Supplemental Material for Re-IQA: Unsupervised Learning for Image Quality Assessment in the Wild

1. Distortion Bank

We used a total of 25 quality distortion augmentations, with five levels for each type of distortion, for training our quality-aware sub-module. The visual differences among the various levels and types of distortions can be seen in Figure 1, 2, 3, and 4. The details of each distortion type are expanded below.

- `Resize Bicubic`: Downsize the image and upsize it back to the original size using bicubic interpolation.
- `Resize Bilinear`: Downsize the image and upsize it back to the original size using bilinear interpolation.
- `Resize Lanczos`: Downsize the image and upsize it back to the original size using Lanczos filter-based interpolation.
- `Pixelate`: Downsize the image and upsize it back to the original size using nearest-neighbor interpolation.
- `Motion Blur`: Emulates motion blur by filtering using a line kernel.
- `Gaussian Blur`: Filters the image with a Gaussian kernel.
- `Lens Blur`: Filters the image with a circular kernel.
- `Mean Shift`: Shifts the mean intensity of the image by adding a constant value to all pixel values and truncating to the original value range.
- `Contrast`: Changes the contrast of the image by applying a non-linear Sigmoid-type adjustment curve on the RGB values.
- `Unsharp Masking`: Increases the sharpness of an image by using unsharp masking.
- `Jitter`: Randomly scatters image data by warping each pixel with small random offsets.
- `Color Block`: Inserts homogenous randomly colored blocks at random locations in the image.
- `Non-eccentricity`: Randomly offsets small patches in the image by small displacements.
- `JPEG Compression`: Applies standard JPEG compression.
- `White Noise (RGB space)`: Adds Gaussian White noise in the RGB space.
- `White Noise (YCbCr space)`: Adds Gaussian white noise in the YCbCr space.
- `Impulse Noise`: Adds salt and pepper noise in the RGB space.
- `Multiplicative Noise`: Adds speckle noise in the RGB space.
- `Denoise`: Adds Gaussian white noise to the RGB image and then applies a randomly chosen blur filter (Gaussian or box blur) to remove noise.
- `Brighten`: Increases the brightness of the image by applying a non-linear curve fitting to avoid changing extreme values.
- `Darken`: Similar to `Brighten`, but decreases pixel values.
- `Color Diffuse`: Applies Gaussian blur on the color channels (a and b) in the LAB-color space.
- `Color Shift`: Randomly translates the green channel and blends it into the original image masked by a gray level map which is the normalized gradient magnitude of the original image.
- `Color Saturate`: Multiplies the saturation channel in the HSV-color space by a factor.
- `Saturate`: Multiplies the color channels in the LAB-color space by a factor.

