Distilling Spatially-Heterogeneous Distortion Perception for Blind Image Quality Assessment
- Supplementary Material -

1. Distortion Classes for Degradation Modeling

Following [2], we employed 24 different types of quality distortion augmentations for training our quality-aware submodule, with each distortion type applied at four intensity levels. An example of an image synthesized using the block-wise degradation model is shown in Figure 1. The specifics of each distortion type are detailed below.

- **Resize Bicubic**: Shrinks the image and then enlarges it back to its original dimensions using bicubic interpolation.
- **Resize Bilinear**: Reduces the image size and restores it to its original resolution using bilinear interpolation.
- **Resize Lanczos**: Compresses the image and resizes it back to its initial size using Lanczos filter interpolation.
- **Pixelate**: Reduces the resolution and scales the image back up using nearest-neighbor interpolation.
- **Motion Blur**: Simulates the effect of motion blur by applying a linear filter.
- **Gaussian Blur**: Smooths the image by convolving it with a Gaussian kernel.
- **Lens Blur**: Applies a circular blur filter to mimic the depth of field effects.
- **Mean Shift**: Alters the average intensity of the image by adding constant and clipping values to the valid range.
- **Contrast**: Adjusts the image contrast by applying a non-linear sigmoid-like transformation to the RGB channels.
- **Unsharp Masking**: Enhances the image's edges by applying an unsharp mask technique.
- **Jitter**: Distorts the image by randomly shifting pixel positions slightly.
- **Color Block**: Overlays randomly sized and colored solid blocks on the image.
- **Non-eccentricity**: Introduces small random displacements to patches of the image.
- **JPEG Compression**: Simulates compression artifacts by saving and reloading the image in JPEG format.

Mode	Methods	KonIQ		LIVEC	
		PLCC	SRCC	PLCC	SRCC
20%	DEIQT	0.888	0.908	0.792	0.822
	LoDa	0.907	0.923	0.815	0.854
	SHDIQA	**0.915**	**0.928**	**0.840**	**0.866**
40%	DEIQT	0.903	0.922	0.838	0.855
	LoDa	0.922	0.935	0.849	0.879
	SHDIQA	**0.929**	**0.942**	**0.869**	**0.889**
60%	DEIQT	0.914	0.931	0.848	0.877
	LoDa	0.928	0.940	0.869	0.891
	SHDIQA	**0.933**	**0.945**	**0.891**	**0.910**

Table 1. Data-efficient learning validation with the training set containing 20% , 40% and 60% of the images.

- **White Noise (RGB space)**: Injects Gaussian white noise into the image's RGB channels.
- **White Noise (YCbCr space)**: Adds Gaussian white noise to the YCbCr color space channels.
- **Impulse Noise**: Introduces salt-and-pepper noise, creating random black and white pixels.
- **Multiplicative Noise**: Applies speckle noise by multiplying pixel values with random noise.
- **Denoise**: Adds Gaussian noise and applies a blur filter (Gaussian or box) to smooth it out.
- **Brighten**: Increases overall brightness using a non-linear transformation that preserves extreme values.
- **Darken**: Reduces brightness in a similar manner to Brighten, but decreases pixel intensities.
- **Color Diffuse**: Blurs the color channels (a and b) in the LAB color space with a Gaussian filter.
- **Color Shift**: Shifts the green channel and blends it with the original image using a gradient-based mask.
- **Saturate**: Amplifies the color intensity in the LAB space by scaling its color channels.

