AugmentIQ: Revolutionizing Image Quality Assessment with Advanced Data Augmentation and Dynamic Data Loading Techniques

Minghao Liu, *Student Member, SJTU,*

***Abstract*—In the contemporary realm of Image Quality Assessment (IQA), the confluence of image quality metrics and the alignment of generated images with textual prompts is pivotal. This paper introduces an innovative IQA model, naming AugmentIQ, a synergistic integration of advanced methodologies from two seminal works: the nuanced image quality measurement techniques of Re-IQA and the incisive alignment evaluation of ImageReward. Our model not only assesses the aesthetic and technical quality of images but also quantifies their semantic congruence with given textual descriptors. This dual-capability framework represents a significant advancement in automated image evaluation, offering a more holistic, human-centric approach to assessing image generation models. It addresses the growing need for comprehensive tools capable of understanding the multifaceted nature of image quality in the era of AI-driven image synthesis. Our code and finetuned models can be obtained at [1].**

***Index Terms*—Image Quality Assessment (IQA),Text-to-Image Synthesis ,Semantic Image Analysis ,Human Preference Feedback ,Deep Learning in Image Processing, Perceptual Quality Metrics ,Structural Similarity Index (SSIM) ,Content and Quality Alignment ,Visual Perception Modeling**

I. Introduction

THE look and feel are two factors that humans consider when interpreting an image, and understanding these elements has been a long-standing issue in computer vision. The look of an image is often related to quantifiable attributes that directly affect the content delivery, such as exposure and noise level. In contrast, the feel of an image is an abstract concept immaterial to the content and cannot be easily quantified. It is of interest to explore the possibility of a universal understanding of both look and feel, as it can save efforts on manual labeling and facilitate the development of vision tasks such as restoration. Considerable efforts have been devoted to the assessment of both the quality perception (i.e., look) and the abstract perception (i.e., feel) of images. The evolution of Image Quality Assessment (IQA) from its nascent pixel-focused analyses to the current complex, context-driven evaluations reflects the rapid advancements in image generation technologies. Traditional IQA models predominantly focused on objective quality metrics, often neglecting the subjective, perceptual aspects integral to human interpretation. With the advent of AI-driven image generation, particularly text-to-image models, like GANs([2],[3],[4]), Autoregressive models([5], ([6]), Diffusion models(([7],[8],[9]), there is an exigent need for IQA models that not only assess image fidelity but also the alignment of generated content with human expectations and textual descriptions. This paper presents an integrated IQA model that builds upon the methodologies of Re-IQA and ImageReward, amalgamating the assessment of image quality with the evaluation of text-image alignment. This model marks a significant stride in the field, catering to the nuanced demands of contemporary image generation and consumption.

Minghao Liu is with Shanghai JiaoTong University (e-mail: lmh209@sjtu.edu.cn).

II. Related work

The evolution of Image Quality Assessment (IQA) is a testament to the burgeoning complexity and capabilities of image generation and processing technologies. Our integrated model, drawing upon the advancements of Re-IQA([10]) and ImageReward([11]), is built upon a rich legacy of research in the field, each contributing vital insights into different aspects of perceptual quality assessment.

A. Early Developments in Objective IQA

Historically, IQA models primarily focused on objective quality metrics. A seminal contribution in this domain was the Structural Similarity Index (SSIM)([12]) introduced by Wang et al. This model marked a departure from traditional error-sensitivity approaches, offering a method that considered changes in structural information, luminance, and contrast. Another notable metric, the Peak Signal-to-Noise Ratio (PSNR), though simplistic, set the groundwork for subsequent, more nuanced models.

B. Bridging Objective Metrics with Human Perception

Recognizing the limitations of purely objective assessments, researchers began to focus on bridging these metrics with human perceptual quality. Sheikh and Bovik's work on Information Fidelity Criteria ([13]) exemplifies this trend. They aimed to align objective quality assessments with human visual perception, acknowledging that image quality is inherently subjective. Similarly, Mittal et al.'s BRISQUE model[14]introduced a no-reference approach to IQA, relying on natural scene statistics and not requiring a pristine reference image.

C. Advancements in Text-to-Image Synthesis Evaluation

With the advent of complex text-to-image synthesis models, the focus of IQA further evolved. The development of models like DALL-E([5],[6]) by OpenAI highlighted the need for assessing the semantic alignment between text prompts and generated images, a challenge distinct from traditional IQA. ImageReward[11], a model from my integrated framework, directly addresses this by evaluating the alignment score between prompts and images. This method builds upon the insights from previous works that combined natural language processing with image analysis, reflecting a more holistic approach to IQA.

D. Contributions

Our Contribution is listed as follows:

1) Our integrated model, synthesizing the methodologies of Re-IQA and ImageReward, represents the next step in this evolutionary path. It not only incorporates the technical advancements in assessing image fidelity and aesthetic quality but also introduces a novel dimension of evaluating text-image semantic congruence.
2) This integration signifies a broader trend in IQA research, one that acknowledges the multi-dimensional nature of image quality in the age of AI and seeks to develop assessment tools that are as dynamic and multifaceted as the images they evaluate.

