BURST PERCEPTION-DISTORTION TRADEOFF: ANALYSIS AND EVALUATION

Danna Xue[⋆†‡] *Luis Herranz*[†‡] *Javier Vazquez Corral*[†‡] *Yanning Zhang*[⋆]

[⋆] School of Computer Science, Northwestern Polytechnical University, Xi'an, China
[†] Computer Vision Center, Barcelona, Spain
[‡] Department of Computer Science, Universitat Autònoma de Barcelona, Barcelona, Spain

ABSTRACT

Burst image restoration attempts to effectively utilize the complementary cues appearing in sequential images to produce a high-quality image. Most current methods use all the available images to obtain the reconstructed image. However, using more images for burst restoration is not always the best option regarding reconstruction quality and efficiency, as the images acquired by handheld imaging devices suffer from degradation and misalignment caused by the camera noise and shake. In this paper, we extend the perception-distortion tradeoff theory by introducing multiple-frame information. We propose the area of the unattainable region as a new metric for perception-distortion tradeoff evaluation and comparison. Based on this metric, we analyse the performance of burst restoration from the perspective of the perception-distortion tradeoff under both aligned bursts and misaligned bursts situations. Our analysis reveals the importance of inter-frame alignment for burst restoration and shows that the optimal burst length for the restoration model depends both on the degree of degradation and misalignment.

Index Terms— burst image restoration, perception-distortion tradeoff, inter-frame alignment

1. INTRODUCTION

Image restoration aims, given an image that has experimented some degradation process, to restore it to obtain the original image. This problem has been widely studied in the literature for many years and relates to several sub-problems, such as denoising, super-resolution, deblurring, *etc.*

Recently, the trend has moved towards burst image restoration. Burst is a sequence of images captured in rapid succession. The main reason for this shift is the current ubiquity of smartphones, since these devices can easily acquire this sequential data and process it to produce better-quality images. Burst image restoration has the advantage that multiple frames provide complementary information to the reference one, leading to higher resolution [1, 2, 3, 4, 5, 6], lower noise level [7, 5], and higher dynamic range [8], while also introducing uncertainty caused by motion or camera shake [9]. This misalignment problem introduced by multiple images may lead to restored images with ghost artefacts, and blurry [10]. Recent works [3, 5] explicitly align burst images by estimating optical flows [11], or implicitly by deformable convolutions [2, 4, 6]. In practice, even after these alignment methods, two images are rarely perfectly aligned due to the degradation and the appearance of artefacts.

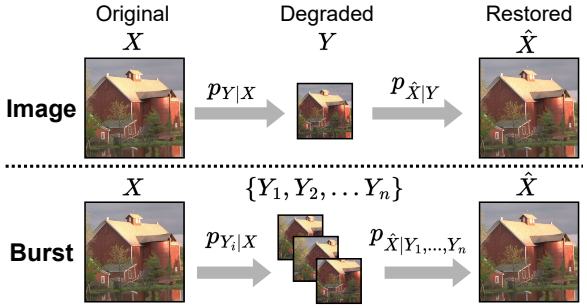

Fig. 1: **Single Image Restoration** versus **Burst Restoration**. Burst introduces the problem of misalignment between images in the burst. See details in Section 2.

The evaluation of image restoration is generally carried out from two aspects: the perceptual quality and the distortion. Blau *et al.* [12] first characterized the Perception-Distortion (P-D) Tradeoff in single image restoration. More specifically, they proved that distortion and perceptual quality are at odds with each other so that no image restoration algorithm can optimize the two indicators to the best at the same time in practice. The P-D curve comprehensively shows the upper bound and range of continuous changes of two types of evaluation criteria. The generative-adversarial-nets (GANs) provide a principled way to approach the P-D bound by varying the hyperparameter between distortion loss and per-

This paper was supported by Grant PID2021-128178OB-I00 funded by MCIN/AEI/10.13039/501100011033, ERDF "A way of making Europe", the Departament de Recerca i Universitats from Generalitat de Catalunya with reference 2021SGR01499, the "Ayudas para la recualificación del sistema universitario español" financed by the European Union-NextGenerationEU (JVC), and Ramon y Cajal grant RYC2019-027020-I (LH), by National Natural Science Foundation of China under Grant No.U19B2037, and by Natural Science Basic Research Program of Shaanxi Province Program No.2021JCW-03 (YZ).

ception loss, thus producing estimators along the P-D curve, and therefore obtaining the P-D tradeoff curve. In [13, 14] authors prove that the P-D curve can be acquired by linear interpolation between two models, which greatly simplifies the steps to obtain the P-D curve. Similar tradeoffs are also proved existing in classification [15] and image compression [16]. However, these works only focus on single-image tasks, and the perception-distortion tradeoff in the case of multiple images has not been studied yet.

Our work studies the perception-distortion tradeoff from multiple images, thus generalizing the case of a single image. In particular, we focus on the case of burst image restoration with relatively stable noise and camera shaking. Through the analysis, we found that using more images does not always lead to better reconstruction quality due to the misalignment between each image. The optimal burst length (*i.e.* the number of images in a burst) for restoration depends on the shake and noise levels.

