מכון ויצמן למדע
WEIZMANN INSTITUTE OF SCIENCE

Thesis for the degree
Master of Science

עבודת גמר (תזה) לתואר
מוסמך למדעים

Submitted to the Scientific Council of the
Weizmann Institute of Science
Rehovot, Israel

מוגשת למועצה המדעית של
מכון ויצמן למדע
רחובות, ישראל

By
Itai Antebi

מאת
איתי ענתבי

הערכת איכות תמונה ללא תמונת התייחסות ובהיעדר דוגמאות למידה

Zero-Shot No-Reference Image Quality Assessment

Advisor:
Prof. Michal Irani

מנחה:
פרופ' מיכל אירני

August 2023

אלול תשפ"ג

Acknowledgements

I would like to express my special thanks and gratitude to Prof. Michal Irani for giving me the opportunity to perform research under her supervision. She not only provided exceptional professional guidance during our work together, but also displayed remarkable compassion during a personally challenging time. I'd like to also express my appreciation to Dr. Shai Bagon that accompanied the research since its inception. It was a great honor working alongside him, and I am proud to call him a friend as well as a colleague. Furthermore, I'd like to thank my associates in my research group, for fruitful conversations and helpful insights along the way. Lastly, I'd like to thank my future wife for the endless support and love, without whom I never would have achieved all that I have.

Abstract

Within the domain of computer vision and image processing, researchers frequently grapple with inverse problems, which encompass tasks such as super-resolution, image enhancement, deblurring, and dehazing, among others. These problems involve the complex task of reconstructing high-quality images from degraded inputs. A critical aspect of this endeavor lies in assessing the quality of these reconstructed images. The common approach has been to measure image quality by comparing the results against a ground truth reference image, employing metrics such as Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), etc. This approach relies on the assumption that a pristine reference image exists. However, this assumption is only true for synthetically degraded images, initiated from a high-quality image. On the other hand, when a degraded image is captured in real-world scenarios, a reference image is inaccessible. As a result, the need for No-Reference Image Quality Assessment (NR-IQA) has emerged. NR-IQA methods have shown promise in assessing image quality without reference images, achieving remarkable success on established benchmarks. Nonetheless, they primarily rely on prior training from datasets, and their generalization performance on out-of-distribution data remains a significant challenge. In this thesis, we introduce a novel approach: ***zero-shot NR-IQA****. To the best of our knowledge, we are the first to do so. As a zero-shot method, it does not suffer from over-fitting and poor generalization, as current state-of-the-art methods do. We base our work on the similarity of patches across scales in a natural image. We study the patch recurrence relation between consecutive scales and show that in a high-quality image, it is preserved across different scales, where it is violated in degraded images. We employ this phenomenon to construct an image-specific score, measuring the deviation between an evaluated image's patch recurrence relation at its original resolution and that at a lower resolution. Empirical results from our experiments validate the efficacy of our approach, demonstrating that our scoring mechanism aligns consistently with human judgments across a diverse range of images and degradations.*

1. Introduction

Image Quality Assessment (IQA) is the task of obtaining an objective measure to evaluate the quality of an image, which best mimics humans' perception. Many computer vision algorithms are designed for a human consumer at the end of the pipeline. Therefore, the ability to estimate a human's judgment of an image is of great interest. Achieving a good IQA could be useful for many computer vision tasks: compression, super-resolution, denoising, deblurring, image enhancement, image generation, etc.

IQA models are traditionally categorized according to the additional information supplied to the model. Full-reference IQA (FR-IQA) models assume the accessibility of the evaluated image's high-quality reference counterpart. Popular examples are Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM) [28] and perceptual similarity. These metrics are frequently used when possible. Unfortunately, these metrics could not be used for authentically degraded images, for which there exists no reference image. The lack of information concerning the reference image is generally a realistic condition. Thus, emerged the need to assess the quality of an image without a reference. In No-reference IQA (NR-IQA), also known as blind IQA, only the evaluated image is available to the model. NR-IQA is the most challenging category of IQA, and is widely researched and continuously improves.

In order to learn how humans perceive the quality of images, and to set a uniform standard for IQA methods, some datasets were curated: LIVE [25], TID2008 [21], CSIQ [9], TID2013 [20], LIVE_MD [5], KADID-10k [10], etc. These datasets contain degraded images with human-labeled quality scores, called Mean Opinion Scores (MOS). An IQA method is evaluated on a given dataset by measuring the correlation between the method's scores and the GT human-labeled MOS scores on the dataset's images. State-of-the-art NR-IQA methods today reach near impeccable results. This is, however, only when they are evaluated on the same dataset they were trained on. Evaluating a method on one dataset after it was trained on another results