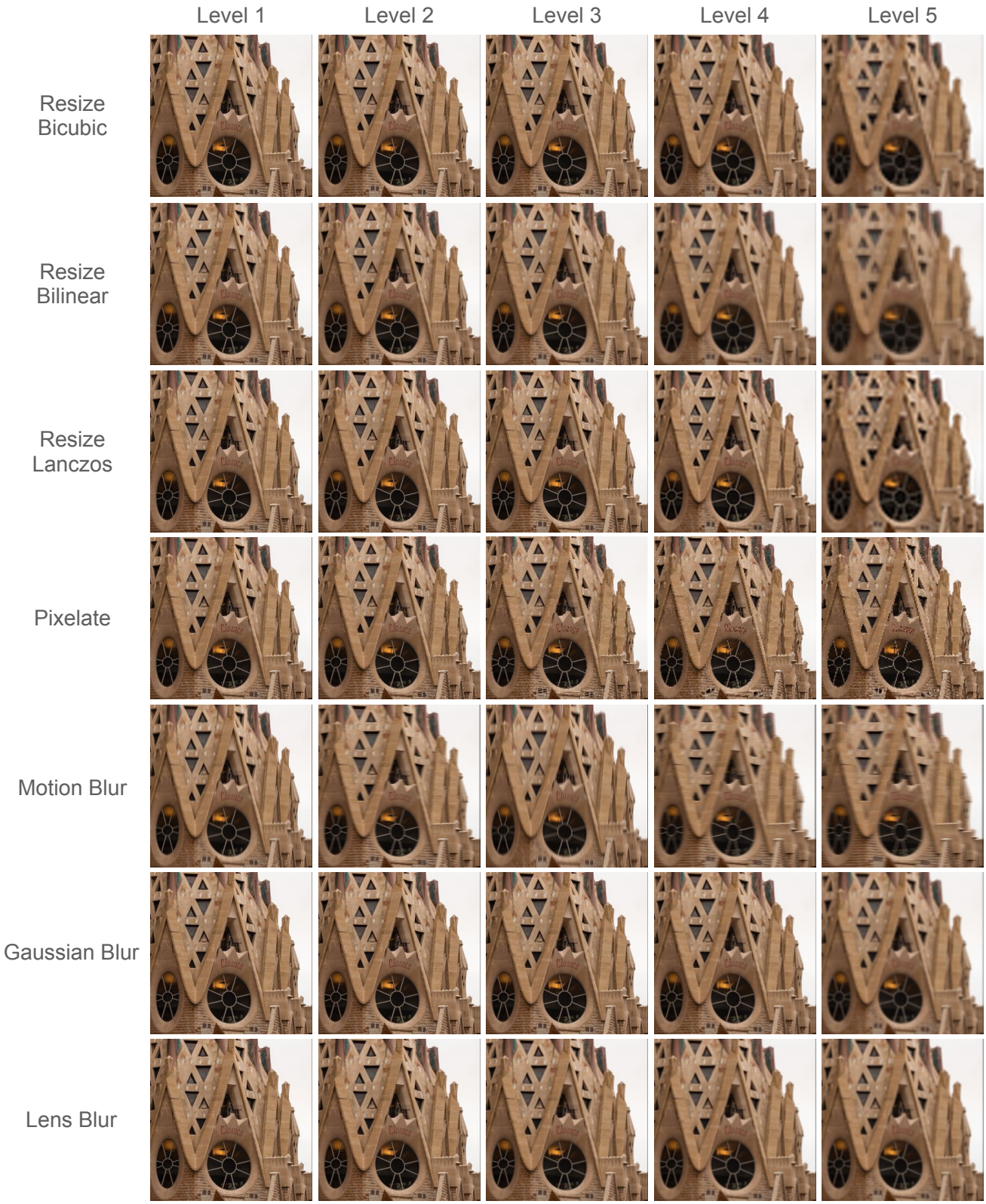

Figure 1. Distortion Bank. Best viewed zoomed.

