Image Intrinsic Scale Assessment: Bridging the Gap Between Quality and Resolution

Supplementary Material

Vlad Hosu[1,*] Lorenzo Agnolucci[2,†,*] Daisuke Iso[1] Dietmar Saupe[3]

[1] Sony AI - [name.surname]@sony.com
[2] University of Florence, Italy - [name.surname]@unifi.it
[3] University of Konstanz, Germany - [name.surname]@uni-konstanz.de

S1. IISA Task: Additional Details

S1.1. Sensitivity of the IIS

In this section, we compare the relative sensitivity of the quality ratings and the IIS. Here, by *sensitivity* we refer to the precision of an annotation tool in detecting variations in image quality. The classical measurement tool in IQA is the rating scale, such as the 100-point "continuous" scale or the discrete 5-point Absolute Category Ratings (ACR). In IISA we employ a 100-point scale corresponding to the rescaling values to measure the IIS. Here, we normalize both quality ratings and scale values to the interval $[0, 1]$.

To conduct our analysis, we consider the KonX dataset [11]. We aim to study the connection between the change in quality relative to the change in scale. Recall that KonX provides quality scores, in the form of a Mean Opinion Score (MOS), for the same images annotated at three resolutions: 512×384, 1024×768, and 2048×1536. We compute the quality differences between pairs of corresponding rescaled images across the three resolutions and plot them against the MOS of the higher-resolution image in each pair, as shown in Fig. S1.

We restrict our analysis to image scale pairs where the MOS at the higher resolution is below 0.85. For these, the average quality increases with downscaling. The ratio of the resolutions in a pair gives the downscaling factor, which in turn gives the variation in scale, referred to as ΔS. Thus, halving the resolution means $\Delta S = 0.5$ and on average corresponds to a quality increase (referred to as ΔQ) of 0.038, while downscaling to $S = 0.25$ means $\Delta S = 0.75$ and leads to $\Delta Q = 0.076$. To measure the sensitivity, we employ the concept of leverage γ, defined as the ratio between the change in image scale and the change in image quality, *i.e.*, $\gamma = |\Delta S|/|\Delta Q|$. Hence, $\gamma > 1$ indicates that a large change in the scale results in a smaller change in quality, and vice-versa for $\gamma < 1$. In our case, the leverage γ is 6.6 and 19.7 for the two scale changes of $\Delta S = 0.5$ and $\Delta S = 0.75$, respectively.

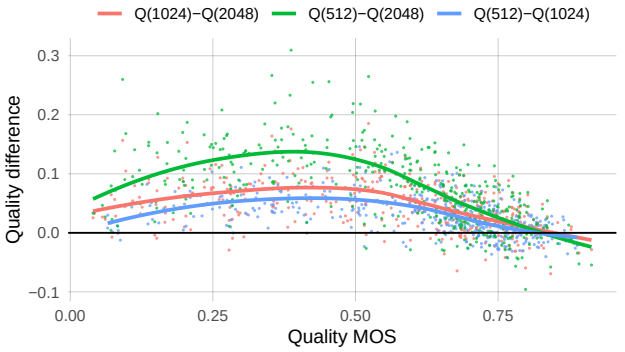


Figure S1. Quality differences between pairs of rescaled images belonging to the KonX dataset against the MOS of the higher-resolution image in each pair. The MOS for an image at a specific resolution is denoted as Q(image-width). We plot trend lines for each resolution pair.

In Sec. 4.2, we discuss the precision of subjective measurements, indicated by the confidence intervals of the aggregated annotations. The precision levels for IISA and NR-IQA are comparable. However, the leverage factor $\gamma > 1$ implies that minor changes in quality result in larger variations in scale. Therefore, achieving a specific precision in measuring scale equates to an even finer precision in measuring quality. Thus, γ acts as a leverage or magnifying factor. This indicates that the sensitivity of our annotation tool for IIS is higher when detecting subtle differences in quality compared to traditional quality rating scales, making IISA suitable for fine-grained quality assessment.

[*] Equal contribution.
[†] Work done during an internship at Sony AI.