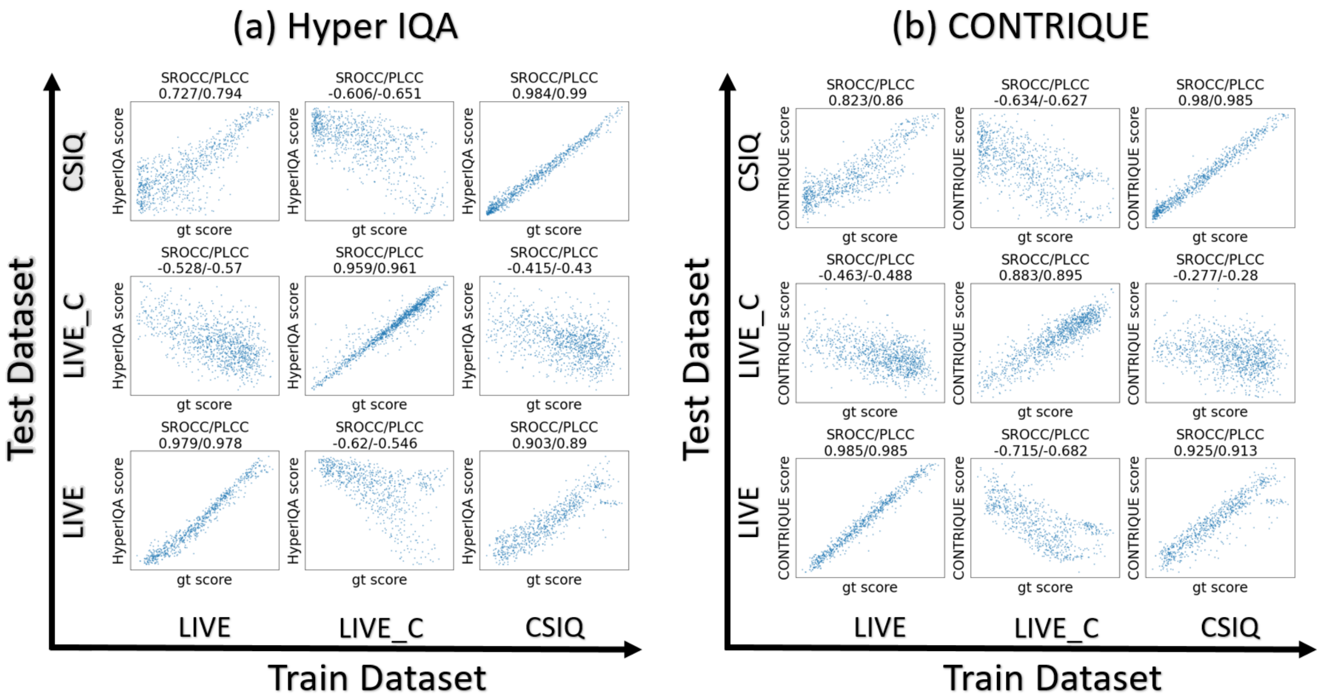

Figure 1: **Poor Generalization Capabilities.** Correlation between GT scores and predicted scores, of two SotA NR-IQA methods: (a) HyperIQA [27], and (b) CONTRIQUE [15]. Each plot in the figures shows the correlation between the GT scores of the test dataset (x-axis) and the predicted scores of the method (y-axis). The subtitle indicated the Spearman and Pearson correlation metrics. The different plots correspond to training and testing the method on different datasets. When tested on the same dataset they were trained upon, they reached an impressive correlation, as can be seen in the plots along the main diagonal. However, this correlation is lost when attempting to generalize outside the trained dataset, as can be seen in the plots outside the main diagonal.

in very poor correlation, as can be seen in Figure 1. This suggests that the generalization capabilities of SotA methods across datasets are weak. Without generalizing outside the trained dataset, an IQA method would have little practical uses.

We believe the limited generalization capabilities of previous work are in part due to the limitations of the datasets. Acquiring an IQA dataset proves to be a cumbrous and costly endeavor, resulting in a scarcity of such datasets, with the existing ones often containing a limited number of examples. Our method, on the other hand, is not trained on any data. We are the first to introduce a **Zero-Shot NR-IQA**. We consider the very low-resolution version of an image as a high-quality version of itself, by utilizing the tendency of degradations to mostly affect the high-frequencies of an image. We are therefore able to assess the quality of an image by measuring the deviation between the characteristics of the evaluated image and its low-resolution high-quality counterpart. The selected characteristic leverages the property of patch recurrence across scales in high-quality images. We showed that unlike common belief, while natural images share similar patches across scales, they do not share the same patch distribution. We addressed the shift in distribution between two consecutive scales, measuring how many patches recur how many times, as a recurrence histogram. We observed that a recurrence histogram is preserved across scales for a high-quality natural image, while it is violated for a degraded image. In order to construct the recurrence histogram, we generalized the well-known Sliced-Wasserstein-Distance. Our ultimate score, which is entirely image-specific, correlates well with multiple degradations.

The rest of the thesis is organized as follows: in section 2 we examine related work, in section 3 we describe our method, section 4 is dedicated to demonstrating our experimental results and in section 5 we summarize and discuss further work.

2. Related work

Assessing the quality of an image is a challenging task. First, there are many types of degradations; a variety of blurring effects, noises, blockiness, quantization, ringing, aliasing, etc. Images could suffer from a single degradation or a mixture of multiple degradations. Second, in a real-world scenario, authentic degradations behave differently than synthetic degradations (for example, a local degradation that affects only a small part of the image). In addition, recent studies have shown that people's perceived degradation severity is also dependent on the image content. Therefore, over the past decades, considerable work has been done in an attempt to solve this difficult problem.

FR-IQA metrics have been used for many years in image quality assessment. The long-established mean-squared-error (MSE) metric measures the pixel-wise average squared difference between a reconstructed image and the reference image. PSNR is based on MSE but expressed in decibels, resulting in a more interpretable measure. The SSIM [28] leverages the strong inter-dependencies of pixels. It considers luminance, contrast, and structure in various windows of an image in an attempt to capture visual distortions. Multi-scale SSIM (MS-SSIM) [29] combines the SSIM index over several versions of the image at various scales. The visual information fidelity (VIF) index [24] takes a different approach, quantifying the information that is present in the reference image, and measures how much of this reference information can be extracted from the distorted image. Lately, with the emergence of deep learning, perceptual similarity metrics that use deep features have been invented. They use pre-trained neural networks, such as VGG [26], ResNet [4], or similar architectures, to extract high-level features from images. These features capture abstract information from the images, allowing a semantically meaningful comparison between a reconstructed image and a reference. Although all these metrics are effective and are still in active use today, they require access to a reference image for evaluation. However, in many practical scenarios, obtaining a reference image for comparison is impossible. Thus, NR-IQA methods were developed to evaluate the quality of an image without relying on a pristine reference image.

Most NR-IQA models use the same method: a feature extractor component followed by a regression component. The regression component utilizes one of two primary approaches: it either regresses the features of distorted images to their ground truth MOS, or it regresses features from pristine images to model natural scene statistics (NSS). The latter case assumes that natural images occupy a subspace of the entire space of possible images. Images were presumed to be of low quality if their statistics deviated greatly from the NSS model. While both components play important roles, the primary emphasis of research was directed towards the feature extraction component.

Traditional methods used hand-crafted features that are sensitive to distortions. BLIINDS [22] and BLIINDS-II [23] use discrete cosine transform coefficients for feature extraction, BRISQUE [17] use mean subtracted contrast normalized coefficients, and DIIVINE [19] use wavelet coefficients. The latter has a two-stage solution which first identifies the type of degradation and only then estimates the quality, making the method distortion-aware. CORNIA [30] were among the first to extract quality representative features with no human labels, resulting in an opinion-unaware model. A learned codebook from local patches was at the base of their technique. NIQE [18] had no exposure to distorted images and was trained solely on a corpus of high-quality images. Inspired by NIQE [18], IL-NIQE [31] followed the same path and enriched the features with other quality-aware features such as gradient features, log-Gabor filter responses, and color statistics.