In summary, the development of my integrated IQA model is informed and enriched by these diverse but interconnected strands of research. It stands as a culmination of years of exploration and innovation in the field, marking a new era in image quality assessment that is attuned to the complexities of modern image generation and processing technologies.

III. Our Architecture and Methods

A. Content and quality aware metrics for image evaluation

Following the setting of [10], I reformulate the question and made several improvements. The core of this proposed framework lies in the amalgamation of two primary components: a content-aware encoder C and a quality-aware encoder Q. Both encoders are functionally represented as deep neural networks, with C focusing on extracting high-level semantic features from images, and Q dedicated to discerning fine-grained quality attributes. The pipeline for reiqa can be best generalized in 1.

1. **Content-Aware Encoder** C: , the content-aware encoder is designed as $C : I \rightarrow F_C$, where $I \in R^{3\times H\times W}$ represents the input image and F_C denotes the high-level feature representation of the content. This encoder leverages a contrastive learning approach, similar to that used in the MoCo framework ([15]), to ensure that the extracted features F_C accurately encapsulate the semantic essence of the image. With this contrastive fashion, the original paper trains this network on ImageNet[16].

2. **Quality-Aware Encoder** Q: The quality-aware encoder is mathematically expressed as $Q : I \rightarrow F_Q$, where F_Q signifies the nuanced quality feature vector. Drawing from the

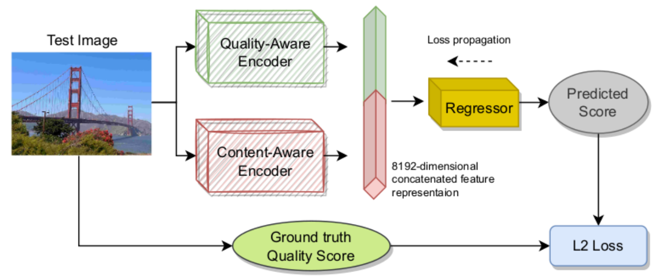

Figure 1. IQA score prediction uses two encoders trained for complementary tasks of learning content and quality aware image representations. The encoders are frozen while the regressor learns to map image representations to quality predictions.

Fig. 1: ReIQA model pipeline: they propose an unsupervised low-level image quality representation learning framework that generates features complementary to high-level representations of image content. I demonstrate how the "Mixture" of the two enables Re-IQA to produce image quality predictions that are highly competitive with existing SOTA traditional.

principles of no-reference quality assessment[17], Q is trained to identify and quantify various aspects of image quality, such as noise, sharpness, and color fidelity, with a non-reference fashion. To get an more detailed comprehension on how they select the positive samples and negative samples, I strongly recommend you to refer to the original paper.

3. **Integration and Regression Model**: The crux of the methodology is the integration of these two encoders into a unified framework. The model concatenates the outputs F_C and F_Q to form a composite feature vector $F = [F_C; F_Q]$. This vector is then mapped to a perceptual quality score S through a linear regression model R, formulated as $R : F \rightarrow S$. This mapping is crucial, as it translates the combined content and quality features into a single, interpretable quality score that aligns with HVS.

4. **Training and Optimization**: The training process involves optimizing the encoders C and Q and the regression model R using a dataset D comprising images and their corresponding human-rated quality scores, more specifically, **std_quality**, **mos_quality**. The optimization objective is defined to minimize the discrepancy between the predicted quality scores and the ground-truth human ratings, typically employing a loss function such as Mean Squared Error (MSE).

B. Text-to-image alignment and correlation score

The pipeline for ImageReward can be best generalized in 2.

1. **Representation of Image and Text**: Let $I \in R^{3\times H\times W}$ represent a generated image, and T represent the corresponding textual prompt. The model first embeds both the image and the text into a shared feature space, similar to CLIP[18] and more advanced BLIP[19], upon which the ImageReward has been built on. This embedding is realized through functions

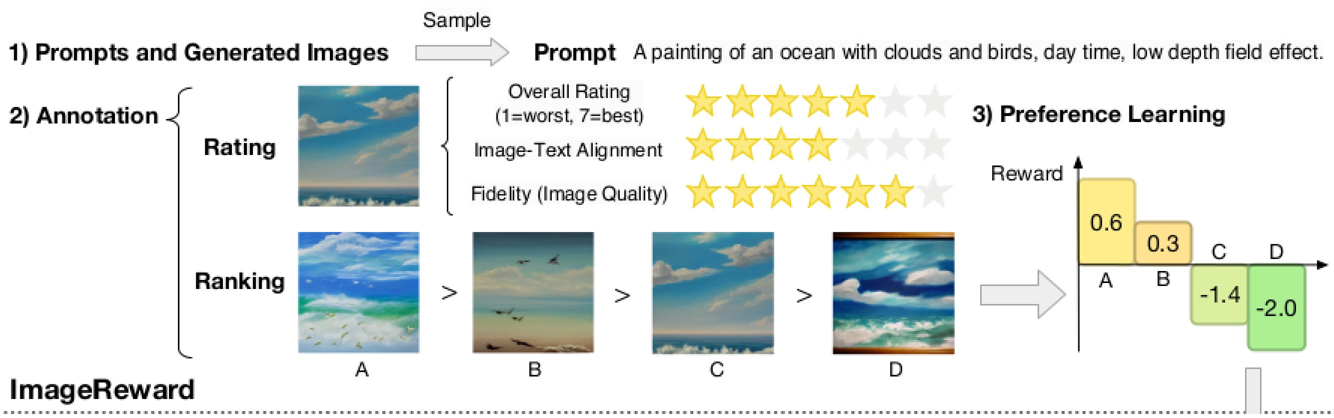

Fig. 2: An overview of the ImageReward[11] model, ImageReward's annotation and training pipeline, consisting of data collection, annotation, and preference learning. At the issue of their paper, they have reported achieving SOTA result.

