Teacher-Guided Learning for Blind Image Quality Assessment Supplementary Material

Zewen Chen[1,2], Juan Wang[1], Bing Li[1✉], Chunfeng Yuan[1], Weihua Xiong[4], Rui Cheng[4], and Weiming Hu[1,2,3]

[1] NLPR, Institute of Automation, Chinese Academy of Sciences
[2] School of Artificial Intelligence, University of Chinese Academy of Sciences
[3] CAS Center for Excellence in Brain Science and Intelligence Technology
[4] Zeku Technology (Shanghai) Corp
{chenzewen2022,jun_wang}@ia.ac.cn, {bli,cfyuan,wmhu}@nlpr.ia.ac.cn, wallace.xiong@gmail.com, chengrui@zeku.com

1 Overview

In this supplementary material, we provide more details about our full model and the variant models in Sec. 2 and Sec. 3. Then we supply more experimental results, including individual distortion evaluations in Sec. 4, comparisons of small data training in Sec. 5, visualized results of restored images in Sec. 6, and gMAD competition results in Sec. 7.

2 More Details about Our Full Model

Our model consists of two networks: a teacher network (TN) and a student network (SN). TN and SN share the same encoder. TN includes an image restorer, and SN includes an image quality predictor. We adopt the ResNet-50 as the shared encoder, and design a multi-branch convolution (MC) based image restorer and an attention mechanism (Att) based quality predictor. The details about the network architecture for the encoder, MC, Att, image restorer and image quality predictor are presented in Tab. 1, Tab. 2, Tab. 3, Tab. 4 and Tab. 5, respectively. (The convolution uses default parameters unless otherwise noted.)

Table 1. Network architecture of the encoder

Input	Name	Output
Distorted Image	Conv+BN+MaxPool	R0
R0	ResNet Layer1	R1
R1	ResNet Layer2	R2
R2	ResNet Layer3	R3
R3	ResNet Layer4	R4

Table 2. Network architecture of the image restorer.

Input	Name	Description	Output
R4	Conv	MC w/o the 1st concat	C1
C1	UpSample	upsample(×2)	U1
U1, R3	MC	-	O1
O1	UpSample	upsample(×2)	U2
U2, R2	MC	-	O2
O2	UpSample	upsample(×2)	U3
U3, R1	MC	-	O3
O3	UpSample	upsample(×2)	U4
U4, R0	MC	-	O4
O4	UpSample	upsample(×2)	U5
U5	Conv	conv kernel size=1	Restored Image

Table 3. Network architecture of the MC.

Input	Name	Description	Output
R3&U1/R2&U2/R1&U3/R0&U4	Concate	dim=1	F1
F1	Inception_Path1	conv kernel size=1	I1
F1	Inception_Path2	conv kernel size=1	I2
		BN RELU	
		conv kernel size=5 padding=2 groups=out_ch / 2	
F1	Inception_Path3	conv kernel size=1	I3
		BN RELU	
		conv kernel size=3 padding=1 groups=out_ch / 2	
		BN RELU	
		conv kernel size=1	
F1	Inception_Path4	MaxPool kernel size=3 stride=1 padding=1	I4
		conv kernel size=1	
I1, I2, I3, I4	Concate	dim=1	D1
D1	Conv1	conv kernel size=3 padding=1 groups=out_ch / 2	D2
		BN RELU	
D2	Conv2	conv kernel size=3 padding=1 groups=out_ch / 2	D3
		BN RELU	
D3	Conv3	conv kernel size=3 padding=1 groups=out_ch / 2	O1/O2/O3/O4
		BN RELU	

Table 4. Network architecture of the Att.

Input	Flow	Operation	Output
X_{enc}, X_{MC}	Concate	-	X_{fuse}
X_{fuse}	Conv	conv kernel size=1	Q
X_{fuse}	Conv	conv kernel size=1	K
X_{fuse}	Conv	conv kernel size=1	V
Q, K	BMM	matrix product Softmax	X_{Att}
V, X_{Att}	BMM	matrix product	Out

Table 5. Network architecture of the image quality predictor.