During the last decade, deep CNN models revolutionized the field of computer vision. They achieved tremendous success across many tasks, and NR-IQA is no exception. However, having access to a large-scale dataset is usually a necessary condition for the success of a CNN. Unfortunately, in the field of NR-IQA, acquiring such datasets is expensive, takes a lot of time, and requires a lot of work. This results in a handful of datasets, most of whom contain roughly 1,000 labeled images only. These amounts of data are not sufficient for training a deep CNN from scratch. Consequently, most methods turn to transfer learning and use the small IQA datasets solely for fine-tuning. BIECON [6] uses reference images during training to supervise feature extraction. MEON [14], DB-CNN [32], and CONTRIQUE [15] first learn to identify distortion types from large-scale available datasets before learning quality prediction. RankIQA [11] and dipIQ [13] synthetically generate couples of images where their relative quality is known and not their quality score, allowing them to learn how to rank images according to their quality. Kim *et al.* [7] showed that image features extracted from architectures like AlexNet [8] or Resnet [4] pre-trained on image classification can be very useful for quality assessment. DB-CNN [32] employ a deep bilinear model for NR-IQA that works for both synthetically and authentically distorted images, utilizing pre-trained VGG-16 [26] architecture for the authentic distortions. HyperIQA [27] introduced a hyper-network architecture to condition the quality prediction on the image's content.

All previous methods are eventually trained on IQA datasets. While the logic is clear, it necessarily induces undesired properties and limitations; Not only are they unable to generalize to distortions not presented in the trained dataset, but they are often unable to generalize even to images with the same distortions sampled from another dataset (see Figure 1).

In this work, we propose a Zero-Shot NR-IQA method. Not training on any dataset releases us from such limitations. In addition, as opposed to other methods, our method is explainable. We explicitly define our premises. This is a great advantage as it enables us to know in advance in which cases our method might fail and why.

3. Method

Our method works in the following manner: Given an input image whose quality we wish to assess, we first consider the low-resolution[1] version of the image to be a high-quality (though smaller) version of that image. In 3.1 we justify why such assumption holds. Our score would measure how much the input image's characteristics deviate from the high-quality version; the more it deviates - the lower the quality of the input image is considered to be. The characteristics we use in our method build upon the patch recurrence property. It has been shown that patches of high-quality natural images recur abundantly both within and across scales. We observe *how many patches* recur *how many times* between two neighboring resolutions of an image. We address this recurrence as the Recurrence Histogram of the image at that resolution. In 3.2 we explain the Recurrence Histogram and supply intuition. In order to construct the recurrence histogram, we relied on Sliced-Wasserstein Distance and generalized it. In 3.3 we give the technical details of how we construct the recurrence histogram. Our ultimate score would be the KL-divergence between the Recurrence Histograms of the input image and its low-resolution high-quality counterpart.

3.1. Low-resolution supervision

As a zero-shot method, given only an evaluated image with unknown quality, from where could we derive information concerning the quality of the image? Image distortions often affect mostly the higher frequencies of the image. For example, blurring artifacts are expressed in the absence of high-frequencies, and noising artifacts are expressed in redundant high-frequencies. When shrinking down an image, in order to avoid aliasing, it is low-pass filtered before it is sub-sampled. The resulting image is a low-resolution image that is not sensitive to the original distortion.

Valid intuition could be found in the pixel space. A blur smears edges over multiple pixels in the high-resolution image. In Figure 2, for example, the edge of the roof outlined in red spreads over a few pixels in high quality, and over multiple pixels once blurred. Shrinking down the image causes the edge to spread out over fewer pixels, resulting in a sharper image at low resolution. Notice in Figure 2 how the low-resolution version of the blurred image resembles the high-quality counterpart. Similarly, shrinking down a high-resolution image with zero-mean noise, averages out the noise, resulting in a cleaner image at low-resolution. This is demonstrated in Figure 2 by the smooth sky outlined in blue. While at high resolution they are corrupted

[1] In our terminology, we distinctly differentiate two terms that in other places might be used interchangeably: Resolution, and Quality. Resolution refers to the size of an image, i.e. number of pixels. Quality, on the other hand, refers to the perception of the image considering its resolution. An image could be very large and very noisy, making it high-resolution and low-quality at the same time. Similarly, an image could be low-resolution and high-quality at once.

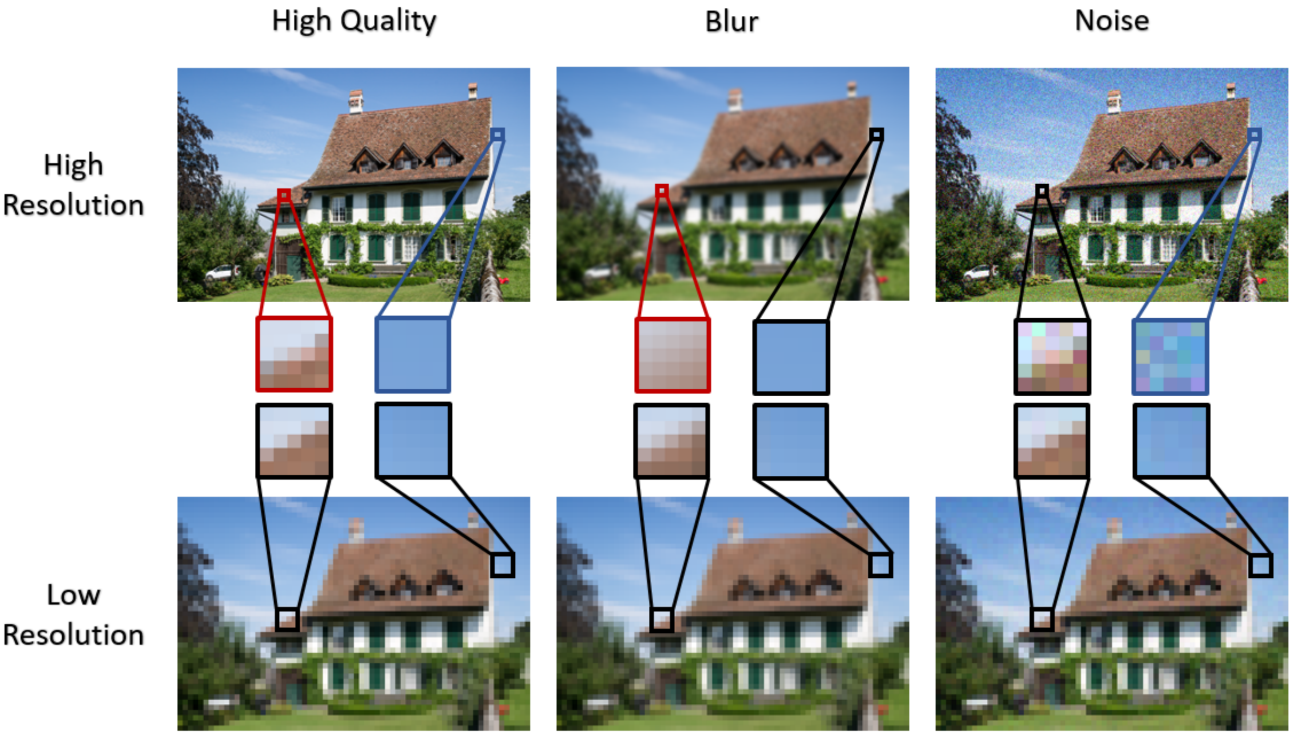

Figure 2: **Low-resolution is of higher quality.** Top row - high-resolution image under different distortions. Bottom row - same images at 1/5 resolution. At high resolution, degradations are apparent. The edge of the roof is smeared by the blur (outlined in red), and the smooth sky is corrupted by noise (outlined in blue). At low resolution, these degradations vanish, all images resemble the high-quality version, and the patches are sharp and clean.

by additive white Gaussian noise, at low resolution the noise is balanced and the clear skies emerge. This phenomenon has been previously shown by [33, 16, 12].