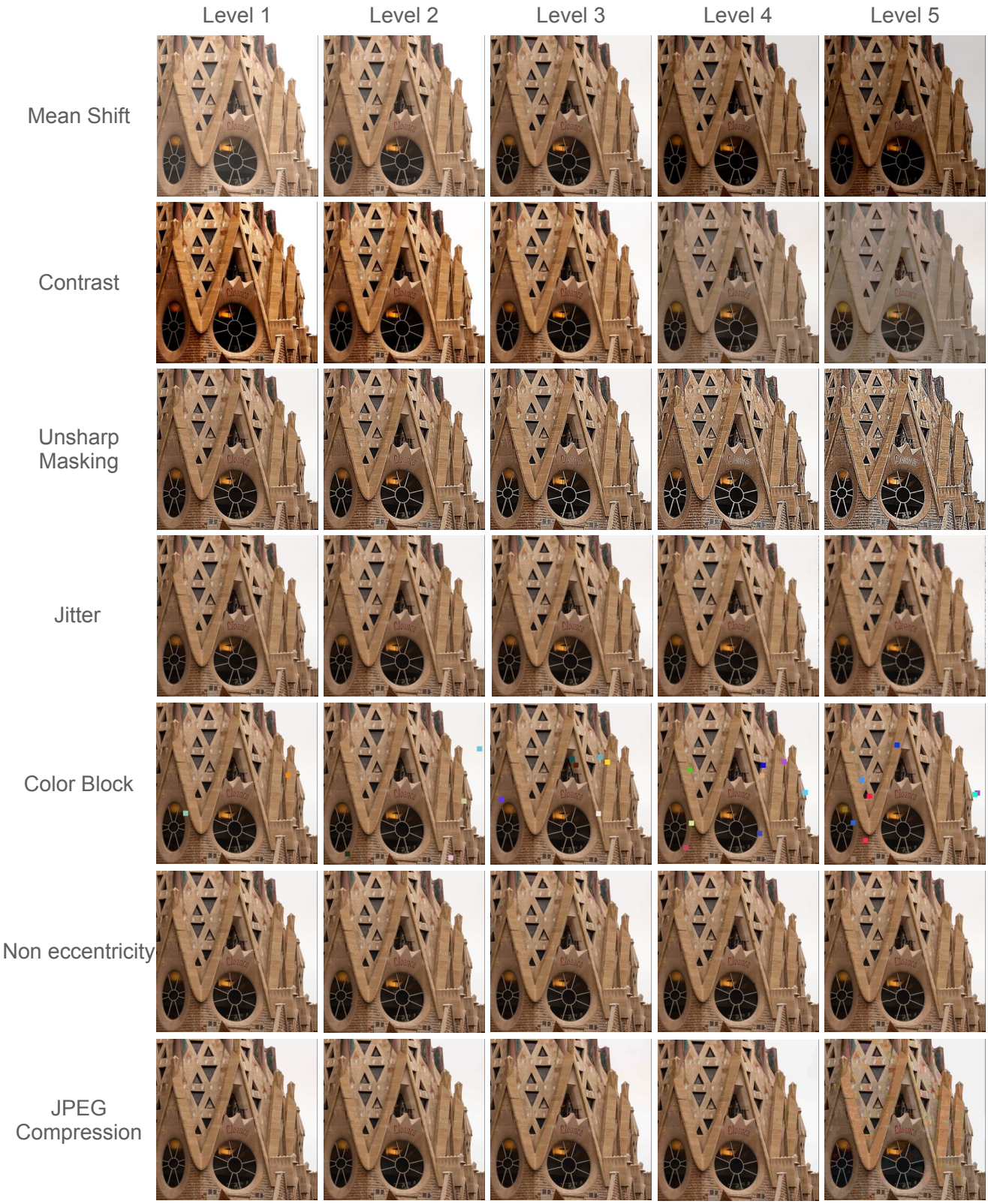

Figure 2. Distortion Bank - Continued. Best viewed zoomed.

