# Supplementary Material for Test Time Adaptation for Blind Image Quality Assessment

Subhadeep Roy* Shankhanil Mitra* Soma Biswas Rajiv Soundararajan

Indian Institute of Science, Bengaluru, India

subhadeeproy2000@gmail.com, {shankhanilm, somabiswas, rajivs}@iisc.ac.in

## A. Experiments with Transformer Parameters Adaptation

We evaluate the performance of TTA-IQA by updating transformer parameters for better adaptation. For TReS [3] and MUSIQ [5], we incorporate the transformer as a part of feature extractor. Thus, only the last fully connected (FC) layer works as the quality regressor. Current literature [6] shows that layer normalization (LN) parameters of transformers are a good choice for test time adaptation. In the case of vision transformer, the CLS token is also used for adaptation. Table 6 shows the performance on all four datasets by optimizing various parameters using the combination of rank and group contrastive loss. We observe that adaptation of transformer parameters alone gives a performance equivalent to the adaptation of the batch normalization (BN) parameters of convolutional neural network (CNN) backbone. Thus, it is possible to update models that only use a transformer and achieve significant gains using TTA.

## B. Visualizing Images that Justify Need for Both Rank and GC Loss

In Section 4.4, we justify the need for both the rank and GC loss for effective TTA. Here we give a few visual examples of images corresponding to that analysis. In Figure 7, we observe that the images have very poor quality. Hence, distorting these images further creates distorted versions that have perceptually indistinguishable quality ratings. On the other hand, Figure 8 shows similar quality images. Here, as the images have almost similar visual quality, it is difficult to form two different quality groups based on pseudo-labels given by the source model.

*Authors contributed equally to this work

## C. Performance of TTA-IQA on Other Databases

### C.1. Performance evaluation with synthetic database as source database

In the main paper, we reported performances where the source model is trained on camera captured LIVEFB [8] database and tested on various authentic and synthetic databases. In Table 7, we provide more such evaluations with respect to different intra and inter domain comparisons. In particular, we present results when TReS [3] is trained on LIVE FB and evaluated on more intra domain datasets such as SPAQ [1] and LIVEC [2]. We also present results when TReS is trained on a synthetic dataset such as LIVE-IQA [7] and tested on authentic as well as other datasets containing restored images. We observe that TTA-IQA gives a reasonable performance gain over the baseline even when there is domain shift between source (synthetic) data and target (authentic) data.

### C.2. Performance on Low-Light Restored Database

To understand the impact of larger domain shifts, we also evaluate on a new database DSLR [4], where images captured in low light are restored via various image restoration algorithms. Since novel distortions are generated while restorting such low-light images, we evaluate the performance of TTA-IQA with source database as LIVEFB and target datasbase as DSLR database. We see that TTA-IQA helps improve the performance of most of the methods.

## References

[1] Yuming Fang, Hanwei Zhu, Yan Zeng, Kede Ma, and Zhou Wang. Perceptual quality assessment of smartphone photography. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3674–3683, 2020. 1

| Backbone | Database / Method | KONIQ SRCC | KONIQ PLCC | PIPAL SRCC | PIPAL PLCC | CID2013 SRCC | CID2013 PLCC | LIVE-IQA SRCC | LIVE-IQA PLCC |
|---|---|---|---|---|---|---|---|---|---|
| TReS | Baseline | 0.6520 | 0.6955 | 0.3845 | 0.4078 | 0.5272 | 0.6463 | 0.5435 | 0.4450 |
| | BN Only | 0.6731 | 0.7151 | 0.4392 | 0.2710 | 0.6173 | 0.6800 | 0.6707 | 0.6006 |
| | LN Only | 0.6694 | 0.7176 | 0.4128 | 0.2690 | 0.6193 | 0.6850 | 0.5998 | 0.5458 |
| | BN+LN | 0.6621 | 0.7059 | 0.4417 | 0.3484 | 0.6123 | 0.6758 | 0.6723 | 0.5948 |
| MUSIQ | Baseline | 0.6304 | 0.6802 | 0.3190 | 0.3414 | 0.5173 | 0.6032 | 0.2596 | 0.3351 |
| | BN Only | 0.6588 | 0.7174 | 0.3772 | 0.3744 | 0.5275 | 0.6126 | 0.3350 | 0.3954 |
| | LN Only | 0.6582 | 0.7155 | 0.3757 | 0.3762 | 0.5403 | 0.6109 | 0.3661 | 0.4044 |
| | CLS Only | 0.6618 | 0.7207 | 0.3776 | 0.3739 | 0.5491 | 0.6212 | 0.3511 | 0.3993 |
| | BN+LN | 0.6552 | 0.7145 | 0.3736 | 0.3732 | 0.5329 | 0.6095 | 0.3767 | 0.4028 |
| | BN+LN+CLS | 0.6598 | 0.7172 | 0.3751 | 0.3764 | 0.5253 | 0.6048 | 0.3569 | 0.3999 |