Input	Flow	Operation	Output
ResNet Layer4	Conv	conv kernel size=1 RELU	X_{enc}
O1	Conv+AvgPool	conv kernel size=1 RELU	X_{MC1}
		AvgPool kernel size=2 stride=2	
O2	Conv+AvgPool	conv kernel size=1 RELU	X_{MC2}
		AvgPool kernel size=4 stride=4	
O3	Conv+AvgPool	conv kernel size=1 RELU	X_{MC3}
		AvgPool kernel size=8 stride=8	
X_{enc},X_{MC1}	Att	–	X_{enc}
X_{enc},X_{MC2}	Att	–	X_{enc}
X_{enc},X_{MC3}	Att	–	P1
P1	Conv	conv kernel size=1 RELU	P2
P2	Conv	conv kernel size=1 RELU	P3
P3	Conv	conv kernel size=1 RELU	P4
P4	Conv	conv kernel size=1 RELU	P5
P5	Conv	conv kernel size=7	Predicted Score

3 More Details about the Variant Models

In this section, we provide more details about the variant models, including the *w/o path-1*, *w/o path-2*,*w/o TNL*, *w/o MC* and *w/o Att*, which are defined in Sec 4.5 of our manuscript. Tab. 6 lists the components of the variant models and our full model.

For the variant model *w/o path-1*, the first prior knowledge transfer path, which connects the shared encoder (ResNet-50) and the image quality predictor is removed.

For the variant model *w/o path-2*, the second prior knowledge transfer path, which connects the image restorer and the image quality predictor is removed.

For the variant model *w/o TNL*, it has the same network architecture with our full model, while the difference is that it removes the first-phase training, and the model is only trained for BIQA.

For the variant model *w/o MC*, each MC is replaced by three convolutions in the image restorer.

For the variant model *w/o Att*, each Att is replaced by one convolution in the SN.

Table 6. Components of the variant models and our full model.

	ResNet-50	path 1	Image restorer	path 2	MC	Training for image restoration	Att
w/o _path-1	✓		✓	✓	✓	✓	✓
w/o _path-2	✓	✓			✓	✓	
w/o TNL	✓	✓	✓	✓	✓		✓
w/o MC	✓	✓	✓	✓		✓	✓
w/o Att	✓	✓	✓	✓	✓	✓	
Full Model	✓	✓	✓	✓	✓	✓	✓

4 More Comparisons on Individual Distortions

In this section, we provide more results of individual distortion evaluations. The comparisons with state-of-the-art methods in terms of SROCC on LIVE [1] and CSIQ [2] datasets are shown in Tab. 7, where the highest score for each distortion type is marked in bold. As shown in Tab. 7, our model outperforms the compared models on most distortion types. For the AWGN, JP2K and GB distortion on CSIQ dataset, our model shows a slight lower performance, but it still achieves acceptable results with SROCC of 0.933, 0.947 and 0.941, respectively. However, the performance of the JPEG distortion on LIVE dataset is less satisfying. This is due to the huge difference of distortion levels between the pre-training dataset and the LIVE dataset. In the future work, we will include more distortion types and levels to train the TN to strengthen the robustness of our model against various distortions.

Table 7. Performance comparison of individual distortions on LIVE and SCIQ datasets in terms of SROCC.