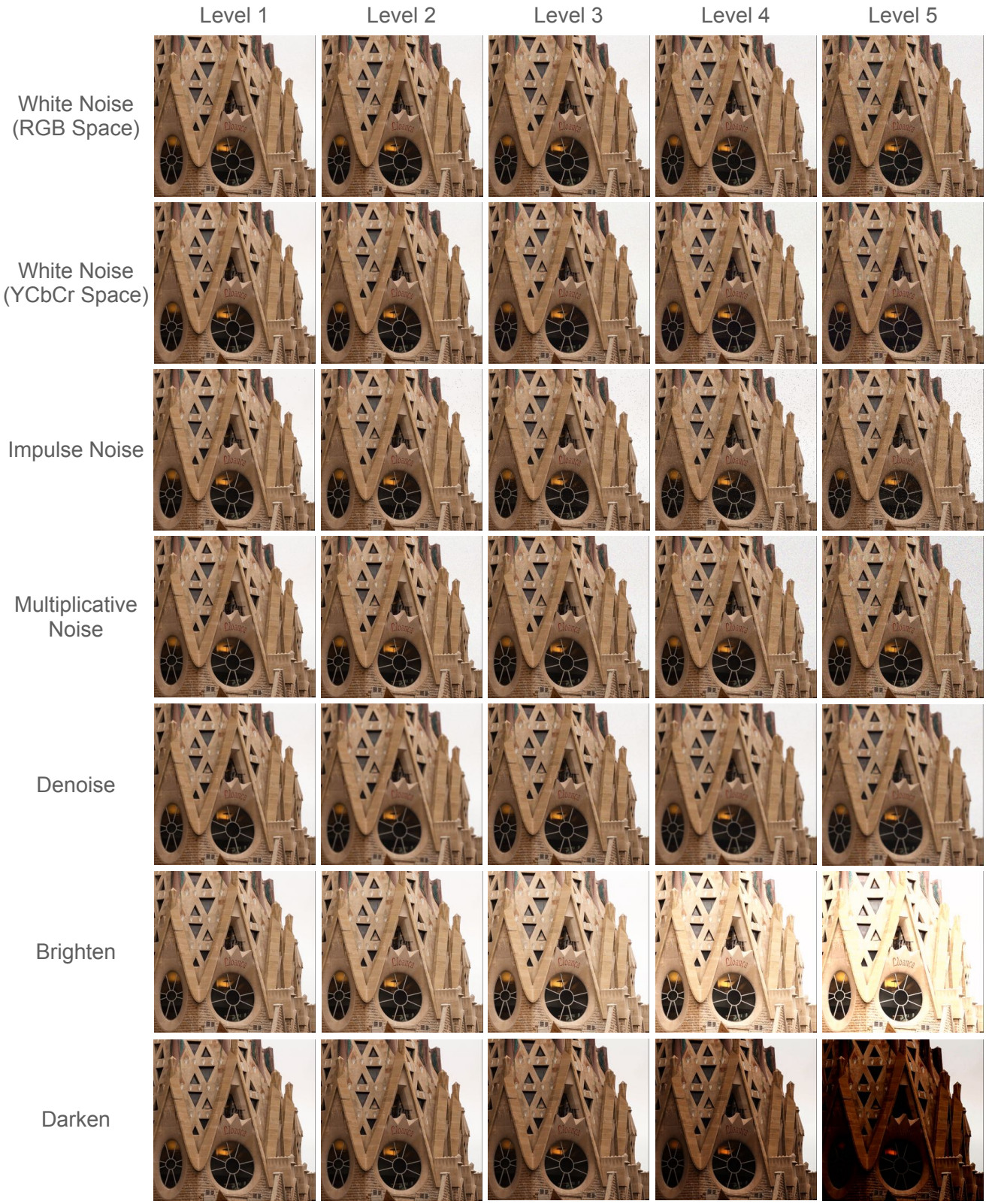

Figure 3. Distortion Bank - Continued. Best viewed zoomed.

Figure 4. Distortion Bank - Continued. Best viewed zoomed.

2. Additional Experiments

In this section, we report results from additional experiments performed using our proposed Re-IQA framework. Section 2.1 illustrates the robust performance of our Re-IQA model when trained and tested on different databases. In Section 2.2, we extend the Re-IQA framework to the Full Reference IQA setting and report performance results for databases with synthetic distortions.

2.1. Cross-Database Performance

We conducted cross-dataset evaluations to demonstrate the robustness of the representation learned in our Re-IQA framework, where training and testing were performed on different datasets. We follow the cross-database evaluation protocol, used by authors of CONTRIQUE. We select four image quality databases CLIVE, KonIQ, LIVE-IQA and CSIQ-IQA, covering both synthetic and authentic distortions and three state-of-the-art NR-IQA algorithms PQR, HyperIQA, and CONTRIQUE, apart from our proposed Re-IQA. We use Spearman's rank order correlation coefficient (SRCC) to evaluate the cross-database performance of all the four models. Results from Table 1 indicate that Re-IQA attains comparable performance to the NR-IQA models across both synthetic and authentic distortions.

Training Database	Testing Database	NR-IQA Algorithms			
		PQR	Hyper-IQA	CONTRIQUE	Re-IQA
CLIVE	KonIQ	0.757	**0.772**	0.676	**0.769**
KonIQ	CLIVE	0.770	**0.785**	0.731	**0.791**
LIVE-IQA	CSIQ-IQA	0.719	0.744	**0.823**	**0.808**
CSIQ-IQA	LIVE-IQA	0.922	**0.926**	0.925	**0.929**

Table 1. Cross Database SRCC Comparison of NR-IQA models. Top 2 Performing Models in each row is in bold. Higher SRCC indicates better performance.