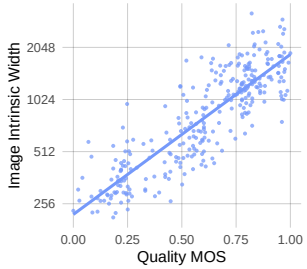


Figure S2. Relationship between image quality and intrinsic width (logarithmic scale) for the KonIQ-10k subset of IISA-DB.

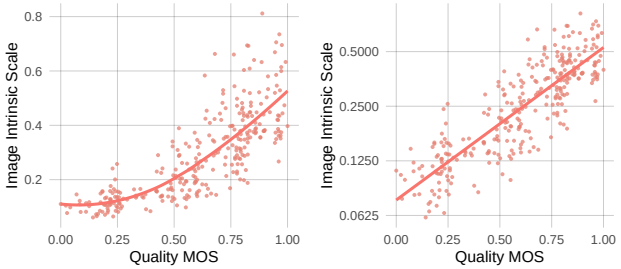


Figure S3. Relationship between image quality and intrinsic scale for the KonIQ-10k subset of IISA-DB, with IIS represented on a linear scale (left) and a logarithmic scale (right).

S1.2. IIS and Quality Scores

We investigate the relationship between the IIS and quality scores, represented by MOS, by analyzing the overlapping images between the KonIQ-10k [6] and IISA-DB datasets. KonIQ-10k comprises quality annotations for images downscaled to a fixed resolution of 1024×768 pixels. In contrast, the KonIQ-10k subset of IISA-DB contains the original high-resolution version of the same images, with the same content and aspect ratio but variable resolutions above 2048×1536 pixels.

By the definition of IIS, high-quality – and thus presumably undegraded – images should have an IIS of 1, as downscaling them can not reduce the visible degradation but merely results in a potential loss of details. Therefore, KonIQ-10k images with near-perfect quality (*i.e.*, with the highest MOS) are expected to correspond to an IIS of 1 at a resolution of 1024×768 pixels. From another perspective, the original high-resolution versions of the images with the highest MOS should have an intrinsic width (*i.e.*, the width of the image downscaled to its IIS) of *at least* 1024 pixels. Indeed, such images could reach a near-perfect quality even when downscaled to a width larger than 1024 pixels. To validate this hypothesis, we plot the intrinsic widths of the images against their MOS in Fig. S2. The results show that the images with the highest quality correspond to intrinsic widths higher than 1024 in almost all cases, thus confirming our hypothesis.

In addition, we plot the IIS of the images against their MOS in Fig. S3. We observe a non-linear relationship between the IIS and quality MOS (Fig. S3, left). On the contrary, the logarithm of the IIS exhibits an approximately linear relationship with the quality scores (Fig. S3, right). This result emphasizes the different nature of the scales of the MOS and the IIS. Indeed, the quality ratings use a perceptually linear scale [2], while rescaling factors – underlying the IIS – are intrinsically non-linear.

S1.3. Discussion on Assumption

In Sec. 6.3 we discuss our assumption that the relationship between image quality and scale follows either a concave-down or monotonic function. Specifically, prior works related to viewing distance [3, 5, 10] align with our assumption. In addition, we empirically test this assumption by analyzing the quality change with resolution in the KonX dataset [11]. For each resolution, the estimated quality MOS have an average confidence interval of approximately 4.6% relative to the rating range, demonstrating good precision and enabling us to draw the following conclusions. First, we observe that for 90% of the KonX images (378 out of 420) the MOS across the three resolutions supports our assumption. Second, for cases where the MOS follows a concave-up function across resolutions, due to the uncertainty of the MOS there is no single image for which we can assert with more than 90% probability that our assumption does not hold. To determine this, we generate MOS values by resampling the individual quality ratings with replacement 100 times from the original pool of per-participant ratings. The fraction of samples that do not support our assumption provides the stated probability. We hypothesize that instances where our assumption appears not to hold may be due to subjective biases from annotators – such as the presentation order of images during annotation, context effects like anchoring caused by the distribution of image quality within the same session, or individual interpretations of the quality scale – as well as the presence of interpolation artifacts from downscaling, including aliasing, moiré patterns, or blurs.

S1.4. IISA Applications: Additional Details

We present scenarios across industry and research where IISA is the perfect tool to optimize quality and resolution trade-offs.