$f_I(I)$ and $f_T(T)$, which map the image and text into vectors in a high-dimensional space.

2. **Alignment Scoring Function**: The core of ImageReward is a scoring function S, defined as $S(I,T)$. This function computes a score reflecting the degree of alignment between the image I and the text T. Conceptually, S is modeled to capture both the perceptual quality of the image and its semantic relevance to the prompt.

3. **Optimization Objective**: They formulate the preference annotations as rankings. I have $k \in [4,9]$ images ranked for the same prompt T (the best to the worst are denoted as $x_1, x_2, \dots x_k$) and get at most C_k^2 comparison pairs if no ties between two images. For each comparison, if x_i is better and x_j is worse, the loss function is thus formulated as:

$$\text{loss}(\theta) = -E_{(T,x_i,x_j)\sim\mathcal{D}}[\log(\sigma(f_\theta(T,x_i) - f_\theta(T,x_j)))]$$

This loss function learns to differentiate the ranking scores produced by ImageReward models. This process fine-tunes the model to align closely with HVS.

C. Image augmentation

I have employed extensive image augmentation[20] approaches, including 26 approaches covering literally the wonderful realm of image augmentation. The augmentation methods employed, ranging from Gaussian Blur to Jitter Effect, can be formalized within the framework of transformation functions T. Each transformation T_i, where $I \in \{1,2,\dots,26\}$ represents a specific augmentation type (e.g., Gaussian Noise, Resize Distortion), acts on an image I to produce a transformed image $I' = T_i(I)$. The extent of each augmentation is parameterized, allowing for multiple levels of transformation intensity, denoted as T_i^e, where e indexes the extent.

Given a dataset D, consisting of images $\{I_1, I_2, ..., I_n\}$, the augmented dataset D' can be represented as:

$$D' = \bigcup_{I\in D} \bigcup_{i,e} T_i^e(I)$$

This extensive augmentation increases the dataset's diversity, providing a richer and more varied gradient landscape during training. The augmentation's impact on model performance is rooted in its ability to simulate a wide range of real-world distortions, thereby enhancing the model's generalization capabilities. This hypothesis aligns with the principles outlined in SimCLR[21] and MoCo v2[15], where augmentations proves effective and significant in learning robust representations.

D. Data Loading Strategy

The data loading strategy involves shuffling the augmented dataset D' and strategically sampling mini-batches to ensure diverse feature representation. Instead of sampling the different augmented versions of same image, shuffling across the whole dataset allows the data loader to produce completely different samples from that gigantic manifolds of data, both in the context of augmentation methods and image samples. This also allows a more varied and diversified gradient, thus allowing for a more robust and reliable training procedure.

E. Loss computing strategy

Due to the specific purpose of my need, I use differentiable Pearson Correlation Coefficient and differentiable Spearman's Correlation Coefficient from `torchmetrics` module to compute my loss, instead of the traditional MSE loss or CrossEntropy loss. And due to current instability in the training process, the loss functions with different datasets and hyperparameters choices may subject to change.

IV. Experiment Results

I have run my experiments on two datasets, the AGIQA-3K database[22] and the AIGCIQA2023 database [23], some samples from the datasets can be seen in 3,4.

The training procedure is rather stable, however the validation graph and the test metrics graph suffers from huge variances and instability. For the training procedure for AGIQA-2023, please refer to 5, for AGIQA-3K, please refer to 6.

I report my test metrics on AGIQA-2023 dataset for quality and content aware features extractor as: `SPCC`:0.3046, PLCC:0.3051. and the text alignment features extractor as: `SPCC`:0.4951, PLCC:0.5461. For fairness, I run the experiment five times and take the avaerage.

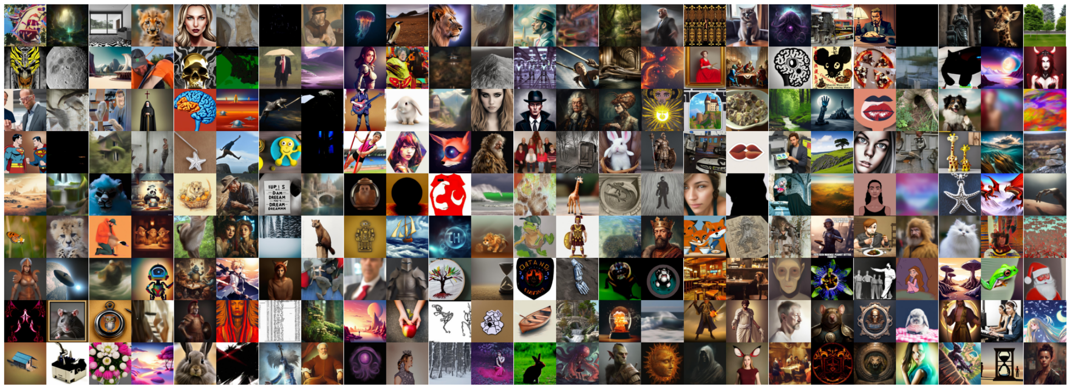

Fig. 3: Sample images from the AGIQA-3K database, where the first to sixth rows show AGIs created by (AttnGAN [3], DALLE2 [6], GLIDE [24], Midjourney, Stable Diffusion[8] and Stable Diffusion XL [25]). Better experience when zoomed in.