2.2. Full Reference Image Quality Assessment

Re-IQA learns image representations in a completely unsupervised setting, thus making it flexible for application in Full Reference Image Quality Assessment (FR-IQA) tasks. Since reference image scores are available in the case of FR-IQA tasks, the score regressor is now trained to predict

Method	Synthetic Distortions							
	LIVE-IQA		CSIQ-IQA		TID-2013		KADID	
	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC	PLCC
PSNR	0.881	0.868	0.820	0.824	0.643	0.675	0.677	0.680
SSIM	0.921	0.911	0.854	0.835	0.642	0.698	0.641	0.633
FSIM	0.964	0.954	0.934	0.919	0.852	0.875	0.854	0.850
LPIPS	0.932	0.936	0.884	0.906	0.673	0.756	0.721	0.713
DRF-IQA	**0.983**	**0.983**	**0.964**	0.960	**0.944**	**0.942**	-	-
CONTRIQUE-FR	0.966	0.966	0.956	**0.964**	0.909	0.915	**0.946**	**0.947**
Re-IQA-FR (content-aware)	0.932	0.947	0.927	0.928	0.877	0.884	0.889	0.898
Re-IQA-FR (quality-aware)	**0.973**	**0.974**	**0.961**	**0.965**	0.905	0.915	0.901	0.903
Re-IQA-FR	0.969	**0.974**	**0.961**	0.962	**0.920**	**0.921**	**0.933**	**0.936**

Table 2. Performance comparison of Re-IQA-FR against various FR-IQA models on IQA databases with synthetic distortions. The top 2 best performing models are in bold. Higher SRCC and PLCC scores imply better performance.

the difference in the scores of the reference image and the distorted image, as described in equation 1 below

$$y_{dmos} = W|h_{ref} - h_{dist}| \tag{1}$$

where y_{dmos} is the differential mean opinion score of the test image, W represents the regressor weights, and h_{ref} and h_{dist} are the image representations generated by the Re-IQA sub-modules for the reference and distorted images respectively. We refer to this modified framework as Re-IQA-FR. Note that only the score regressor is modified and trained for FR-IQA, and the weights of the pre-trained sub-modules of Re-IQA remain unchanged.

The evaluation strategy for FR-IQA remains exactly the same as that of NR-IQA, wherein we split the dataset into 70-10-20 for train-val-test sets and report the median performance scores. We report the performance of Re-IQA-FR on various full reference IQA datasets (refer Table 2), otherwise referred to as synthetic IQA datasets throughout our paper. We compare our model against six popular FR-IQA models available in the literature: PSNR, SSIM, FSIM, LPIPS, DRF-IQA [1] and CONTRIQUE-FR, where the first three are traditional models and the last three are deep learning based models. We also report the scores of the individual sub-modules of Re-IQA-FR.

Results from Table 2 highlight that our proposed Re-IQA-FR model achieves state-of-art performance across all evaluated databases. Another interesting observation is that the SRCC and PLCC scores of the Re-IQA-FR are better than the No-Reference version of Re-IQA across all databases, as reported in Table 2 of the main paper. This can be attributed to the knowledge of extra information available to the Re-IQA-FR framework in the form of the undistorted reference image. We also provide a cross-database comparison between CONTRIQUE-FR and Re-IQA-FR in Table 3. The results suggest that RE-IQA-FR and CONTRIQUE-FR exhibit similar performance, with RE-IQA-FR performing slightly better.

