

Poster Session at Graduate School Information Fair

The Overall Quality Prediction on AI-generated Images (AIGI)

General Background

Introduction

AIGI text-to-image models [1] are rapidly becoming a key to content creation in various domains, including entertainment, art, design, and advertising, and are also being generalized to image editing. Despite potential, recent studies have shown many of the ``uncomfortable'' and ``unnatural'' content. Such contents, known as image distortions in multimedia, present unique challenges to AIGI. We categorized these issues into three distinct levels: semantic level, perceptual level, and axiology level. Some studies [2] [3] have already shown that accurate quality prediction metrics can significantly mitigate these issues. However, the quality prediction accuracy for AIGI remains suboptimal [4].



Fig. 1.

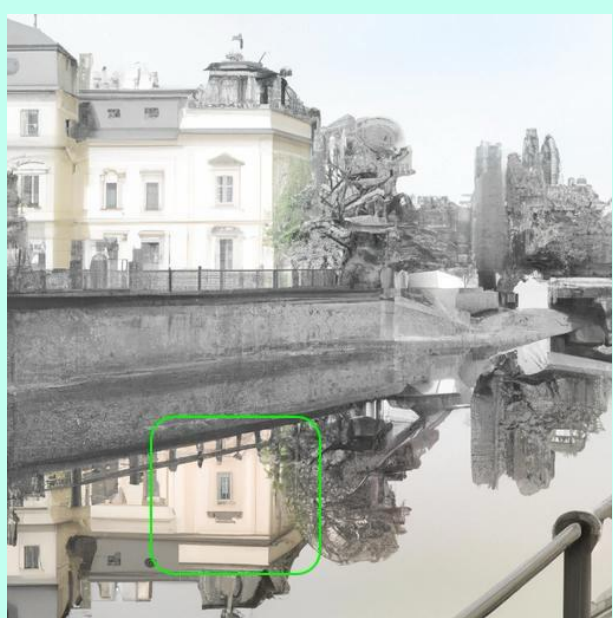


Fig. 2



Fig. 3.



Fig. 4.

Distortions in AIGI

Semantic Level

Fig. 1. Prompt: ``Sheep that looks like broccoli, baroque style'' , but lack of sheep.

Perceptual Level

Fig. 2. Reflection on the water is more clear than the building.

Fig. 3. A man ' s right eye missing a visible pupil.

Axiology Level

Fig. 4. The model focuses on facial features at the expense of other body parts, resulting in an incomplete human head

Meta-learning Design, Encoding Analysis for AIGI Quality Prediction

Research Challenges

The poor performance of existing quality prediction models for AIGI can be attributed to several challenges. *AIGI quality issues often manifest at both pixel and structural levels*, yet distribution-based metrics like IS and FID overlook individual distortions and anomalies. *AIGI distortion diverges from traditional image processing*, models trained on conventional datasets lack the capacity to pinpoint specific new distortions. Moreover, *semantic quality is crucial for AIGI*, but while metrics like CLIPScore enhance interpretability, they struggle to handle the fine-grained nuances in complex prompts. Consequently, this evaluation gap results in overall quality scores that fail to capture the actual perceptual quality of AIGI outputs.

My Approaches

Given the critical need for generalization, meta-learning, or learning to learn, provides a promising solution. We have chosen the Model-Agnostic Meta-Learning (MAML) framework to improve the model ' s adaptability across diverse image quality assessment tasks.

The dual-loop approach allows the model to capture fine-grained, task-specific quality nuances, such as perceptual and semantic inconsistencies, by using diverse task data for quick adaptability.

Quality Models	PLCC	SROCC
ResNet-50	0.337	0.338
Clip	0.353	0.352
PickScore	0.202	0.206
Ours	0.546	0.530

The poor performance of the existing models on unnatrulness prediction!

A Case of Study: Quality Prediction for AIGI Unnatrualness