Therefore, given a high-resolution image, with or without distortion, we could address the low-resolution version of that image as a high-quality image. Obviously, the more the high-resolution input image is degraded, the lower the resolution we must reach in order to obtain a high-quality image (see Figure 3). As a zero-shot method, the low-resolution version of an image is where we derive the information about the characteristics of the high-quality version of the given image.

3.2. Recurrence Histogram

The characteristic we use to compare the evaluated input image against its high-quality counterpart is the recurrence histogram. A recurrence histogram depicts how many patches recur how many times between an image and a 1/2 scaled version of that image.

Given two images at consecutive scales of the image pyramid, we wish to reconstruct a higher-resolution image using a lower-resolution image's patches. Let I^0 denote the input image at the top of the image pyramid, and I^l denote the image at scale l of the image pyramid. I^l is the result of shrinking I^0 by a factor of $(1/2)^l$. Thus, we wish to reconstruct I^l using $I^{(l+1)}$'s patches. It has been shown in [3] that small patches in a high-quality natural image tend to recur many times inside the image, both within the same scale, as well as across different scales. This implies that such reconstruction is possible. However, the question arises, how many times would each patch of $I^{(l+1)}$ be used for such reconstruction?

Since $I^{(l+1)}$ is at half the resolution of I^l, it has $1/4$ the number of patches, concluding that the average patch of $I^{(l+1)}$ would recur 4 times. If all patches of $I^{(l+1)}$ would recur 4 times, it could be said that the two images share the same patch

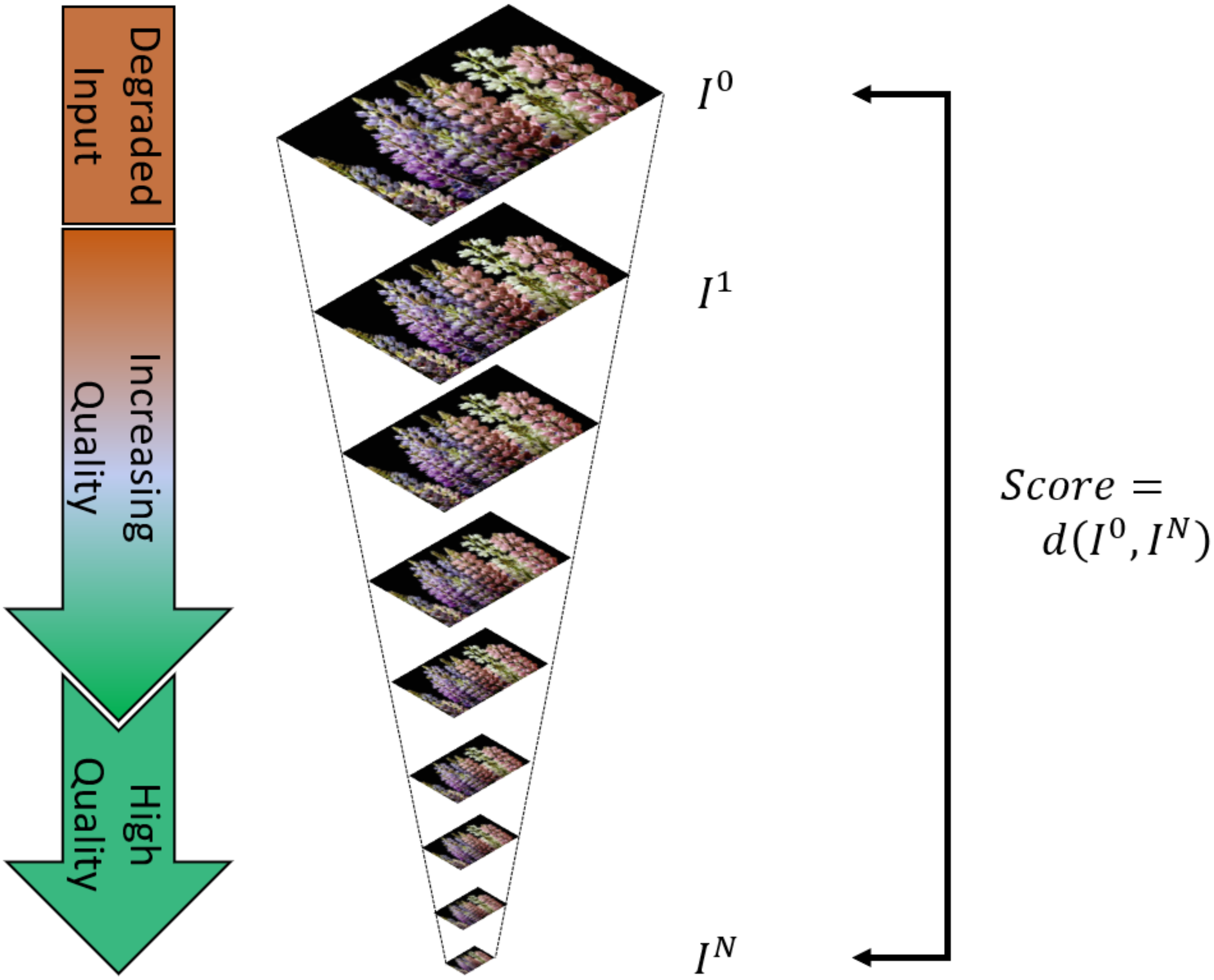

Figure 3: **Method Overview.** As the degraded input gets smaller in resolution, its quality increases. The very low-resolution version of the input image is considered to be a high-quality version of the image. Our score estimates the quality of the input image by measuring the deviation of the characteristics of I^0 from I^N.

distribution. However, as opposed to common belief, while high-resolution and low-resolution images share common *patches*, they do not share the same *patch distribution*, since the frequency of occurrences changes across different scales.

Consider the simple example of I^l containing a 2*2 checkerboard pattern (see Figure 4). It has uniform patches (both black and white), patches containing an edge, and patches containing a corner. In that case, $I^{(l+1)}$ would have the exact same patches. However, the number of uniform patches at $I^{(l+1)}$ would be $1/4$ of that at I^l, the number of patches containing an edge at $I^{(l+1)}$ would be $1/2$ of that at I^l, and both images would have the same number of patches containing a corner. Thus, even in our simple example, the two images possess the exact same patches, but not the same patch distribution, as different patches recur a different amount of times. In a natural image, the situation is ever more complex.