Evaluation Criteria	Training Ratio	HOSA [6]	CORNIA [8]	MetaIQA [10]	HyperNet [5]	DBCNN [9]	KG-IQA [3]	AL-IQA [4]	DEIQT [1]	Proposed SHDIQA
KonIQ										
PLCC	5%	0.73	0.721	0.796	0.800	0.829	0.85	0.859	0.852	**0.892**
	10%	0.751	0.743	0.821	0.842	0.843	0.876	0.888	0.885	**0.907**
	25%	0.777	0.765	0.861	0.883	0.868	0.901	0.915	0.910	**0.930**
SRCC	5%	0.685	0.682	0.761	0.768	0.811	0.825	0.831	0.839	**0.874**
	10%	0.708	0.701	0.788	0.814	0.828	0.851	0.857	0.865	**0.890**
	25%	0.737	0.720	0.830	0.859	0.852	0.877	0.891	0.890	**0.917**
SPAQ										
PLCC	5%	0.806	0.812	0.875	0.867	0.873	0.883	0.894	0.886	**0.903**
	10%	0.827	0.827	0.887	0.885	0.885	0.895	0.903	0.900	**0.916**
	25%	0.848	0.843	0.898	0.901	0.898	0.911	0.917	0.912	**0.924**
SRCC	5%	0.800	0.805	0.872	0.867	0.874	0.877	0.889	0.880	**0.899**
	10%	0.821	0.82	0.885	0.885	0.885	0.891	0.897	0.896	**0.912**
	25%	0.842	0.836	0.895	0.899	0.900	0.907	0.910	0.908	**0.920**

Table 2. Performance on KonIQ and SPAQ. All methods are trained with 5%, 10%, or 25% of the images and tested on the other images. We mark the best result in bold, second-best is underlined.

2. Data-Efficient Learning Validation

In this section, we further validate the effectiveness of our method in efficiently learning from limited data across three datasets. As shown in Table 2, our approach consistently achieves superior PLCC and SRCC performance in data-efficient experiments, outperforming state-of-the-art IQA methods. Notably, our method achieves the best results on two of the datasets. Among the other methods, AL-IQA, which emphasizes selecting representative samples for efficient learning, ranks second in performance. KG-IQA leverages natural scene statistics and the human visual system's principles, while MetaIQA benefits from strong pre-training initialization. In contrast to these methods, our approach does not depend on specific pre-training strategies or selective data filtering, underscoring its robust generalization across diverse BIQA tasks.

To facilitate a fair comparison with existing BIQA methods based on the ViT architecture, particularly in terms of data efficiency, we conducted additional evaluations under similar experimental conditions. Following the data-efficient protocols in LoDa [7] and DEIQT [1], our model was trained on 20%, 40%, and 60% subsets of the data, with each experiment repeated 10 times to report average SRCC and PLCC scores. By incorporating distortion priors through Block-wise Degradation, our method achieved competitive performance with state-of-the-art (SOTA) approaches, even when trained on only 20% of the dataset. Notably, when utilizing 60% of the data, our method outperformed existing SOTA BIQA methods trained on the full KonIQ dataset (Tab.1 in the manuscript). These results highlight the efficiency and effectiveness of our approach in data-constrained scenarios.

	LIVEC		KonIQ	
Method.	PLCC	SRCC	PLCC	SRCC
Vertical	0.922	0.909	0.948	0.937
Horizontal	0.919	0.905	0.947	0.936
Grid	0.919	0.903	0.949	0.938

Table 3. Ablation study on partitioning strategies.

3. Ablation about the Partitioning Strategy

Our initial choice of vertical partitioning was informed by empirical observations of the dataset, particularly in authentic image quality assessment (IQA) datasets, where horizontal distortions may often more prevalent (e.g., blurred cars in Figure 2 and people in Figure 5 of the main paper). To further evaluate the impact of different partitioning strategies, we conducted an ablation study incorporating both horizontal partitioning and a 2×2 grid partitioning, and correspondingly modified the mask matrix in Block-wise Aggregation. The experimental results, summarized in Table 3, indicate that the choice of partitioning strategy does not significantly impact performance across different datasets. Specifically, both horizontal and grid partitioning exhibited performance very similar to that of vertical partitioning. This finding highlights the robustness of our method.