Fig. 4: Samples taken from the AIGC-3K[22] database with corresponding prompts. Better experience when zoomed in.

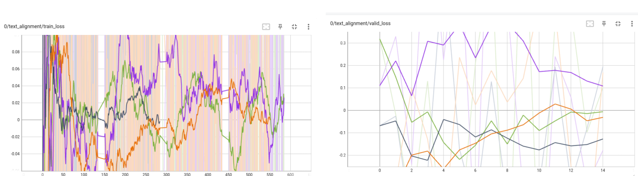

Fig. 5: Evaluation dataset: AGIQA-2023. Left column is the training loss graph, while the right column is the valid loss graph.

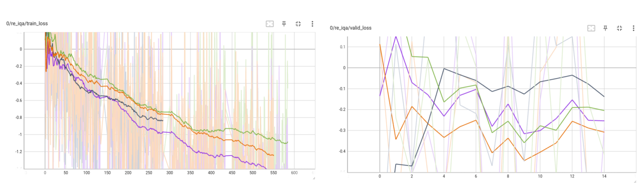

Fig. 6: Evaluation dataset: AGIQA-3K . Left column is the training loss graph, while the right column is the valid loss graph.

I report my test metrics on AGIQA-3K dataset for quality and content aware features extractor as: SPCC:0.4032, PLCC:0.3241., and the text alignment features extractor as: SPCC:0.4719, PLCC:0.5002. For fairness, I run the experiment five times and take the average.

I have arrived at some conclusions after running some parameter fine-tuning results, the problem space of different questions can be quite different, so though my two heads(content and quality aware head and text alignment head) are very heavy and took a while to forward a tensor, its performance on transfer learning across datasets(such as AIGCIQA2023,AIGC3k,DiffusionDB,ImageReward,Pick-A-Pic) can be out of expectation, also it is noteworthy that during the process of constructing the dataset, different settings, prompts, rating schemas can make a big difference in outcomes, so according to my opinion, better results can be achieved by training the model from scratch, instead of frozening weights of some pretrained models designed to perform well only on specific datasets and then fine-tuning some final layers.

Also it is noteworthy that the choices of loss function is also pivotal in the final results, the text to image alignment task can be achieved by training on the `SPCC`, `PLCC` metrics directly, but the content and quality related task seldom boasts a well-formulated metrics, we believe that in our task, a better metrics needs to be chosen for better results.

V. Future work and Practical Applications

A. Future work

Building upon the integrated framework that marries the strengths of Re-IQA and ImageReward, my future endeavors will pivot around refining and extending the model's capabilities. I anticipate delving into the realm of dynamic content, where the evolution of visual elements over time in videos presents a nuanced challenge for IQA models. Leveraging advancements in temporal analysis, similar to those discussed by Wang et al. in their study of video quality assessment [26], my model will aim to discern quality and content alignment in moving images.

Another prospective avenue is the incorporation of cross-modal feedback, synthesizing inputs from auditory and textual domains to provide a holistic sensory evaluation. This approach resonates with the multi-modal promising IQA methodologies proposed by a lot of researchers.

B. Pracical applications and future impacts

The practical applications of my integrated model are manifold and transformative. In digital marketing, where the congruence between visual content and textual description significantly impacts consumer engagement, my model can offer invaluable insights into content optimization. The model's adeptness at understanding human aesthetics and preferences will also revolutionize personalization algorithms in social media platforms like X, Instagram or facebook.

Moreover, in educational technology, where visual aids play a critical role, my model can ensure the quality and relevance of image-based learning materials, echoing the sentiment of Mayer's principles of multimedia learning [27]. Additionally, the model has significant implications for assistive technologies, enhancing content accessibility for visually impaired users by ensuring the alignment and quality of descriptive visual content.

VI. Concluding remarks

In conclusion, the fusion of Re-IQA and ImageReward models represents a significant leap forward in the field of Image Quality Assessment (IQA). This integrated model transcends traditional boundaries by not only evaluating the perceptual quality of images but also ensuring their semantic alignment with textual descriptions. This synergy mirrors the advancements in deep learning-based IQA methods, which have progressively shifted from pixel-based evaluations to understanding perceptual and contextual nuances[28].

Our integrated model stands as a testament to the power of interdisciplinary research, combining insights from human preference studies, computer vision, and machine learning. It embodies the evolving landscape of IQA, where the focus has gradually shifted from objective fidelity measures to subjective human-centric evaluations [29].