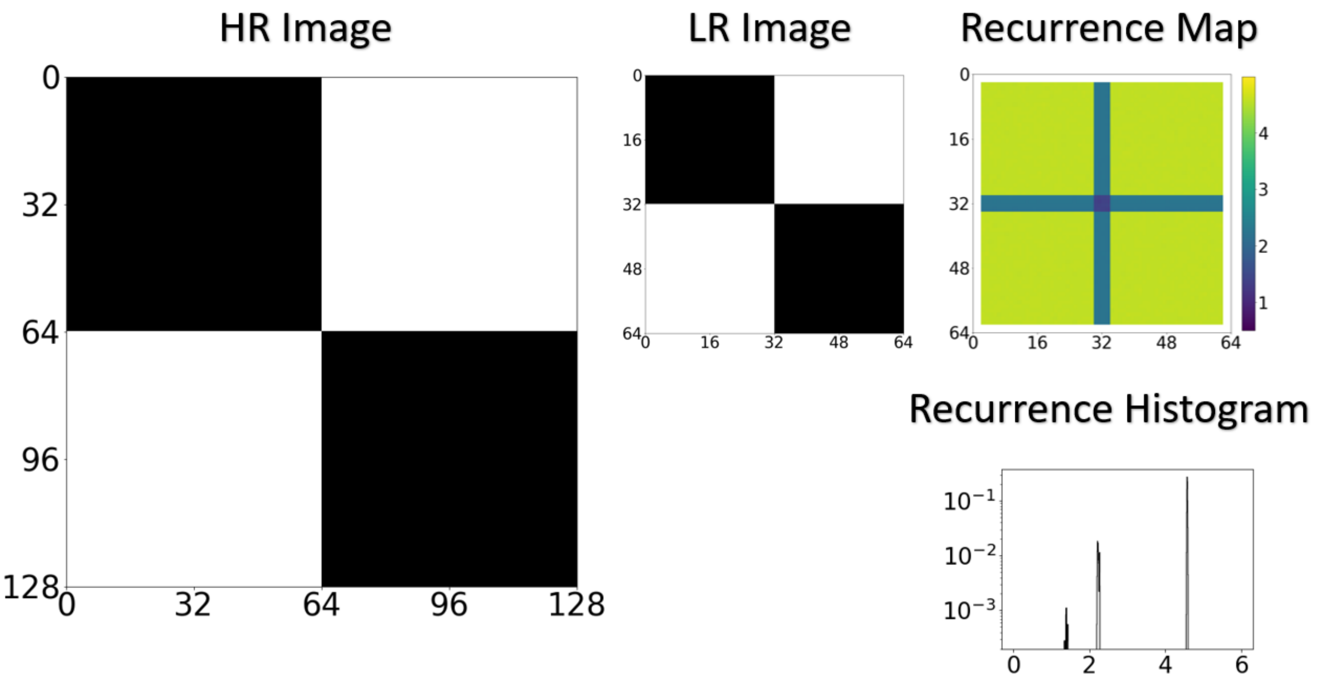

Figure 4: **Recurrence Map and Histogram of Toy Example.** The high-resolution and low-resolution images share the exact same patches. Different patches recur at different amounts of times. The recurrence map shows how many times each patch in the low-resolution image recurs in the high-resolution image. Smooth patches recur more than edge patches, and corner patches recur the least. The recurrence histogram is a characteristic of the high-resolution image.

3.3. Weighted Sliced Wasserstein Distance

How can we construct such a Recurrence Map or Histogram? A naive solution would be to search for each patch in I^l the nearest-neighbor patch in $I^{(l+1)}$ and count for each patch in $I^{(l+1)}$ how many patches in I^l found it as their nearest-neighbor. However, this is computationally expensive and results in a very strict outcome. Our method must be robust to slight changes in pixel values in order for it to hold as a property that is preserved across scales.

We readdress the question as such: what is the best mapping between $I^{(l+1)}$'s patch-distribution to I^l's patch-distribution, when the mapping is restricted to re-sampling alone? In other words, we wish to find some weighting, such that sampling from I^l's patch-distribution would be as close as possible to sampling from the *weighted* $I^{(l+1)}$'s patch-distribution.

We were inspired by previous methods that found sliced-Wasserstein-distance (SWD) useful for measuring the distance between two images' patch-distributions [2]. We generalized SWD and introduced Weighted-SWD (WSWD) by resampling patches according to chosen weights that aren't uniform. In our method, we look for the best weights such that the WSWD is minimized.

In SWD, the patches from both images are projected onto one dimension by a random unit vector, and the Wasserstein distance (WD) is computed on the projections. The SWD is the expectation of the WD over the random unit vector (usually approximated by the mean WD over a selected set of random vectors). Let $P^l, P^{(l+1)}$ be the set of patches from $I^l, I^{(l+1)}$ accordingly, ω a unit vector, and $P^l_\omega, P^{(l+1)}_\omega$ the set of projected patches $P^l, P^{(l+1)}$ onto ω:

$$SWD(P^l, P^{(l+1)}) = \mathop{\mathbb{E}}_{|\omega|=1}(WD(P^l_\omega, P^{(l+1)}_\omega)) \tag{1}$$

However, WD (and therefore SWD as well) is restricted such that the two images must contain the same number of