Dataset	**LIVE**					**CSIQ**					
Type	FF	GB	JP2K	JPEG	WN	AWGN	JPEG	JP2K	FN	GB	CC
BRISQUE [3]	0.828	0.964	0.929	0.965	0.982	0.723	0.806	0.840	0.378	0.820	0.804
ILNIQE [4]	0.833	0.915	0.894	0.941	0.981	0.850	0.899	0.906	0.874	0.858	0.501
HOSA [5]	0.954	0.954	0.935	0.954	0.975	0.604	0.733	0.818	0.500	0.841	0.716
FRIQUEE [6]	0.884	0.937	0.919	0.947	**0.983**	0.748	0.869	0.846	0.753	0.870	0.838
BIECON [7]	0.923	0.956	0.952	**0.974**	0.980	0.902	0.942	0.954	0.884	0.946	0.523
PQR [8]	0.921	0.944	0.953	0.965	0.981	0.915	0.934	0.955	0.926	0.921	0.837
DB-CNN [9]	0.930	0.935	0.955	0.972	0.980	**0.948**	0.940	0.953	0.940	**0.947**	0.870
HyperIQA [10]	0.934	0.926	0.949	0.961	0.982	0.927	0.934	**0.960**	0.931	0.915	0.874
Ours	**0.970**	**0.974**	**0.958**	0.913	**0.983**	0.933	**0.950**	0.947	**0.954**	0.941	**0.917**

5 More Comparisons on Small Training Data

In this section, we provide more comparisons of small data training with HyperIQA[5] [10] and DB-CNN[6] [9]. The SROCC and PLCC curves with respect to the training data ratio on LIVE [1] and CSIQ [2] datasets are shown in Fig. 1 and Fig. 2, respectively. The results are consistent with those shown in our manuscript. We can observe that as the training data ratio decreases, the advantage of our model is more significant in terms of both metrics. Moreover, the scores achieved by our model show a slower decrease compared with another two models. This means that our model can provide a better BIQA performance for scenarios where the annotated data is insufficient.

[5] https://github.com/SSL92/hyperIQA

[6] https://github.com/zwx8981/DBCNN-PyTorch

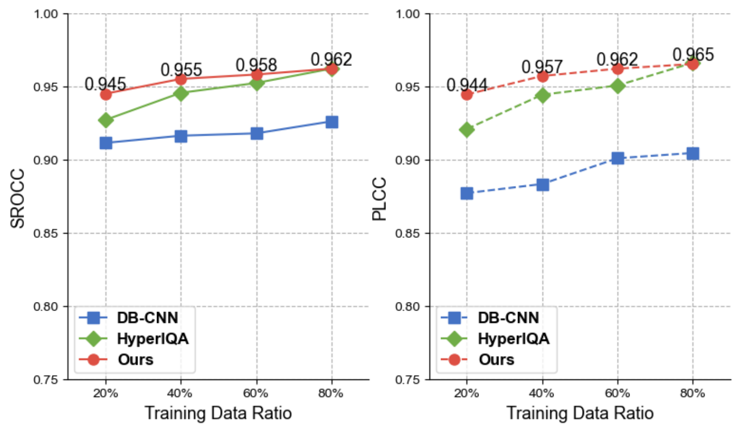

Fig. 1. Comparison of small data training on LIVE in terms of SROCC (left) and PLCC (right).

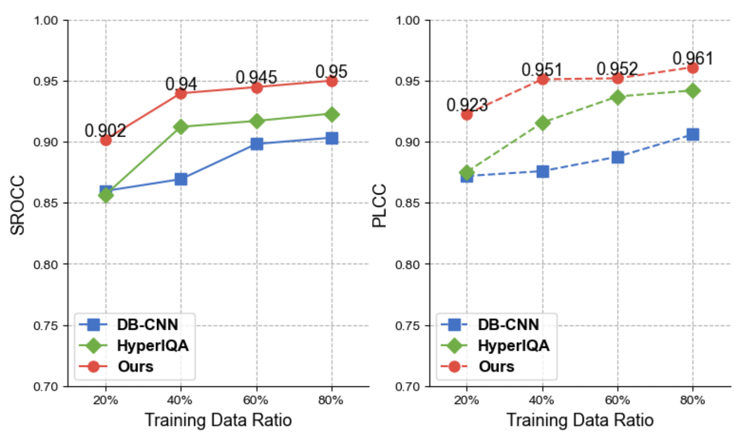

Fig. 2. Comparison of small data training on CSIQ in terms of SROCC (left) and PLCC (right).