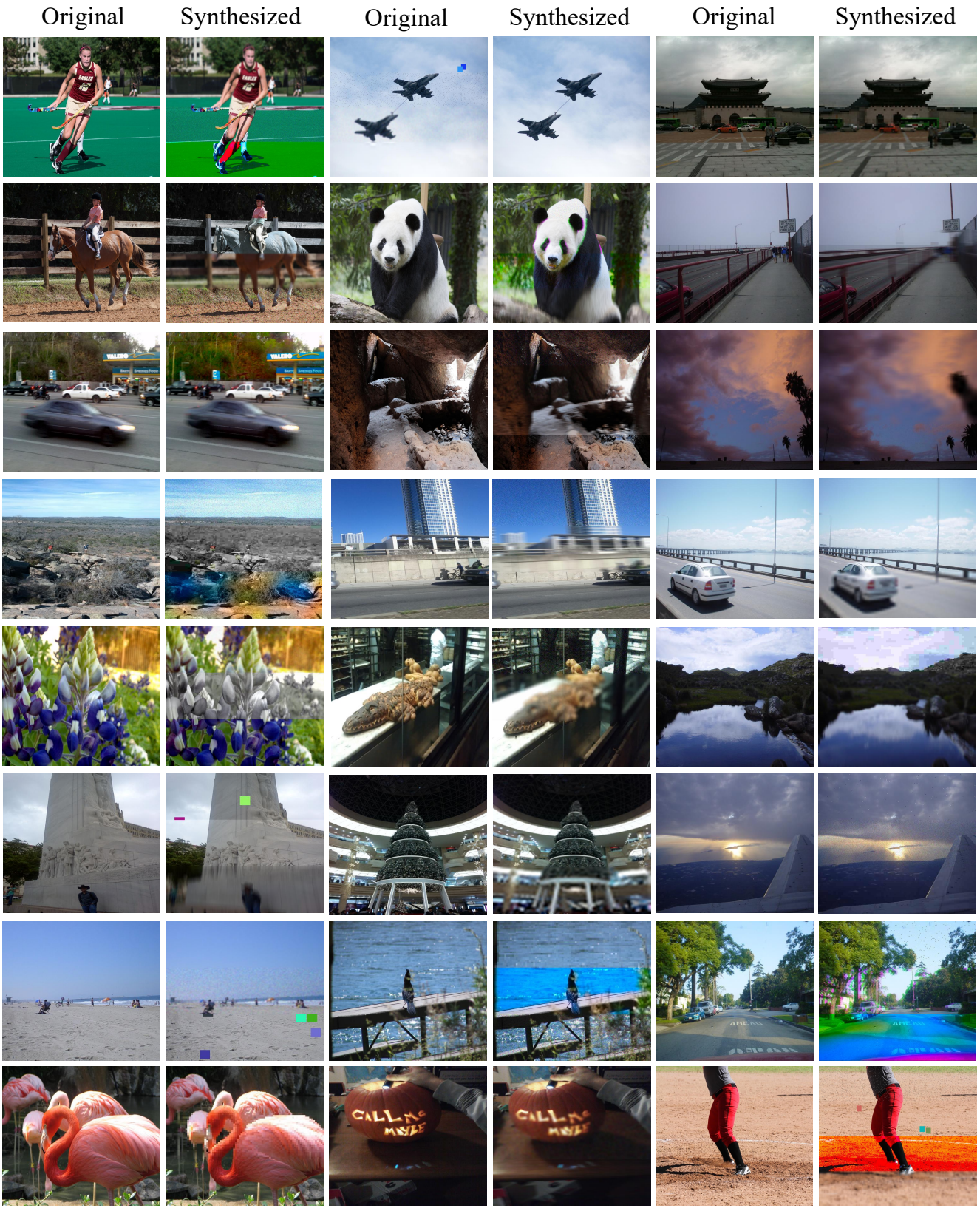

Figure 1. Examples of synthesized images generated by the Block-wise Degradation Model. Zoom in for better visual clarity.

Method	Per Epoch	Total Time
DEIQT	61s / 338s	525s / 3062s
LoDa	85s / 586s	836s / 5798s
Ours	82s / 409s	747s / 3628s

Table 4. Training time on LIVEC / KonIQ dataset (Batch: 16 / 64).