Printing and publishing Printing an image too large can accentuate flaws present in the source image while printing too small sacrifices detail. IISA can guide the selection of print dimensions and resolution (dot-per-inch, DPI). This ensures consistently high-quality prints. Traditionally, this task is managed by an expert operator. IISA automates these decisions, allowing scalable deployment in online printing systems and enabling non-expert users to make optimal choices independently.

Moreover, web developers and UI designers often need

to serve images across devices with different screen sizes and resolutions. Typically, responsive design uses fixed rules, whereas IISA enables content-awareness. For instance, an online image gallery can automatically size each photo based on its intrinsic scale. This ensures that users see images at the best quality for their device while saving bandwidth and load time.

Gaming and graphics rendering Modern games employ dynamic resolution scaling to maintain high FPS. However, choosing the amount of rescaling on each axis can dramatically affect quality. IISA offers a principled solution to this problem. Moreover, if the target FPS is not fixed, a game engine could automatically downscale rendered frames (slightly degraded by aliasing or motion blur) until just before quality starts dropping, ensuring players get the clearest visuals with optimal performance (FPS).

Computational photography Smartphone cameras rely on computational photography to balance resolution and noise. Although sensors may reach 100+MP, phones often merge pixels in low light to produce cleaner lower MP images – effectively searching for the image's intrinsic scale. IISA makes this process explicit and optimal. In tasks like super-resolution or denoising, algorithms can use IISA to determine when further resolution or noise reduction stops improving quality.

Benchmarking IQA methods Traditional IQA metrics predict quality at a fixed resolution, while IISA requires accuracy at multiple scales. Our experiments show that off-the-shelf IQA models – NR-IQA trained on traditional datasets – perform poorly on the IISA task. By directly evaluating alignment with human perception across scales, IISA serves as a valuable benchmark. Performance improvements of NR-IQA methods on the IISA-DB benchmark indicate a deeper understanding of the quality–resolution trade-off, which is critical for both academic research and real-world image processing.

Extending IQA study methodology IISA introduces new methods for subjective image quality evaluation. Traditional IQA often has viewers rate an image's quality at a fixed resolution, which can be difficult for subtle degradations in no-reference settings. By contrast, IISA asks viewers to resize an image until it "looks best", inherently comparing quality across scales. This approach improves sensitivity to minor artifacts (see Sec. S1.1).

Training dataset curation for image restoration Constructing image datasets requires managing images of varying quality. Traditionally, low-quality images are discarded to avoid introducing erroneous priors into restoration models. However, IISA provides a more nuanced approach: rather than discarding low-quality images, it downscales them to their intrinsic scale to maximize quality. This preserves content diversity because discarding images purely on quality can disproportionately exclude dynamic or low-light scenes, which tend to be blurrier and noisier. By applying IISA, we mitigate that bias while maintaining a wider range of content.

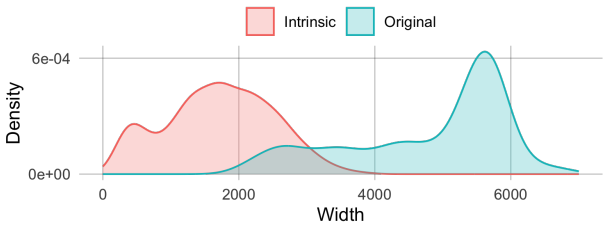


Figure S4. Distribution of the width of the images of the IISA-DB dataset at their original and intrinsic scale.

S2. IISA-DB Dataset: Additional Details

S2.1. Image Curation

To ensure the diversity of images in the proposed dataset, we selected images from two sources: 300 from the KonIQ-10k dataset [6] – which were themselves sourced from Flickr – and 600 from Pixabay. The two sets were chosen to balance the range of intrinsic resolutions (corresponding to the image rescaled to its IIS) in the database. The KonIQ-10k subset comprises lower-quality photos with smaller intrinsic resolutions, whereas the Pixabay images are typically of higher quality and intrinsic resolution.