As I look to the future, the potential applications of this model span diverse sectors, promising not only to enhance user experiences but also to inform content creators and technologists about the multifaceted nature of image quality and alignment. This model, thus, does not merely add to the existing compendium of IQA methodologies but pioneers a new path, setting the stage for future innovations and applications that resonate with the evolving digital ecosystem.

Acknowledgment

I extend my heartfelt gratitude to the esteemed faculty and dedicated teaching assistants of CS3324 for their invaluable guidance and support throughout my journey in image processing. Their profound knowledge, coupled with an unwavering commitment to nurturing curiosity and innovation, has been instrumental in my academic and personal growth. I am deeply appreciative of their efforts in creating a stimulating and enriching learning environment, which has significantly contributed to the development of this paper and my understanding of the field. My sincere thanks to each one of them for inspiring and challenging me to reach new heights in my studies.

References

[1] MinghaoLiu, "AugmentIQ: Revolutionizing Image Quality Assessment with Advanced Data Augmentation and Dynamic Data Loading Techniques," 11 2023. [Online]. Available: https://github.com/Learner209/AugmentIQ

[2] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," 2020.

[3] T. Xu, P. Zhang, Q. Huang, H. Zhang, Z. Gan, X. Huang, and X. He, "Attngan: Fine-grained text to image generation with attentional generative adversarial networks," 2017.

[4] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," 2019.

[5] A. Ramesh, M. Pavlov, G. Goh, S. Gray, C. Voss, A. Radford, M. Chen, and I. Sutskever, "Zero-shot text-to-image generation," 2021.

[6] A. Ramesh, P. Dhariwal, A. Nichol, C. Chu, and M. Chen, "Hierarchical text-conditional image generation with clip latents," 2022.

[7] M. Ding, Z. Yang, W. Hong, W. Zheng, C. Zhou, D. Yin, J. Lin, X. Zou, Z. Shao, H. Yang, and J. Tang, "Cogview: Mastering text-to-image generation via transformers," 2021.

[8] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," 2022.

[9] C. Saharia, W. Chan, S. Saxena, L. Li, J. Whang, E. Denton, S. K. S. Ghasemipour, B. K. Ayan, S. S. Mahdavi, R. G. Lopes, T. Salimans, J. Ho, D. J. Fleet, and M. Norouzi, "Photorealistic text-to-image diffusion models with deep language understanding," 2022.

[10] A. Saha, S. Mishra, and A. C. Bovik, "Re-iqa: Unsupervised learning for image quality assessment in the wild," 2023.

[11] J. Xu, X. Liu, Y. Wu, Y. Tong, Q. Li, M. Ding, J. Tang, and Y. Dong, "Imagereward: Learning and evaluating human preferences for text-to-image generation," 2023.

[12] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[13] H. Sheikh and A. Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430–444, 2006.

[14] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 4695–4708, 2012.

[15] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, "Momentum contrast for unsupervised visual representation learning," 2020.

[16] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "Imagenet large scale visual recognition challenge," 2015.
[17] J. Wu, Z. Xia, Y. Ren, and H. Li, "No-reference quality assessment for contrast-distorted image," 12 2016, pp. 1–5.
[18] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, and I. Sutskever, "Learning transferable visual models from natural language supervision," 2021.
[19] J. Li, D. Li, C. Xiong, and S. Hoi, "Blip: Bootstrapping language-image pre-training for unified vision-language understanding and generation," 2022.
[20] S. Yang, W. Xiao, M. Zhang, S. Guo, J. Zhao, and F. Shen, "Image data augmentation for deep learning: A survey," 2023.
[21] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," 2020.
[22] C. Li, Z. Zhang, H. Wu, W. Sun, X. Min, X. Liu, G. Zhai, and W. Lin, "Agiqa-3k: An open database for ai-generated image quality assessment," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. PP, pp. 1–1, 01 2023.
[23] J. Wang, H. Duan, J. Liu, S. Chen, X. Min, and G. Zhai, "Aigciqa2023: A large-scale image quality assessment database for ai generated images: from the perspectives of quality, authenticity and correspondence," 2023.
[24] A. Nichol, P. Dhariwal, A. Ramesh, P. Shyam, P. Mishkin, B. McGrew, I. Sutskever, and M. Chen, "Glide: Towards photorealistic image generation and editing with text-guided diffusion models," 2022.
[25] D. Podell, Z. English, K. Lacey, A. Blattmann, T. Dockhorn, J. Müller, J. Penna, and R. Rombach, "Sdxl: Improving latent diffusion models for high-resolution image synthesis," 2023.
[26] Z. Wang, H. Sheikh, and A. Bovik, "No-reference perceptual quality assessment of jpeg compressed images," in *Proceedings. International Conference on Image Processing*, vol. 1, 2002, pp. I–I.
[27] R. E. Mayer, "Multimedia learning," ser. Psychology of Learning and Motivation. Academic Press, 2002, vol. 41, pp. 85–139. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0079742102800056
[28] Y. Li, S. Wang, Q. Tian, and X. Ding, "A survey of recent advances in visual feature detection," *Neurocomputing*, vol. 149, pp. 736–751, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231214010121
[29] H. Sheikh, M. Sabir, and A. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3440–3451, 2006.