6 More Samples on Restored Images

In this section, we provide more restored images of our TN. We select seven distortion types from the TID2013 [11], including Block-Wise (BW), Color Aberrations (CA), Non Eccentricity Pattern Noise (NEPN), Gaussian Blur (GB), Gaussian Noise (GN), Impulse Noise (IN) and JPEG compression (JEPG). Fig. 3 illustrates examples of restored images by our TN. As we can see, our model produces visually pleasing restored images for all these distortion types. The result validates the robustness of our model against various distortions.

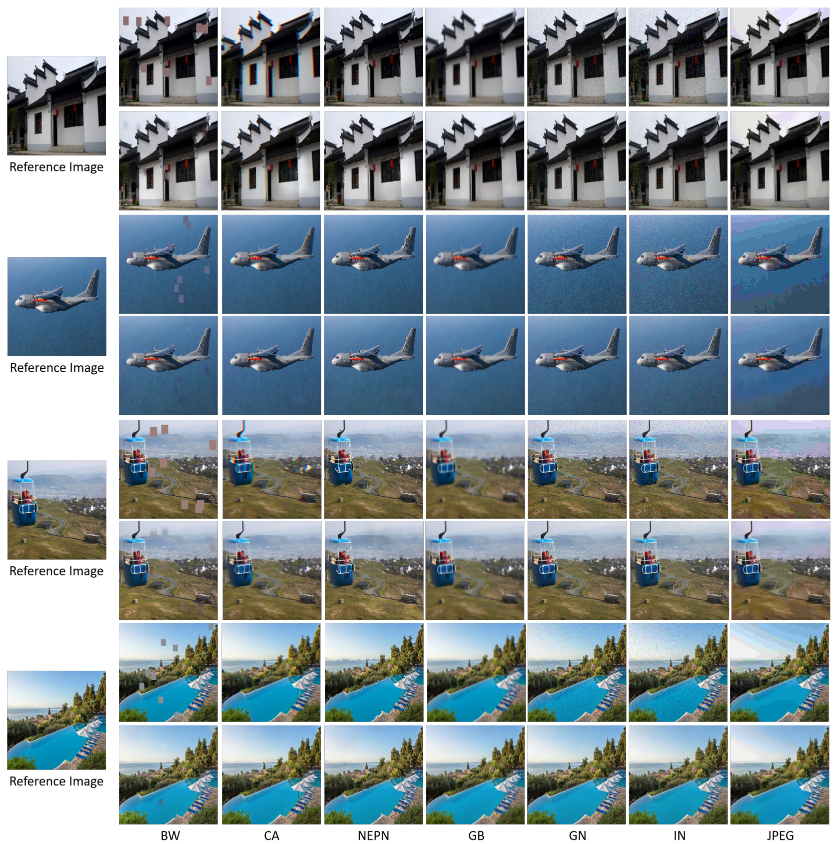

Fig. 3. Examples of restored images by the TN. For each image group, the images in the first row from left to right are corrupted by different types of distortions, and the second row shows the corresponding restored images by the TN.

Meanwhile, we provide more visual results of image restoration with the same distortion type but different levels. We choose three typical distortions - Color Quantization Dither (CQD), Multiplicative Gaussian Noise (MGN), and

Quantization Noise (QN). As shown in Fig. 4, all the restored images from different distortion levels show high quality. Although image restoration is not the objective of this paper, and we do not design complicated networks for it, we still achieve a satisfying performance on image restoration.