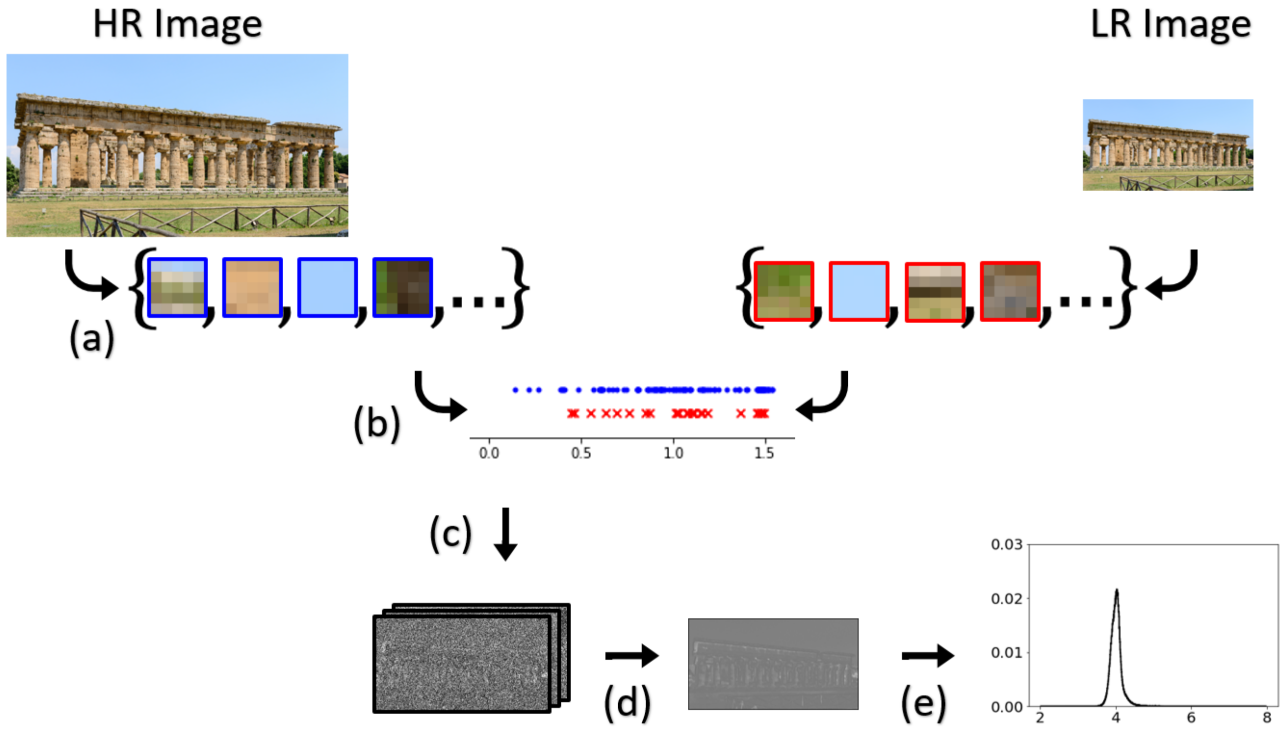

Figure 5: **Patch-recurrence relation between two images.** We compute for a set of images (here an HR image and its corresponding LR counterpart) how many patches recur how many times. (a) Extract patches from the images. (b) Project the patches to a single dimension using a random unit vector. (c) Count for each projected LR patch how many HR projected patches address it as a nearest neighbor. We present this counting as a heat map. (d) Average over the different heat-maps created using different random projections. The objects in the input image suddenly appear in the resulting heat map. (e) Observe the average heat map's values as a histogram.

patches. When computing SWD between two images' patch distributions, the patches from the smaller image are resampled uniformly to compensate for the lack of patches. Let $W^{(l+1)}$ be the weights, indicating how many times each patch in $P^{(l+1)}$ is resampled. We define:

$$WSWD(P^l, P^{(l+1)}, W^{(l+1)}) = SWD(P^l, W^{(l+1)}P^{(l+1)}) \tag{2}$$

We are therefore interested in the best possible weights $W^{(l+1)}$ which bring $W^{(l+1)}P^{(l+1)}$ distribution as close as possible to P^l distribution.

$$W_*^{(l+1)} = \arg\min_{W^{(l+1)}} WSWD(P^l, P^{(l+1)}, W^{(l+1)}) \tag{3}$$

In order to do so, we follow a similar path as SWD: We first project the patches from both images onto one-dimension by a random unit vector (Figure 5 (b)). Then, for each projected patch in P^l_ω, we find the nearest projected patch in $P^{(l+1)}_\omega$, and add 1 to its weight. This gives us the best weighing $W^{(l+1)}$ for the given projection ω (Figure 5 (c)). We can reshape the weights to the shape of $I^{(l+1)}$, by replacing each pixel in $I^{(l+1)}$ with the corresponding weight of the surrounding patch centered at that pixel. These weight maps tend to be very noisy, but when averaged over many random projections, a clear weight map is revealed (Figure 5 (d)). We consider the histogram of $W_*^{(l+1)}$ to be the Recurrence Histogram of I^l (Figure 5 (e)).

4. Experiments

The relation we defined, between two images at neighboring scales of the pyramid, must hold several properties: First, given a high-quality image, we wish the relation would be preserved across the different scales of the pyramid. Second, we hypothesized that degrading the high-resolution image would have little effect on the low-resolution image. This suggests that the relation at the bottom of the pyramid would remain the same, regardless of the high-quality image's degradation severity. Third, as more severe degradation is applied to the high-resolution image, the relation at the top of the pyramid should gradually deviate from the relation at the bottom. A representative example of a relation at different degradations, severities, and scales is showcased at Figure 6.

Observing the shape of the relation histogram of the high-quality image, it resembles a skewed Gaussian distribution. Different high-quality images had different relation histograms: diverse variance, skew, or multi-modal. But none resembled a δ function. This led us to the realization that, unlike common belief, high-quality natural images do not possess the same patch distribution across scales, even though they do share similar patches. In contrast, as the image gets blurred, the relation histogram tends to $\delta(4)$, leading us to the conclusion that blurry natural images do share the same patch distribution across scales.

Evaluating our method on standard general-purpose IQA datasets yielded mediocre results, as seen in Figure 7. On the one hand, the results clearly show a positive correlation, in some cases even beating the State-of-the-Art methods (when they are tested on out-of-distribution data). On the other hand, the correlation is not as significant as we had hoped.

Investigating our results, we revealed a few problems our method encountered when dealing with these datasets. First, they contain distortions that do not follow our premise of affecting mostly high frequencies, such as JPEG compression or brightening. Second, some of the images were degraded so severely, that no down-scaling could result in a high-quality version of the image. Moreover, some images were considered high-quality, even though they had a blurry background. This was usually the case for images that captured a single object close to the camera with a shallow depth of field.

Since it is very simple to degrade many pristine images to multiple severities, we used the images from DIV2K dataset [1], which do not possess a narrow depth of field, and degraded each of them to 8 different severity levels. We ensured the worst degradation severity would be pretty bad, but still kept some information from the original high-quality image. We used multiple degradation types following our premise, some not appearing in other datasets. Our mean score and standard deviation over the images are presented in Figure 8.

As expected and desired, the results show that the more an image is degraded - the more its high-resolution recurrence histogram deviates from its low-resolution recurrence histogram. This is consistent across many images and degradations (blurs, noises, and others). While the score of high-quality images is not exactly 0, it appears to be a very stable local minimum. This implies that the characteristic of natural images we found is robust.

5. Discussion

We presented a zero-shot NR-IQA. To the best of our knowledge, we are the first to do so. We observed that while high-quality images possess similar patches as their low-resolution counterparts, they do not share the same patch distribution. Based on that observation, we defined a patch recurrence relation and showed that for high-quality images the relation is preserved across scale. We utilized the fact that most degradations vanish at lower resolutions to obtain an image-specific high-quality relation. The final score was then calculated by the deviation of the evaluated image's relation at the original resolution from the relation at the low resolution.