APPENDIX

EXPERIMENT SETTINGS

I detail the experimental settings below. These settings encompass various factors that could significantly influence the model's performance. GPU: NVIDIA RTX 3090, GPUs Used: 1, OS: Ubuntu 22.04 LTS, RAM: 128GB DDR4. Language: Python 3.8, Framework: PyTorch 2.0.1. For further details, I encourge readers to visit out github repos. Batch Size: 64, Channel dimensions(number of augmentation methods present in one mini-batch): 80, Learning Rate: 0.001, Optimizer: Adam, Loss Function: Mean Squared Error, Training Epochs: 100, Validation Split: 80/20. Evaluation Metric: Spearman's Rank Correlation, Additional Metrics: Pearson Correlation, RMSE. Seed: 42, Regularization: Dropout, L2, Distributed Training: Yes. The current training procedure is subject to large variances and instability, so it might be difficult to reproduce the optimal result, I believe it would still encompass extensive trials and errors to arrive at the ideal realm.

A DETAILED EXPLORATION ON THE IMAGE AUGMENTATION METHODS I HAVE EMPLOYED

1) Gaussian Blur: Applies a Gaussian filter to blur the image, simulating the effect of viewing through a translucent screen.
2) Lens Blur: Mimics the blur produced by camera lens aberrations or depth-of-field effects.
3) Motion Blur: Simulates the blur effect due to the motion of objects or camera shake during exposure.
4) Color Diffusion: Spreads color in an image to create a soft, dreamlike appearance.
5) Color Shift: Alters the overall color balance and tone of the image.
6) Color Saturation Increase: Enhances the intensity of colors in the image.
7) JPEG Compression: Introduces artifacts typically seen in JPEG compressed images, such as blockiness.
8) Gaussian Noise: Adds normally distributed noise throughout the image, simulating sensor noise.
9) Color Map Noise: Introduces noise based on color mapping distortions.
10) Impulse Noise: Adds sparse but intense disturbances, resembling salt-and-pepper noise.
11) Multiplicative Noise: Applies noise that multiplies pixel values, affecting image brightness and contrast.
12) Noise Reduction: Smoothens the image to reduce noise while preserving details.
13) Brightness Increase: Enhances the overall brightness of the image.
14) Darkness Increase: Reduces the overall brightness, simulating low-light conditions.
15) Mean Shift: Adjusts the mean of the pixel intensity distribution.
16) Resize Distortion: Resizes the image, causing distortion based on the resizing algorithm used.
17) Bilinear Resize Distortion: Resizes the image using bilinear interpolation, a common resizing technique.
18) Nearest Neighbor Resize Distortion: Uses nearest-neighbor interpolation for resizing, preserving hard edges.
19) Lanczos Resize Distortion: Employs the Lanczos resampling method for high-quality image resizing.
20) High Sharpening: Increases the sharpness of the image, enhancing edges and details.
21) Contrast Change: Alters the contrast level of the image.
22) Color Blocking: Introduces large, uniform blocks of color, simulating a low-color-depth effect.
23) Pixelation: Reduces the image resolution, creating a blocky, pixelated appearance.
24) Non-Eccentricity Adjustment: Modifies the aspect ratio of objects in the image without preserving their original shape.
25) Warping via Map: Distorts the image based on a predefined warping map.
26) Jitter Effect: Randomly shifts the pixels of the image, creating a jittery or shaky effect.

Minghao Liu Minghao Liu is an undergraduate majoring in IEEE-CS at Shanghai Jiao Tong University.. His research interests include AI interpretability, ML, Rl and software developemnts. You can reach him on yufeiyang553@gmail.com.