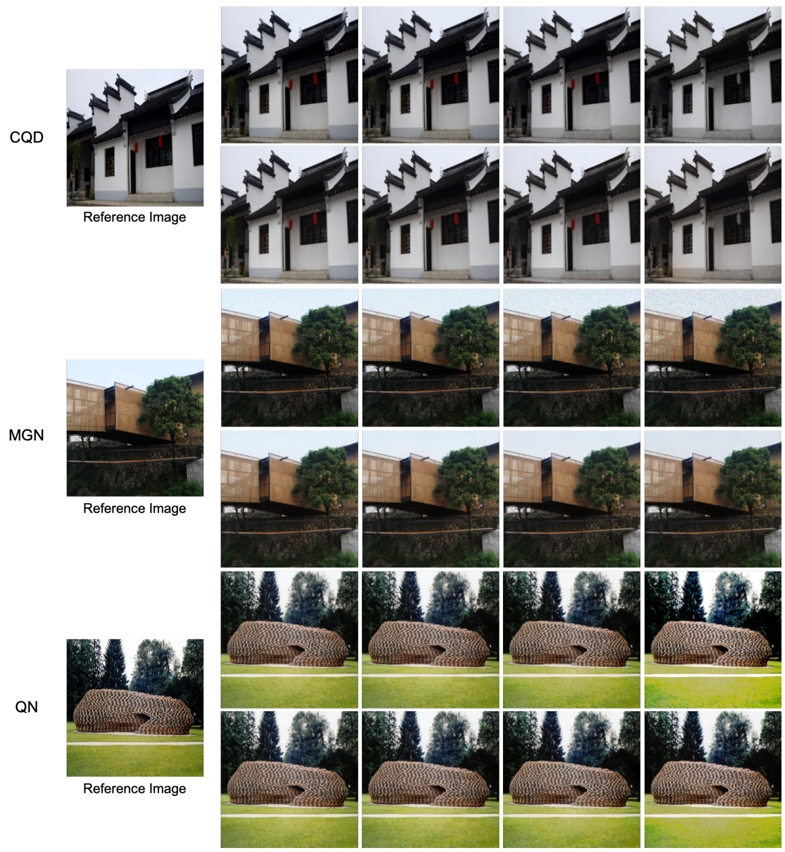

Fig. 4. Examples of restored images by TN. For each image group, the images in the first row from left to right are corrupted by the same type of distortion with increasing levels, and the second row shows the corresponding restored images by TN.

7 More Results on gMAD Competition

In this section, we provide more gMAD competition [12] results compared with tow state-of-the-art BIQA methods, hyperIQA [10] and DBCNN [9]. As shown in Fig. 5 and Fig. 6, it can show good robust performance, whether our model acts as an attacker or a defender.