Aiming for higher quality in the Pixabay subset, we selected newer camera models from a list of 71, focusing on those released after 2010 with full-frame sensors. From a pool of over one million images on Pixabay with EXIF information, we filtered for photos with a width greater than 4,000 pixels, resulting in approximately 18,000 images that met all criteria. We sampled for diversity 600 images, maintaining uniform distributions among binned normalized favorites, likes, downloads, and user tags using a method similar to [11]. Most of these were captured with *Canon EOS 1/5/6D Mark 2/3/4* cameras, using various lenses and capture settings. For the KonIQ-10k subset, we also sampled for diversity regarding quality levels and machine tags – with a confidence greater than 80% – using the same stratified procedure.

The last filtering step we applied was removing images containing identifiable people. Thus, we retained 248 images from KonIQ-10k and 537 from Pixabay. All images have a minimum width of 2048 pixels and are annotated at their original resolution. Fig. S4 illustrates the distribution of image widths at their original and intrinsic scales.

S2.2. Annotation Approach: Slider

As explained in Sec. 4, the subjective side of IISA is similar to the image Just Noticeable Difference (JND) task [9], which aims to determine the smallest level of distortion (*e.g.*, compression amount) at which degradation becomes

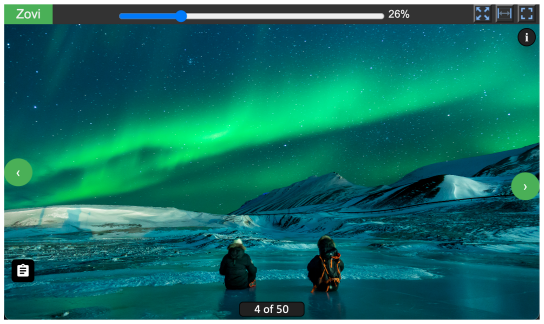


Figure S5. Screenshot of the UI of the ZOVI web application that we developed to annotate the IIS of an image.

perceptible. This level is called the JND and is conceptually analogous to the IIS, which can be interpreted as the lowest downscale factor that maximizes the perceived quality of an image. Given the similarity between the two tasks, we took inspiration from the JND assessment to design the annotation strategy of the IIS.

The literature on JND assessment has proposed various annotation methods, including binary search and slider presentation [9]. Among these, slider presentation has proven more effective, offering lower costs and higher precision. Therefore, as explained in Sec. 4, we adopt a similar approach and develop an annotation tool (ZOVI) – shown in Fig. S5 – that displays a slider that allows the users to downscale the image from its original size ($scale = 1$). In contrast, the binary search method is less efficient as it requires multiple independent participant judgments. For instance, if we were to consider 100 possible scale values the binary search would require $\lceil log_2 100 \rceil = 7$ comparisons. At a median of 3 seconds per judgment, it would take around 21 seconds to determine the IIS of an image. Empirical evidence suggests that the slider presentation is faster, taking 15 seconds per image, and provides more precise results for JND assessment [9].

S2.3. Intrinsic Scale Aggregation Strategy

Following the methodology detailed in Sec. 4, we collect 20 IIS annotations for each image (10 participants $\times$ 2 opinions each). To obtain the ground-truth IIS labels we need a strategy to aggregate the single subjective opinions. We refer to this aggregated IIS value (*i.e.*, the ground-truth one) as the Mean Opinion Intrinsic Scale (MOIS), drawing an analogy to the Mean Opinion Scores (MOS) used for assessing perceived image quality. One might naïvely compute the arithmetic mean of the single IIS opinions, similar to how MOS are computed. However, the scale of the slider of our annotation tool is inherently non-linear. For instance, when an image's size doubles from 50% to 100%, the scale difference on the slider is twice that of when the image doubles from 25% to 50%. Therefore, values from different parts of the slider range should not be equally weighted, as plain averaging would. To address this, we apply a logarithmic transformation ($\log_2$) to the individual IIS values, which linearizes the scale before averaging. After averaging, we then exponentiate the result to revert to the original scale. This approach is equivalent to computing the geometric mean of the individual subjective opinions to obtain the MOIS of each image.

Formally, let $\Omega_j(I)$ be the j-th subjective opinion associated with the image I, with $j = 1, \ldots, 20$. Then, we compute the MOIS $\Omega(I)$ of the image I by using the geometric mean:

$$\Omega(I) = 2^{\frac{\sum_{j=1}^{N} log_2(\Omega_j(I))}{N}} = \sqrt[N]{\prod_{j=1}^{N} \Omega_j(I)} \qquad \text{(S1)}$$

In this way, we account for the non-linearity of the slider scale. The final value $\Omega(I)$ represents the ground-truth IIS value of the image I, or MOIS. Across the 785 images composing our dataset, the average MOIS is 0.347, with per-image MOIS ranging from 0.060 to 0.811.