Our method has multiple advantages over the current state-of-the-art methods: First, since our method is zero-shot and not trained on a particular dataset, it does not suffer from poor generalization capabilities to novel images or distortions, as other methods do. In addition, since our approach is entirely transparent, with clearly defined premises, we are able to anticipate the

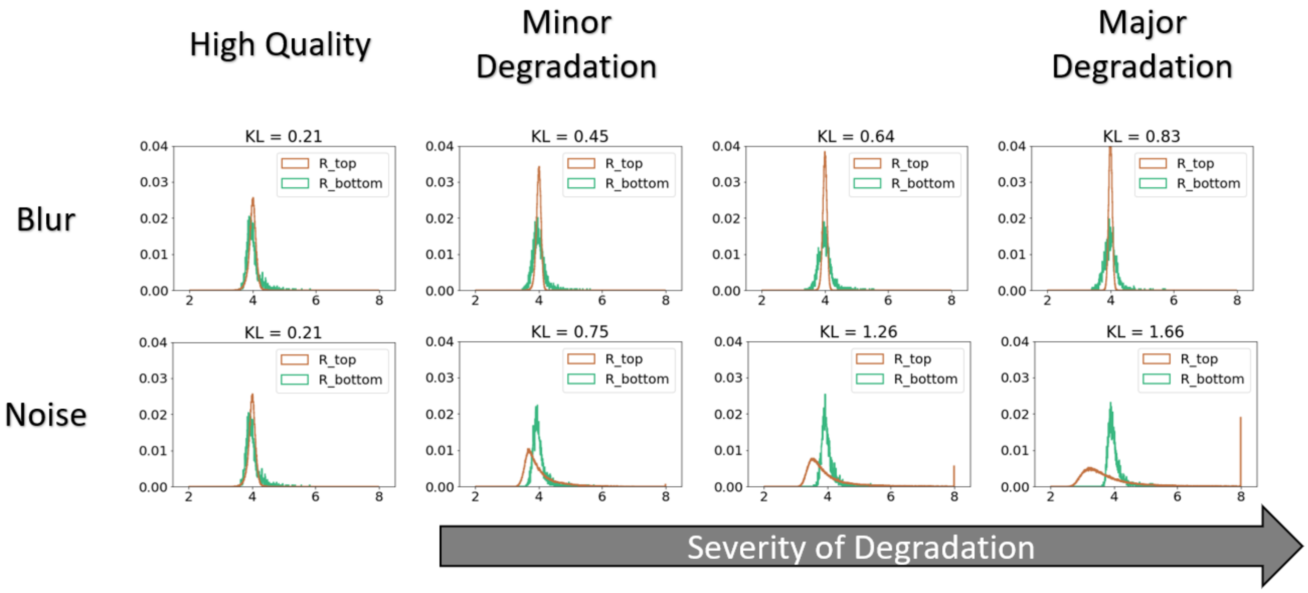

Figure 6: **Relation Patterns.** For a high-quality image, the top and bottom relation histograms are similar, resulting in a low KL divergence. For degraded images, the bottom relation histogram remains stable while the top relation histogram deviates away, increasing the KL divergence.

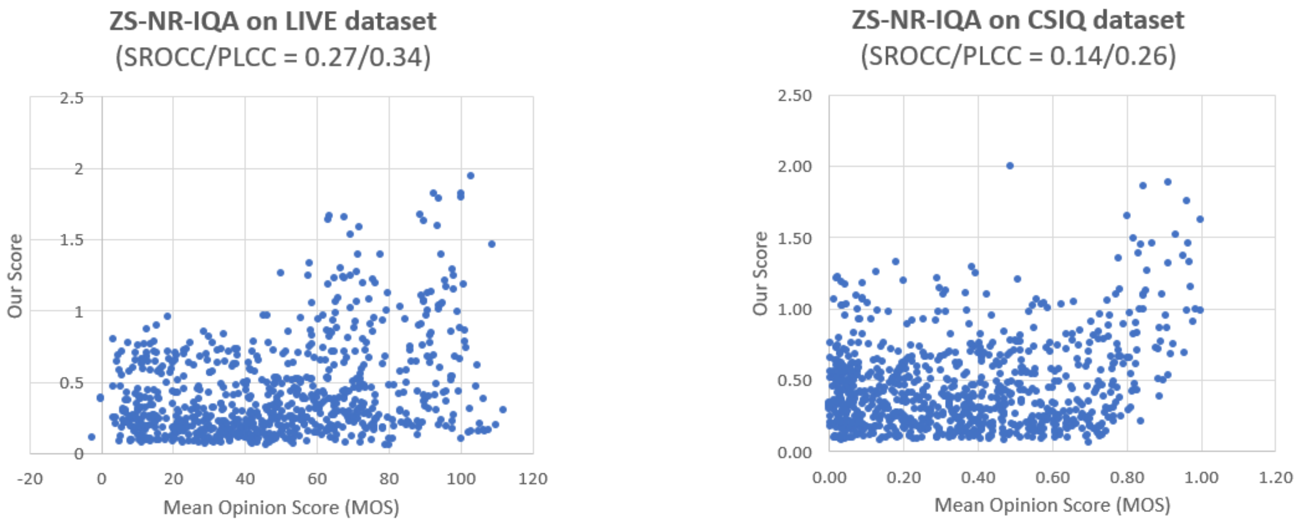

Figure 7: **ZS-NR-IQA on benchmark datasets.** Our method shows a positive correlation on benchmark datasets, some superior to out-of-distribution current State-of-the-Art methods. Nonetheless, our method struggles with some images. See section 4 for further explanations.

scenarios where our method will succeed or encounter challenges.