Method	Params	MACs ↓	Throughput ↑
TReS	152.5M	8.39 G	294 (/s)
LoDa	118.1M	23.0 G	276 (/s)
Ours	24.2M	4.35 G	1113 (/s)

Table 5. Inference efficiency on LIVEC dataset (Batch: 16).

4. Computational Time

We compare the training time and inference efficiency of our method with state-of-the-art (SOTA) approaches on the LIVEC and KonIQ datasets. As shown in Table 4, the additional training cost—relative to DEIQT—mainly comes from the extra time required to load degraded images. Moreover, as illustrated in Table 5, our method achieves faster inference speeds than previous hybrid vision transformer-based BIQA methods, highlighting the efficiency advantage of our approach. These results reflect the inherent trade-off between performance and computational cost.

5. Limitations

Although we have demonstrated the superiority of SHDIQA and found that incorporating synthetic distortions into traditional no-reference image quality assessment is highly beneficial, an important issue remains. Specifically, for certain tasks like underwater and medical imaging, there may be limitations due to the unique distortions (such as shifts and artifacts) that differ significantly from the synthetic distortions used in traditional BIQA. Therefore, exploring how to adapt this framework to such specific applications is a worthwhile direction for future research.

6. Broader Impacts

This paper aims to advance the field of Image Quality Assessment (IQA) and examines its societal implications. Improved IQA models can enhance user experiences on digital platforms by ensuring high-quality images, benefiting sectors like online retail, social media, and digital advertising, where visual content quality influences user engagement and satisfaction. However, IQA models may be vulnerable to adversarial attacks, where image quality ratings are manipulated to deceive users or systems. For example, low-quality ads might be falsely rated as high-quality, misleading consumers and diminishing advertising effectiveness. To address these risks, a possible strategy is to implement monitoring systems to detect and respond to anomalies, ensuring the reliability and integrity of IQA models.

References

[1] Guanyi Qin, Runze Hu, Yutao Liu, Xiawu Zheng, Haotian Liu, Xiu Li, and Yan Zhang. Data-efficient image quality assessment with attention-panel decoder. In *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence*, 2023. 2

[2] Avinab Saha, Sandeep Mishra, and Alan C Bovik. Re-iqa: Unsupervised learning for image quality assessment in the wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5846–5855, 2023. 1

[3] Tianshu Song, Leida Li, Jinjian Wu, Yuzhe Yang, Yaqian Li, Yandong Guo, and Guangming Shi. Knowledge-guided blind image quality assessment with few training samples. *IEEE Transactions on Multimedia*, 2022. 2

[4] Tianshu Song, Leida Li, Deqiang Cheng, Pengfei Chen, and Jinjian Wu. Active learning-based sample selection for label-efficient blind image quality assessment. *IEEE Transactions on Circuits and Systems for Video Technology*, 2023. 2

[5] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3667–3676, 2020. 2

[6] Jingtao Xu, Peng Ye, Qiaohong Li, Haiqing Du, Yong Liu, and David Doermann. Blind image quality assessment based on high order statistics aggregation. *IEEE Transactions on Image Processing*, 25(9):4444–4457, 2016. 2

[7] Kangmin Xu, Liang Liao, Jing Xiao, Chaofeng Chen, Haoning Wu, Qiong Yan, and Weisi Lin. Local distortion aware efficient transformer adaptation for image quality assessment. *arXiv preprint arXiv:2308.12001*, 2023. 2

[8] Peng Ye, Jayant Kumar, Le Kang, and David Doermann. Unsupervised feature learning framework for no-reference image quality assessment. In *2012 IEEE conference on computer vision and pattern recognition*, pages 1098–1105. IEEE, 2012. 2

[9] Weixia Zhang, Kede Ma, Jia Yan, Dexiang Deng, and Zhou Wang. Blind image quality assessment using a deep bilinear convolutional neural network. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(1):36–47, 2018. 2

[10] Hancheng Zhu, Leida Li, Jinjian Wu, Weisheng Dong, and Guangming Shi. Metaiqa: Deep meta-learning for no-reference image quality assessment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14143–14152, 2020. 2