Table 6: Comparison of TTA-IQA using popular transformer based NR IQA methods on authentically and synthetically distorted datasets.

| Train on | LIVEFB | | | | LIVE-IQA | | | |
|---|---|---|---|---|---|---|---|---|
| Test on | PIPAL | KonIQ-10k | SPAQ | LIVEC | PIPAL | CID2013 | KonIQ-10k | LIVEC |
| Baseline | 0.385 | 0.652 | 0.707 | 0.726 | 0.402 | 0.519 | 0.521 | 0.563 |
| TTA-IQA | **0.428** | **0.658** | **0.755** | **0.728** | **0.449** | **0.523** | **0.522** | **0.565** |

Table 7: SRCC performance evaluation of TTA-IQA with TReS backbone trained on LIVE-IQA database

| | TRES | MUSIQ | HYPER-IQA | META-IQA |
|---|---|---|---|---|
| Baseline | 0.535 | 0.404 | 0.496 | 0.591 |
| Rotation | 0.529 | 0.425 | 0.479 | 0.540 |
| TTA-IQA | 0.586 | 0.450 | 0.493 | 0.608 |

Table 8: SRCC performance analysis of TTA-IQA on DSLR database.


[2] Deepti Ghadiyaram and Alan C. Bovik. Massive online crowdsourced study of subjective and objective picture quality. *IEEE Transactions on Image Processing*, 25(1):372–387, 2016. 1

[3] S Alireza Golestaneh, Saba Dadsetan, and Kris M Kitani. No-reference image quality assessment via transformers, relative ranking, and self-consistency. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 3209–3218, 2022. 1

[4] Vignesh Kannan, Sameer Malik, Nithin C. Babu, and Rajiv Soundararajan. Quality assessment of low-light restored images: A subjective study and an unsupervised model. *IEEE Access*, 11:68216–68230, 2023. 1

[5] Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang. Musiq: Multi-scale image quality transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 5148–5157, October 2021. 1

[6] Takeshi Kojima, Yutaka Matsuo, and Yusuke Iwasawa. Robustifying vision transformer without retraining from scratch by test-time class-conditional feature alignment. In Lud De Raedt, editor, *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*, pages 1009–1016. International Joint Conferences on Artificial Intelligence Organization, 7 2022. Main Track. 1

[7] H.R. Sheikh, M.F. Sabir, and A.C. Bovik. A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Transactions on Image Processing*, 15(11):3440–3451, 2006. 1

[8] Zhenqiang Ying, Haoran Niu, Praful Gupta, Dhruv Mahajan, Deepti Ghadiyaram, and Alan Bovik. From patches to pictures (paq-2-piq): Mapping the perceptual space of picture quality. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. 1

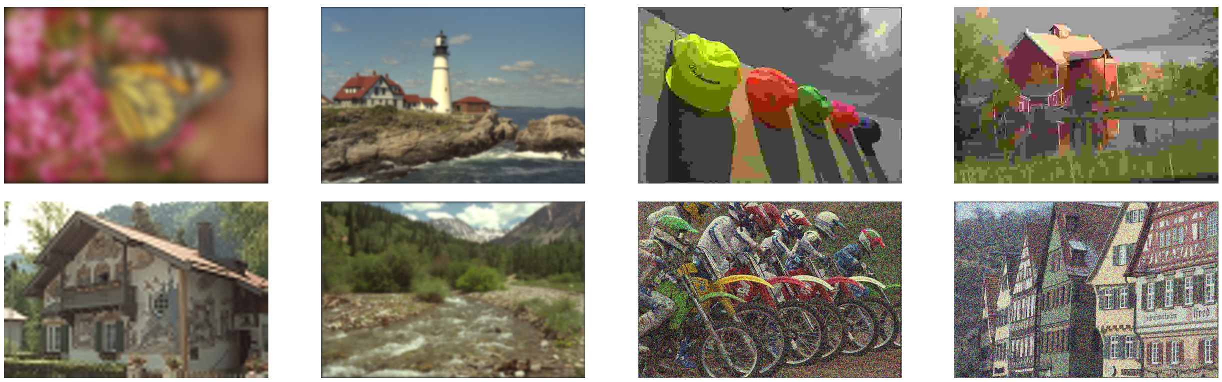

Figure 7: Examples of highly distorted images in which GC loss is more effective than rank loss

Figure 8: Examples of similar quality images in which rank loss is more effective than GC loss