We believe there is great potential in pursuing the trajectory of zero-shot in the research field of NR-IQA. For example, our algorithm is currently not end-to-end differentiable. Overcoming this constraint could enable many deep learning methods to use the score as part of their training. Normalizing the score ranges of different degradations, and broadening the span of our work to include authentic distortions are more examples of fascinating advancement. We look forward to the forthcoming

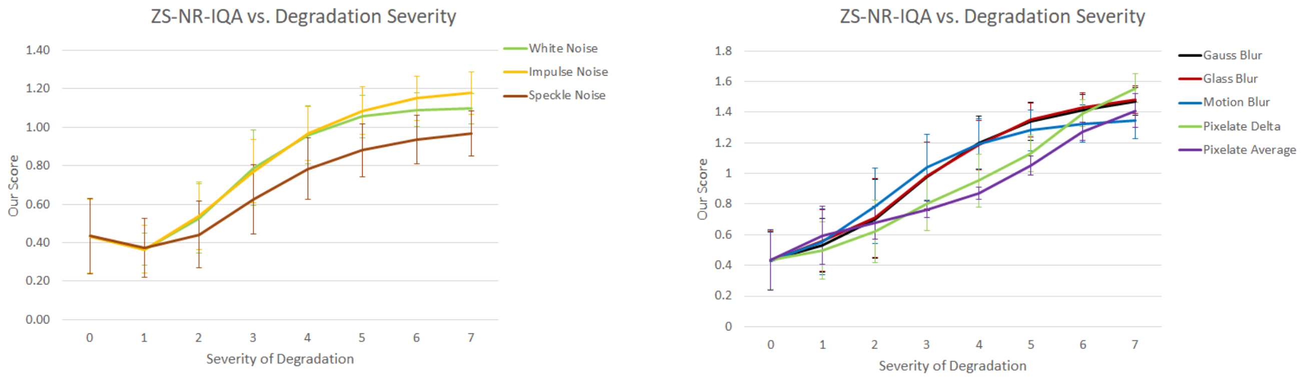

Figure 8: **ZS-NR-IQA on multiply-degraded DIV2K.** The graphs present the mean ZS-NR-IQA score and standard deviation over the images. Our score is highly correlative to the severity of degradation. The trend is clear when experimenting with many images and multiple types of degradations.

strides in this field, as new discoveries and innovations continue to shape the trajectory of our collective knowledge.

References

[1] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 126–135, 2017. 11

[2] Ariel Elnekave and Yair Weiss. Generating natural images with direct patch distributions matching. In *European Conference on Computer Vision*, pages 544–560. Springer, 2022. 9

[3] Daniel Glasner, Shai Bagon, and Michal Irani. Super-resolution from a single image. In *2009 IEEE 12th international conference on computer vision*, pages 349–356. IEEE, 2009. 7

[4] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 5, 6

[5] Dinesh Jayaraman, Anish Mittal, Anush K Moorthy, and Alan C Bovik. Objective quality assessment of multiply distorted images. In *2012 Conference record of the forty sixth asilomar conference on signals, systems and computers (ASILOMAR)*, pages 1693–1697. IEEE, 2012. 3

[6] Jongyoo Kim and Sanghoon Lee. Fully deep blind image quality predictor. *IEEE Journal of selected topics in signal processing*, 11(1):206–220, 2016. 5

[7] Jongyoo Kim, Hui Zeng, Deepti Ghadiyaram, Sanghoon Lee, Lei Zhang, and Alan C Bovik. Deep convolutional neural models for picture-quality prediction: Challenges and solutions to data-driven image quality assessment. *IEEE Signal processing magazine*, 34(6):130–141, 2017. 6

[8] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012. 6

[9] Eric C Larson and Damon M Chandler. Most apparent distortion: full-reference image quality assessment and the role of strategy. *Journal of electronic imaging*, 19(1):011006–011006, 2010. 3

[10] Hanhe Lin, Vlad Hosu, and Dietmar Saupe. Kadid-10k: A large-scale artificially distorted iqa database. In *2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–3. IEEE, 2019. 3

[11] Xialei Liu, Joost Van De Weijer, and Andrew D Bagdanov. Rankiqa: Learning from rankings for no-reference image quality assessment. In *Proceedings of the IEEE international conference on computer vision*, pages 1040–1049, 2017. 6

[12] Or Lotan and Michal Irani. Needle-match: Reliable patch matching under high uncertainty. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 439–448, 2016. 7

[13] Kede Ma, Wentao Liu, Tongliang Liu, Zhou Wang, and Dacheng Tao. dipiq: Blind image quality assessment by learning-to-rank discriminable image pairs. *IEEE Transactions on Image Processing*, 26(8):3951–3964, 2017. 6

[14] Kede Ma, Wentao Liu, Kai Zhang, Zhengfang Duanmu, Zhou Wang, and Wangmeng Zuo. End-to-end blind image quality assessment using deep neural networks. *IEEE Transactions on Image Processing*, 27(3):1202–1213, 2017. 5

[15] Pavan C Madhusudana, Neil Birkbeck, Yilin Wang, Balu Adsumilli, and Alan C Bovik. Image quality assessment using contrastive learning. *IEEE Transactions on Image Processing*, 31:4149–4161, 2022. 4, 5

[16] Tomer Michaeli and Michal Irani. Blind deblurring using internal patch recurrence. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part III 13*, pages 783–798. Springer, 2014. 7

[17] 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. 5

[18] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a “completely blind” image quality analyzer. *IEEE Signal processing letters*, 20(3):209–212, 2012. 5

[19] Anush Krishna Moorthy and Alan Conrad Bovik. Blind image quality assessment: From natural scene statistics to perceptual quality. *IEEE transactions on Image Processing*, 20(12):3350–3364, 2011. 5

[20] Nikolay Ponomarenko, Lina Jin, Oleg Ieremeiev, Vladimir Lukin, Karen Egiazarian, Jaakko Astola, Benoit Vozel, Kacem Chehdi, Marco Carli, Federica Battisti, et al. Image database tid2013: Peculiarities, results and perspectives. *Signal processing: Image communication*, 30:57–77, 2015. 3

[21] Nikolay Ponomarenko, Vladimir Lukin, Alexander Zelensky, Karen Egiazarian, Marco Carli, and Federica Battisti. Tid2008-a database for evaluation of full-reference visual quality assessment metrics. *Advances of modern radioelectronics*, 10(4):30–45, 2009. 3

[22] Michele A Saad, Alan C Bovik, and Christophe Charrier. A dct statistics-based blind image quality index. *IEEE Signal Processing Letters*, 17(6):583–586, 2010. 5

[23] Michele A Saad, Alan C Bovik, and Christophe Charrier. Dct statistics model-based blind image quality assessment. In *2011 18th IEEE International Conference on Image Processing*, pages 3093–3096. IEEE, 2011. 5

[24] Hamid R Sheikh and Alan C Bovik. A visual information fidelity approach to video quality assessment. In *The first international workshop on video processing and quality metrics for consumer electronics*, volume 7, pages 2117–2128. sn, 2005. 5

[25] 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. 3

[26] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 5, 6

[27] 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. 4, 6

[28] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 3, 5

[29] Zhou Wang, Eero P Simoncelli, and Alan C Bovik. Multiscale structural similarity for image quality assessment. In *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, volume 2, pages 1398–1402. Ieee, 2003. 5

[30] 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. 5

[31] 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

[32] 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. 5, 6

[33] Maria Zontak, Inbar Mosseri, and Michal Irani. Separating signal from noise using patch recurrence across scales. In *proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1195–1202, 2013. 7