	2.0	2.0	-0.3	-0.2	12.3	7.2	21.7	8.0
	TOPIQ (4 versions)																				-2.9	11.2	23.4	9.5	4.8	15.5	3.6	10.9	7.2	19.2			3.5	4.0	3.6	5.4	5.5	5.7	5.9	-0.4	-0.2	12.3	5.7	20.7	7.9
	DBCNN																				-3.9	10.7	23.8	8.1	4.9	15.1	3.1	11.4	7.4	18.5			3.4	3.6	3.7	4.3	3.7	2.6	2.5	-0.6	-0.2	11.5	5.7	19.2	7.2
	ARNIQA (3 versions), LQIE-MIX																				8.8	9.4	18.3	6.9	4.5	11.1	2.3	8.5	5.2	14.7			3.6	3.6	3.3	4.9	2.0	2.0	1.2	-0.7	-0.1	11.1	7.3	19.0	6.7
MUSIQ(KONIQ), MUSIQ(AVA)																				0.1	9.7	18.1	7.9	4.2	11.4	2.7	8.9	5.4	12.9	2.8	3.2	2.9	4.5	3.4	4.3	3.3	-0.9	-0.6	8.5	5.7	17.4	6.2			
TRES (FLIVE)																				-2.1	9.6	18.7	5.9	4.3	11.9	2.8	8.8	4.0	16.0	3.6	3.6	3.9	4.7	2.6	1.7	1.1	-0.6	0.0	8.1	5.1	17.5	6.0			
MANIQA (KONIQ)																				-1.9	7.1	18.2	5.0	3.8	10.2	2.1	7.3	7.5	15.6	3.1	3.2	3.2	4.8	0.8	2.4	1.5	-0.8	-1.1	11.6	6.1	15.3	5.7			
CNNIQA																				-1.8	7.2	16.8	6.6	3.8	10.4	2.6	7.6	6.0	14.8	3.0	3.3	3.1	5.1	-0.2	1.3	0.4	-0.7	-1.1	9.4	6.1	16.5	5.5			
MUSIQ (AVA)																				2.4	5.7	12.2	5.8	2.6	7.0	2.3	5.1	4.0	7.7	1.9	2.2	1.9	3.3	2.7	3.8	3.3	0.2	-0.3	6.4	5.2	12.5	4.4			
CLIP-IQA+																				3.0	4.6	11.9	4.3	2.9	5.4	1.0	3.7	3.1	10.2	2.7	2.9	2.4	3.8	-0.5	1.7	0.2	-1.4	-0.4	9.7	5.7	14.5	4.2			
ARNIQA (KADID)																				-6.6	3.0	8.8	4.2	2.6	3.5	0.9	3.6	2.1	7.2	2.1	2.3	1.9	2.9	-0.6	1.6	0.3	-1.3	-1.0	7.0	6.4	10.5	3.4			
DiffusionDPO+ IQA-Adapter (TOPIQ, LAION-AES) + Neg. G.																				6.2	13.1	40.9	16.5	5.6	20.2	3.0	12.6	10.0	25.2	4.3	5.5	5.1	9.5	6.9	6.3	5.1	-0.9	0.6	17.7	7.8	28.1	11.3			
BeautifulPrompt + IQA-Adapter(TOPIQ,LAION-AES)																				-6.3	15.0	32.1	13.5	6.2	20.3	2.9	15.6	10.8	25.6	4.3	5.0	5.0	7.8	5.3	5.6	5.1	1.3	-5.0	17.7	9.9	27.4	10.8			
DiffusionDPO+ IQA-Adapter (TOPIQ, LAION-AES)																				3.1	10.6	36.9	14.5	5.0	18.6	3.2	11.2	7.8	23.1	4.0	5.0	4.4	8.4	6.0	5.7	4.7	-0.7	0.8	15.7	6.6	25.5	10.0			
Prompt Weighting + IQA-Adapter(TOPIQ,LAION-AES)																				3.0	14.1	27.5	12.3	5.4	18.2	3.3	14.4	10.0	22.0	3.8	4.9	5.0	7.3	3.8	3.7	3.7	0.4	0.0	13.6	8.0	23.9	9.5			
Q-Refine + IQA-Adapter (TOPIQ, LAION-AES)																				20.8	-0.7	2.2	4.8	3.2	-1.2	-2.4	-2.0	2.8	2.7	2.7	2.6	2.7	5.0	-2.2	2.7	1.0	-0.5	-0.5	8.5	8.4	14.6	3.4			
DiffusionDPO																				2.3	-2.5	19.0	7.6	0.6	7.3	2.6	0.5	-0.7	10.2	2.1	2.2	1.6	4.5	3.7	4.8	4.0	-0.7	0.8	7.3	1.9	13.6	4.2			
BeautifulPrompt																				-2.6	3.3	8.0	1.9	1.5	3.2	0.4	3.9	0.6	7.1	1.3	0.7	0.2	0.7	2.6	3.0	2.0	0.8	-4.9	5.9	4.5	10.6	2.7			
Gradient-based, TRES(FLIVE)																				-0.2	4.1	8.6	0.1	2.3	5.8	1.6	2.8	5.2	9.7	1.6	1.5	5.0	1.2	-0.7	-6.3	-4.1	-0.7	-3.7	-0.9	5.1	5.2	2.0			
Prompt Weighting																				-2.8	0.3	1.2	0.7	0.4	0.2	0.3	0.4	0.6	1.9	0.4	0.4	0.3	0.7	0.4	0.8	0.5	0.0	0.2	0.8	2.1	2.1	0.8			
Q-Refine																				13.2	-5.7	-5.2	0.0	0.5	-6.3	-3.9	-6.5	0.7	-3.3	0.9	0.5	0.6	2.3	-2.4	1.4	0.5	-0.0	-0.2	3.6	5.7	3.8	0.0			
Evaluation Quality Metric																				ARNIA (KADID)	TOPIQ (KONIQ)	MANIQ (KONIQ)	ARNIA (KONIQ)	MUSIQ (KONIQ)	HYPER-IQA	CNNQA	DBCNN	MANIQ (PPAL)	TOPIQ (SPAQ)	MUSIQ (FLIVE)	TOPIQ (FLIVE)	TRES (FLIVE)	ARNIA (FLIVE)	MUSIQ (AVA)	NIMA (AVA)	TOPIQ (AVA)	LAION-AES	HPSv2	CLIP-IQA+	LQIE MIX	LQIE	Average			
IQA-Adapters trained with different IQA models										Combinations with IQA-Adapter										Other methods																									