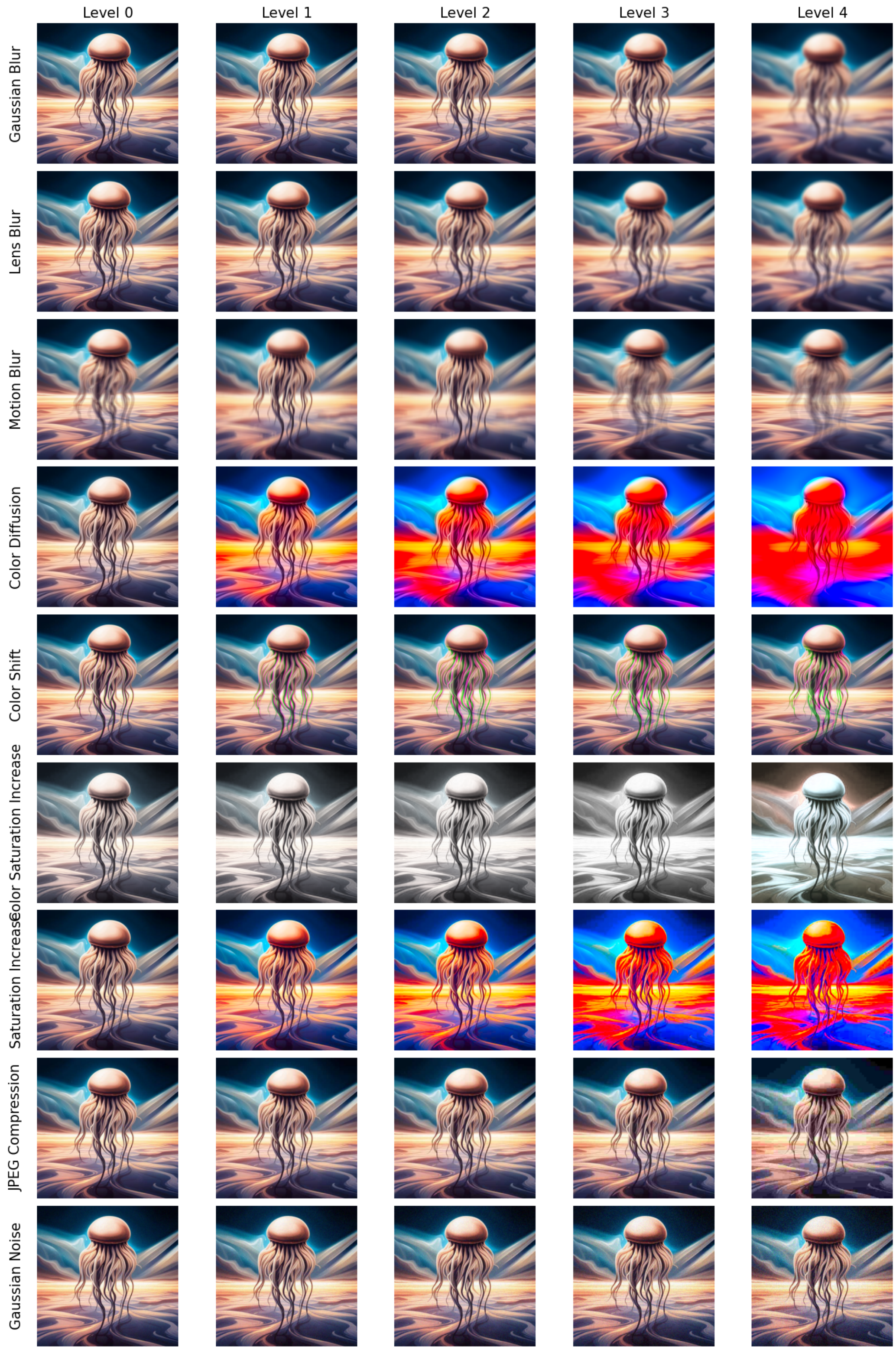

Fig. 7: Several distortions methods employed in the paper.