S2.4. Examples of Image-IIS Pairs

IISA-DB is designed to be diverse, featuring images with varying content and quality levels. Fig. S7 presents examples of image-IIS pairs from our dataset. Since we cannot display the images at their original size, we have cropped relevant sections and shown them at their original scale. Note that due to rescaling in the PDF viewer, the images may not appear exactly as they did to participants in the subjective study, *i.e.*, at a 1:1 ratio of image to native screen pixels. However, the scale ratio between the original and intrinsic image crops remains consistent.

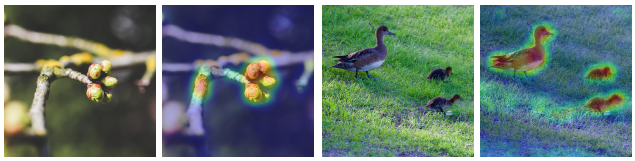
S2.5. Examples of Attention Maps

We visualize the attention maps of the TOPIQ model trained with our WIISA approach in Fig. S6, using the method described in the original paper [1]. Fig. S6 (a) shows an image featuring a small foreground object in focus against a blurry background, while Fig. S6 (b) depicts a high-semantic object surrounded by high-frequency texture. The model attends to in-focus and high-semantic regions, suggesting that IIS predictions are primarily driven by the degradation of high-level semantic content.

S3. Additional Experimental Results

S3.1. Implementation Details

During the training of each baseline, we extract square center crops with a size of 1536 pixels. We use data augmentation techniques that do not affect image quality, namely horizontal and vertical flips, with a probability of 0.5. We use Lanczos interpolation to generate the weakly labeled



(a) (b)

Figure S6. Visualization of TOPIQ's model attention.

Method	SRCC	PLCC	RMSE	MAE
TOPIQ (SPAQ)	0.042	0.088	0.323	0.290
TOPIQ (UHD-IQA)	**0.054**	**0.166**	**0.290**	**0.255**

Table S1. Evaluation of the performance on the IISA-DB dataset of the zero-shot multi-scale IISA approach based on the TOPIQ [1] model. $(\cdot)$ indicates the pre-training dataset. Best scores are highlighted in bold.

image-IIS pairs with our approach. We set the number of weak labels n_{wl} to 2 and the downscaling threshold δ to 0.65. During testing, we feed each model the image at its original scale as input. We carry out the experiments on an NVIDIA H100 80GB GPU.

S3.2. Zero-Shot Multi-scale IISA

Given the formulation of IIS reported in Eq. 1, we can employ the quality scores predicted by a pre-trained NR-IQA method to estimate the IIS automatically. Specifically, given n_s uniformly sampled scales s in the range $[s_{lb}, 1]$, we can use a pre-trained NR-IQA model to assess the quality of each downscaled version I^s of an input image I and then find the scale for which the predicted quality is the highest. This would be an estimate for the IIS of image I.

Following the evaluation protocol described in Sec. 6.1, we assess the performance of this zero-shot multi-scale approach on the proposed IISA-DB dataset. We employ two versions of the TOPIQ [1] model pre-trained on the SPAQ [4] and UHD-IQA [7] datasets, which feature high-resolution images similar to those in IISA-DB. We use $n_s = 100$ scales and report the results in Tab. S1. We observe that the zero-shot multi-scale approach achieves unsatisfactory performance, regardless of the pre-training dataset. We attribute this to NR-IQA models struggling to handle the change in perceptual quality caused by downscaling, as noted in [8, 11]. In addition, the multi-scale approach requires multiple model forward passes to obtain a single IIS prediction, which can be inefficient.

S4. Limitations

The annotation process for IISA is time-consuming, requiring a median of 15 seconds per image versus 3 seconds for NR-IQA. Moreover, the significant effort and concentration required often make it challenging for typical crowdsourcing workers. In our pilot experiments, we found a high disqualification rate (about 90%) among participants from Amazon Mechanical Turk, highlighting the need for more qualified but expensive expert annotators – the latter participated in our experiments. Despite these challenges, the superior sensitivity of IISA justifies its use in scenarios requiring highly precise quality judgments.