Table 3. Cross Database SRCC Comparison of CONTRIQUE and Re-IQA-FR

Training Dataset	Testing Dataset	CONTRIQUE-FR	Re-IQA-FR
LIVE-IQA	CSIQ-IQA	0.855	0.856
CSIQ-IQA	LIVE-IQA	0.937	0.942

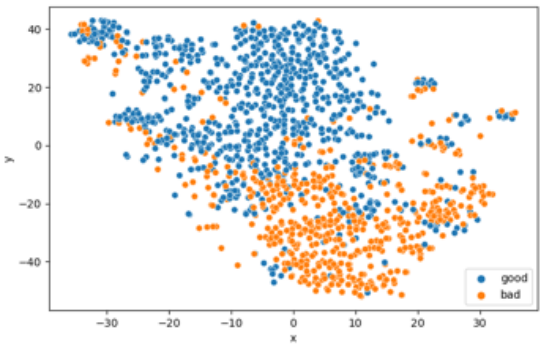

Figure 5. t-SNE visualization of learned quality-aware representations for 1338 'Images in the Wild' from KonIQ.

3. Visualization

We utilized the TSNE() function from the sklearn.manifold library to conduct t-SNE experiments and present the findings in Figure 5 (Main Paper). The function's default parameters were used, except for $perplexity$ and n_iter. While we selected $perplexity = 50$ and $n\_iter = 3000$ to achieve optimal visual outcomes, our results remain reliable even when $perplexity = 10$ and $n\_iter = 1000$ were used.

In Figure 5, we demonstrate the efficacy of the representations generated by the quality-aware sub-module of Re-IQA in segregating high-quality images from low-quality images. We use 2D t-SNE plots to visualize the high-dimensional learned feature embeddings. The orange

points represent lower-quality images (scored less than 50 on a scale of 0-100), while the blue points represent higher-quality images (scored more than 50 on a scale of 0-100). We chose 1338 images from the KonIQ dataset, such that there is an equal representation of low and high-quality images available for the visualization. As can be seen in the figure, the orange points cluster toward one corner, whereas the blue points cluster toward the opposite corner. The overlapping region of the clusters represents those images for which the quality-aware model's learned representations are likely to produce quality scores that do not agree with the ground truth.

4. Discussion on Hypotheses

Our unsupervised learning framework to learn quality-aware features requires us to define a decision boundary between 'similar-quality' and 'different-quality' labels. As discussed in Section 3.2 (Main Paper), we set up our learning framework by assuming the following three hypotheses to be true:

Hyothesis 1: Image quality can vary in different locations based on the image's content. Thus, PQAF varies within an image itself. If we assign PQAF to an image patch x and denote it as $PQAF_x$, then $PQAF_x$ varies only a small amount between neighboring patches. However, $PQAF_x$ may vary significantly between two distant patches depending on the content. Our approach generates 'similar-quality' samples by extracting overlapping patches from an image in the training database. We conducted an ablation study to assess the impact of varying the degree of overlap between the two patches sampled from an image on the performance of the Re-IQA model. Detailed results of this study can be found in Table 1 of the Main Paper. We do not explicitly sample from distant patches in an image.

Hypothesis 2: This hypothesis serves as a primary basis in the design of the Re-IQA-QA module. The PQAF of any two randomly selected images are 'different,' which assumes that the scenes depicted in the images differ. However, this does not enforce any restrictions on the quality scores of the two images. Without this hypothesis, the contrastive loss cannot be computed between features obtained from crops of 2 different source images, and the network will not learn to compare two images from different sources during training.

Hypothesis 3: Two different distorted versions of the same image have different PQAF. This hypothesis is like an axiom. Its absence results in 2 images with the same content but different distortions generating the same PQAF, which would be technically wrong because it would imply two images with different distortions have the same perceptual quality. This is important because it ensures that the PQAF accurately reflects the perceptual quality of an image, even when comparing images with the same content but different distortions.

5. Acknowledgement

The authors acknowledge the Texas Advanced Computing Center (TACC), at the University of Texas at Austin for providing HPC, visualization, database, and grid resources that have contributed to the research results reported in this paper. URL: http://www.tacc.utexas.edu

References

[1] Woojae Kim, Anh-Duc Nguyen, Sanghoon Lee, and Alan Conrad Bovik. Dynamic receptive field generation for full-reference image quality assessment. *Trans. Img. Proc.*, 29:4219–4231, jan 2020. 6