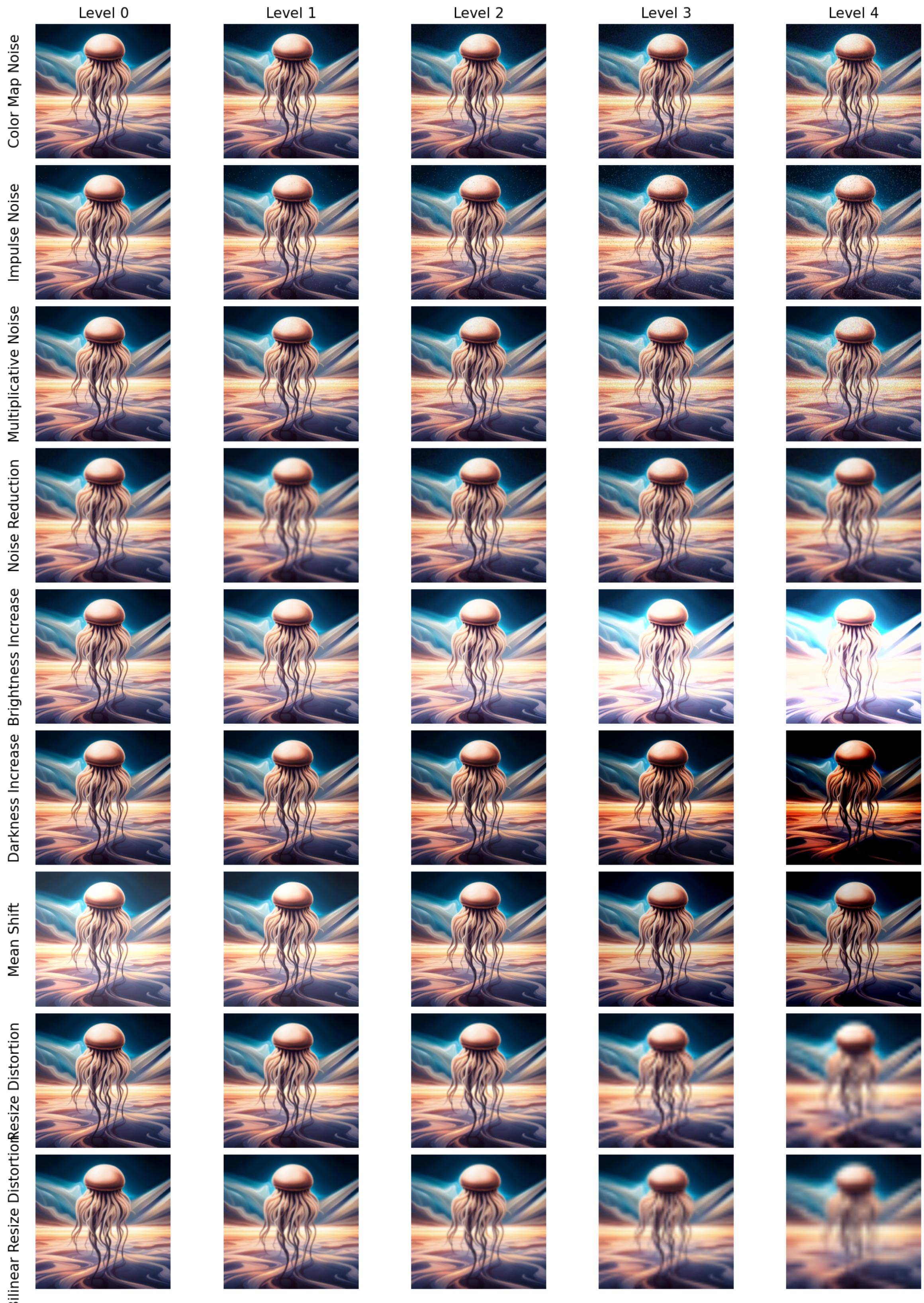

Fig. 8: Several distortions methods employed in the paper.

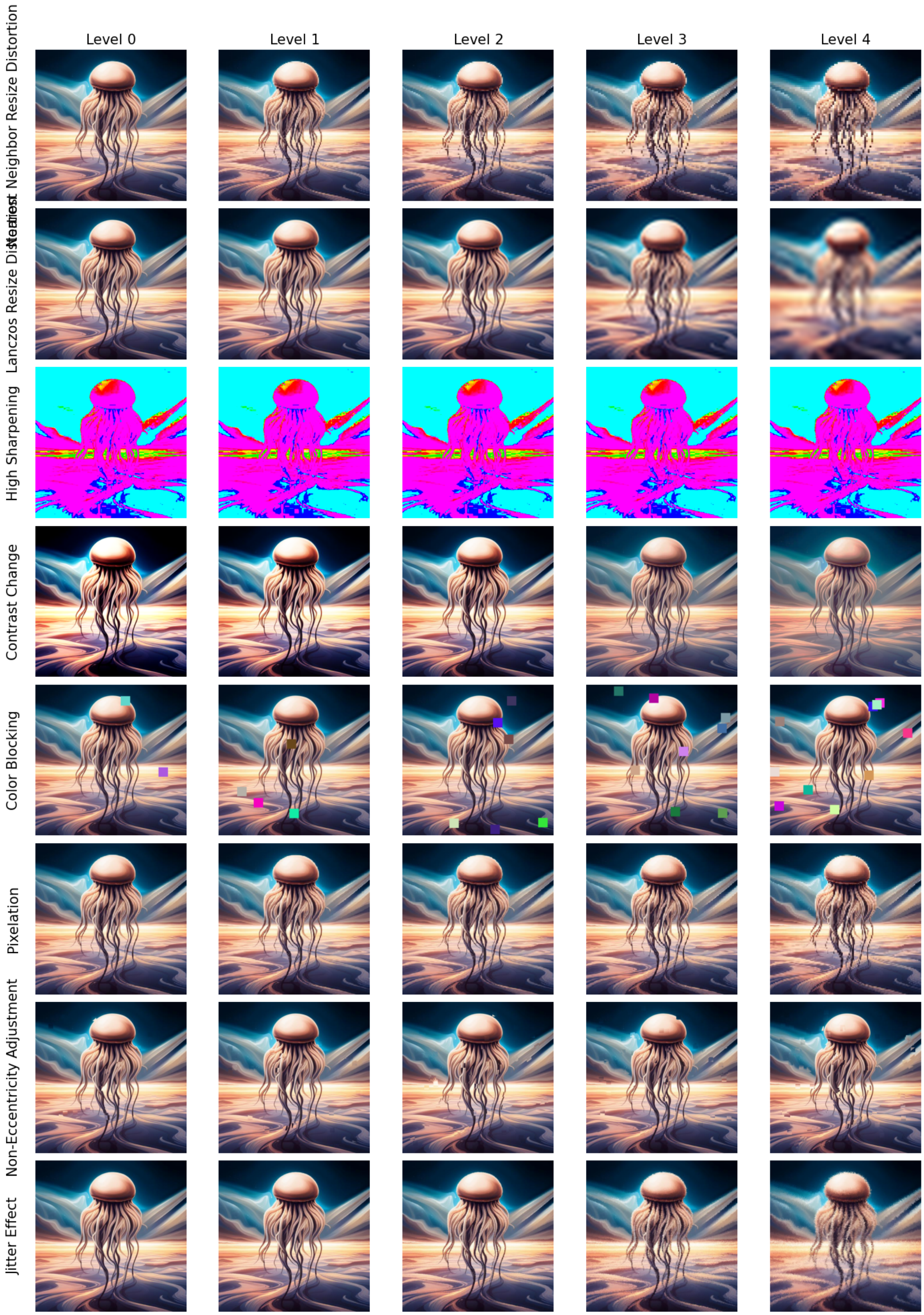

Fig. 9: Several distortions methods employed in the paper.