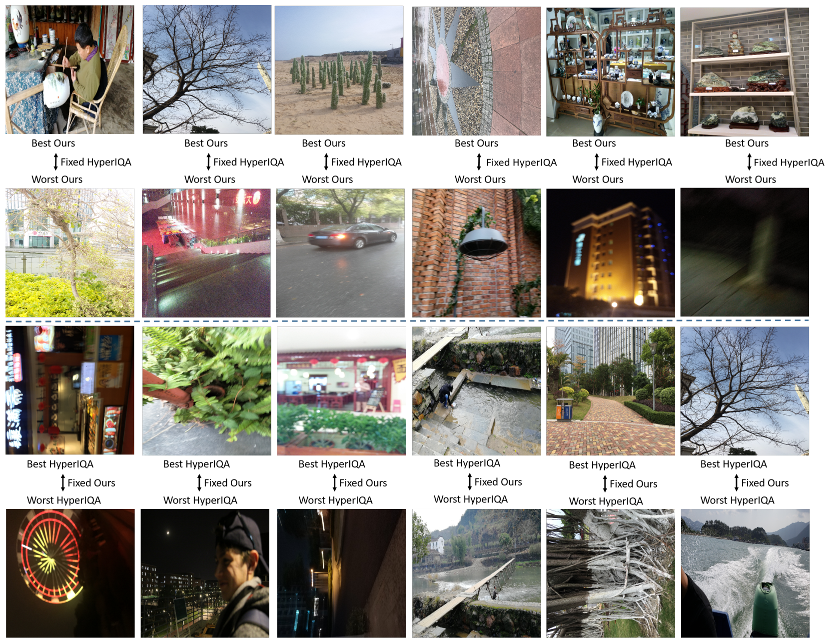

Fig. 5. The gMAD competition against hyperIQA [10] on SPAQ dataset

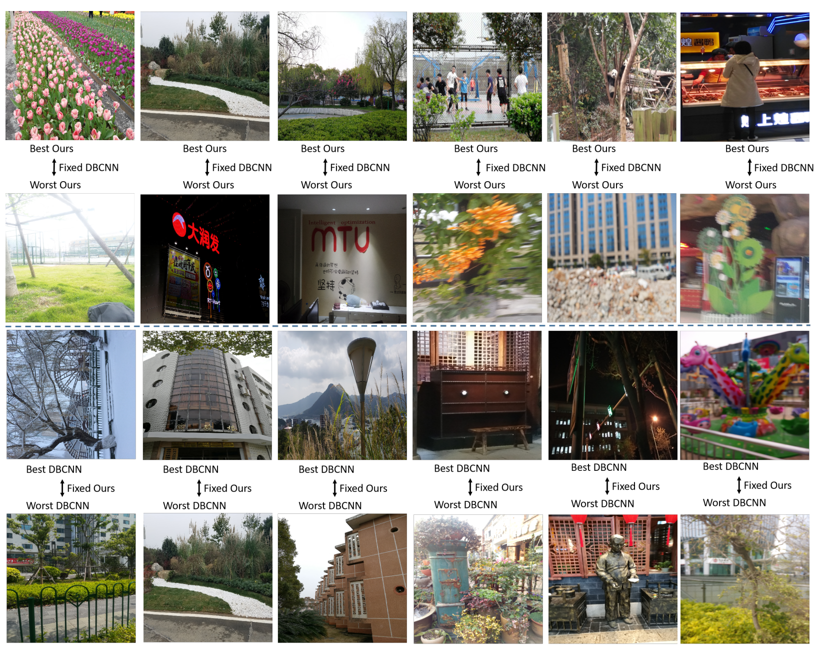

Fig. 6. The gMAD competition against DBCNN [9] on SPAQ dataset

References

1. Hamid R Sheikh, Muhammad F Sabir, and Alan C Bovik. A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Transactions on image processing*, 15(11):3440–3451, 2006.
2. Eric Cooper Larson and Damon Michael Chandler. Most apparent distortion: full-reference image quality assessment and the role of strategy. *Journal of electronic imaging*, 19(1):011006, 2010.
3. Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. No-reference image quality assessment in the spatial domain. *IEEE Transactions on image processing*, 21(12):4695–4708, 2012.
4. Lin Zhang, Lei Zhang, and Alan C Bovik. A feature-enriched completely blind image quality evaluator. *IEEE Transactions on Image Processing*, 24(8):2579–2591, 2015.
5. 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.
6. Deepti Ghadiyaram and Alan C Bovik. Perceptual quality prediction on authentically distorted images using a bag of features approach. *Journal of vision*, 17(1):32–32, 2017.
7. Jongyoo Kim and Sanghoon Lee. Fully deep blind image quality predictor. *IEEE Journal of selected topics in signal processing*, 11(1):206–220, 2016.
8. Hui Zeng, Lei Zhang, and Alan C Bovik. A probabilistic quality representation approach to deep blind image quality prediction. *arXiv preprint arXiv:1708.08190*, 2017.
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.
10. 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 (CVPR)*, June 2020.
11. Nikolay Ponomarenko, Oleg Ieremeiev, Vladimir Lukin, Karen Egiazarian, Lina Jin, Jaakko Astola, Benoit Vozel, Kacem Chehdi, Marco Carli, Federica Battisti, et al. Color image database tid2013: Peculiarities and preliminary results. In *european workshop on visual information processing (EUVIP)*, pages 106–111. IEEE, 2013.
12. Kede Ma, Qingbo Wu, Zhou Wang, Zhengfang Duanmu, Hongwei Yong, Hongliang Li, and Lei Zhang. Group mad competition-a new methodology to compare objective image quality models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1664–1673, 2016.