When collecting subjective IIS annotations, we employed Lanczos interpolation to rescale the images. While such an interpolation method guarantees high-quality results, different algorithms could be considered. For example, one could employ faster but lower-quality methods such as bilinear, or higher-quality, albeit slower, algorithms such as the full 2D Lanczos one. While the experiments reported in Sec. 6.3 show that the interpolation algorithm does not make a significant difference for predictive IISA models, future work could extend our study by examining the effects of different interpolation methods for subjective IISA.

The scope of this study is focused on the impact of low-level distortions in User-Generated Content (UGC) images. Consequently, the applicability of our findings to specialized domains governed by different perceptual criteria, such as text-heavy images where readability is the primary quality indicator, remains largely unexplored. Extending the proposed framework to these areas is a promising direction for future work, which would involve adapting protocols to capture domain-specific quality factors.

S5. Future Work

Our work suggests several promising directions for future research, including: 1) extending our dataset to incorporate more types of images, such as super-resolved, synthetically distorted, and computer-generated images; 2) conducting additional subjective studies – similar to those of KonX – to futher validate our assumption related to how image quality changes with scale; 3) analyzing the impact of different types of distortion on the IIS.

References

[1] Chaofeng Chen, Jiadi Mo, Jingwen Hou, Haoning Wu, Liang Liao, Wenxiu Sun, Qiong Yan, and Weisi Lin. TOPIQ: A Top-down Approach from Semantics to Distortions for Image Quality Assessment. *IEEE Transactions on Image Processing*, 2024. 4, 5

[2] Robert F DeVellis and Carolyn T Thorpe. *Scale Development: Theory and Applications*. Sage Publications, 2021. 2

[3] Ruigang Fang, Dapeng Wu, and Liquan Shen. Evaluation of Image Quality of Experience in Consideration of Viewing Distance. In *IEEE China Summit and International Conference on Signal and Information Processing (ChinaSIP)*, pages 653–657. IEEE, 2015. 2

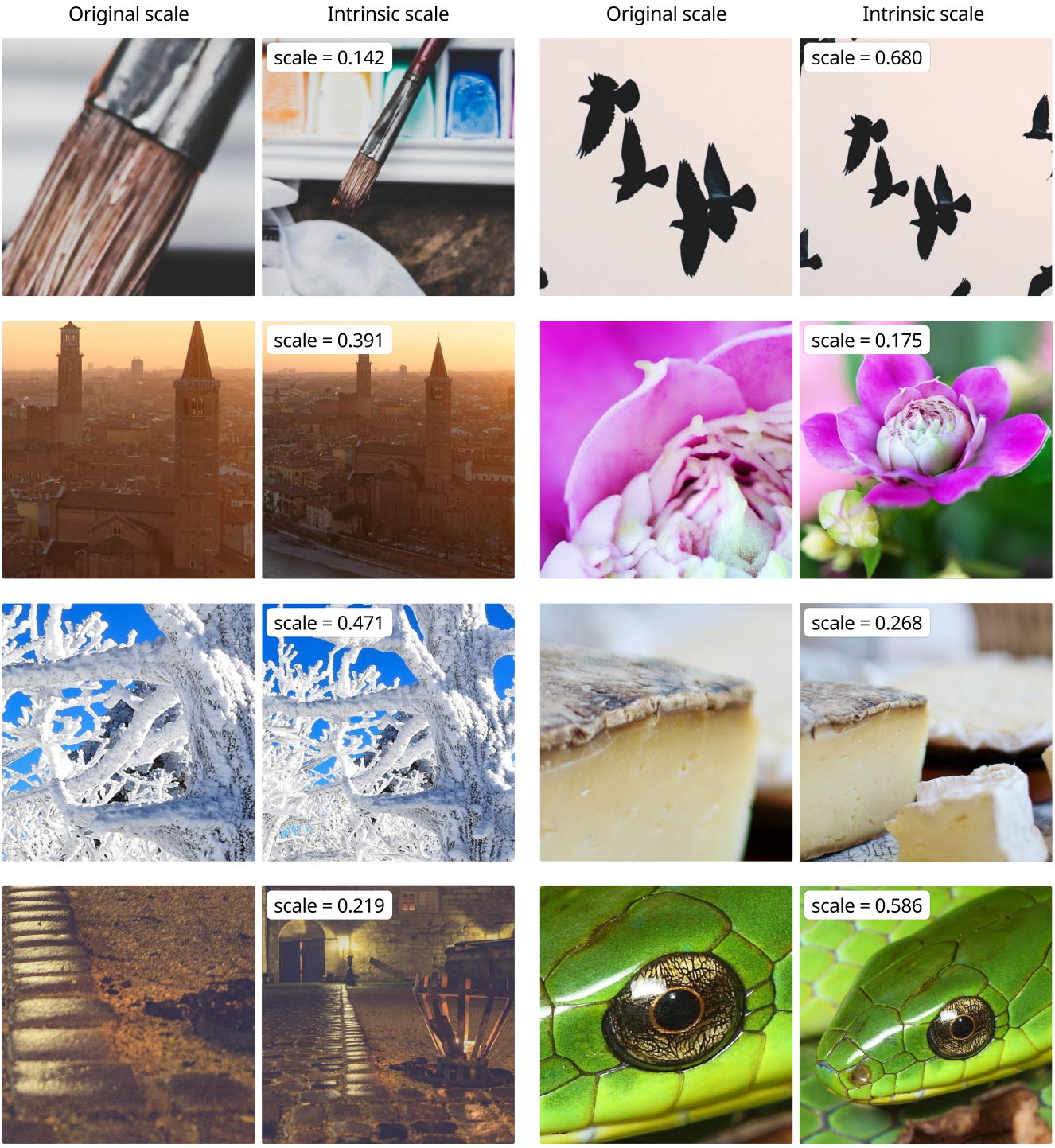


Figure S7. Examples of image-IIS pairs from the proposed IISA-DB dataset. Pairs of images are displayed in two columns. For each pair, the image on the left represents a crop from the original, while the image on the right depicts a crop from the original downscaled to the intrinsic scale. The content regions overlap between the two crops in each pair.

[4] Yuming Fang, Hanwei Zhu, Yan Zeng, Kede Ma, and Zhou Wang. Perceptual Quality Assessment of Smartphone Photography. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3677–3686, 2020. 5

[5] Ke Gu, Min Liu, Guangtao Zhai, Xiaokang Yang, and Wen-

jun Zhang. Quality Assessment Considering Viewing Distance and Image Resolution. *IEEE Transactions on Broadcasting (ToB)*, 61(3):520–531, 2015. 2

[6] Vlad Hosu, Hanhe Lin, Tamas Sziranyi, and Dietmar Saupe. KonIQ-10k: an Ecologically Valid Database for Deep Learning of Blind Image Quality Assessment. *IEEE Transactions on Image Processing*, 29:4041–4056, 2020. 2, 3

[7] Vlad Hosu, Lorenzo Agnolucci, Oliver Wiedemann, and Daisuke Iso. UHD-IQA Benchmark Database: Pushing the Boundaries of Blind Photo Quality Assessment. *arXiv preprint arXiv:2406.17472*, 2024. 5

[8] Huang Huang, Qiang Wan, and Jari Korhonen. High Resolution Image Quality Database. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3105–3109. IEEE, 2024. 5

[9] Hanhe Lin, Guangan Chen, Mohsen Jenadeleh, Vlad Hosu, Ulf-Dietrich Reips, Raouf Hamzaoui, and Dietmar Saupe. Large-scale Crowdsourced Subjective Assessment of Picturewise Just Noticeable Difference. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(9):5859–5873, 2022. 3, 4

[10] Xinwei Liu, Marius Pedersen, and Jon Yngve Hardeberg. CID: IQ - a New Image Quality Database. In *International Conference on Image Signal Processing (ICISP)*, pages 193–202. Springer, 2014. 2

[11] Oliver Wiedemann, Vlad Hosu, Shaolin Su, and Dietmar Saupe. KonX: Cross-Resolution Image Quality Assessment. *Quality and User Experience*, 8(1):8, 2023. 1, 2, 3, 5