In summary, the contribution of this paper is threefold:

- We propose the Burst Perception-Distortion Tradeoff by introducing multiple-frame information.
- We propose AUR as a new method for multi-frame restoration evaluation, which comprehensively reflects the perception and distortion quality of the restored image.
- We analyse the Burst P-D tradeoff under the influence of image noise and shake, and found the effect of inter-frame misalignment on burst restoration.

2. THE PERCEPTION-DISTORTION TRADEOFF

2.1. Single image perception-distortion tradeoff

The original P-D tradeoff formulation considers the case of a single degraded image y, which is observed according to some conditional distribution $p_{Y|X}$, where $x \sim p_X$ would be the underlying true original image. This formulation assumes that the degradation is not reversible, *i.e.* cannot be estimated from y without error, which is typically the case in image restoration. Thus, given the degraded image y, a restored image $\hat{x}$ is estimated according to the conditional distribution $p_{\hat{X}|Y}$. The problem setting is described in Fig. 1 (top).

Two performance metrics are defined: *distortion* $E[\Delta(\hat{x}, x)]$ that measures how similar the restored image is to the actual original image, and *perception* (*i.e.* perceptual quality) $d(p_{\hat{X}}, p_X)$ that measures the divergence between the distribution of reconstructed images $p_{\hat{X}}$ and the distribution of natural images p_X. The perception-distortion function of the restoration task is given by

$$P(D) = \min_{p_{\hat{X}|Y}} d(p_X, p_{\hat{X}}) \quad \text{s.t. } E[\Delta(X, \hat{X})] \leq D. \quad (1)$$

The main finding in this formulation is that the region under the P-D function is not attainable, and the P-D function represents points where an improvement of one metric implies a worsening of the other.

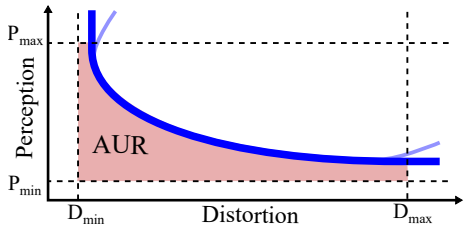

Fig. 2: Illustration of the Area of the Unattainable Region (AUR) within the ranges $[P_{min}, P_{max}]$ and $[D_{min}, D_{max}]$. The P-D curve is extended when derivative is 0 or inf (from the light blue curve to dark blue curve) to avoid the ill effects caused by model training.

2.2. Burst perception-distortion tradeoff

In our case, we generalize the previous formulation to the case in which a burst of n degraded images $\{y_1, y_2, ..., y_n\}$ is observed from the same underlying image x, each y_i being a sample from the distribution $p_{Y_i|X} = p_{Y|X}$, since we assume them independent and identically distributed. Critically for our analysis, there exists camera shake that may result in small misalignments between the images. Then, given the sequence of degraded images $\{y_1, y_2, ..., y_n\}$, a restoration algorithm estimates a restored image $\hat{x}$ according to the conditional distribution $p_{\hat{X}|Y_1,Y_2,...,Y_n}$ (see Fig. 1 (bottom)). The burst perception-distortion function is thus defined as

$$\begin{aligned} &P(D) = \min_{p_{\hat{X}|Y_1,Y_2,...,Y_n}} d(p_X, p_{\hat{X}}) \\ &\text{s.t. } E[\Delta(X, \hat{X})] \leq D. \end{aligned} \quad (2)$$

Note that this formulation generalizes the single image P-D function and introduces the new misalignment problem between images in the burst.

2.3. Area of the unattainable region

While the P-D plane and P-D functions are the main tools to compare the performance of restoration algorithms, we propose the *area of the unattainable region* (AUR), that is, the area under the P-D function as metric for more convenient comparison (see Fig. 2). This metric summarizes the performance in one single value. While AUR can be applied to the single image case, it is particularly convenient to study the influence of factors, such as burst length, in the burst case.

Since the AUR could be infinite, we define it within a range of perception and distortion values of interest $[P_{min}, P_{max}]$ and $[D_{min}, D_{max}]$, respectively, as

$$\text{AUR} = \int_{D_{min}}^{D_{max}} \hat{P}(D) dx \quad (3)$$

where $\hat{P}(D)$ is $P(D)$ clamped to the range $[P_{\min}, P_{\max}]$.

3. EXPERIMENTS

For the experiments, we focus on the burst super-resolution task, which is a common and representative problem with three degradation factors: noise, (camera) shake and downsampling. In this case, the i^{th} observed image in a burst with n images is related to the (unknown) original image via the following relation

$$y_i\left[\mathbf{u}\right] = x\left[\alpha\mathbf{u} + \nu_i\right] + \epsilon_i\left[\mathbf{u}\right], \tag{4}$$

where $\mathbf{u}$ represents the coordinates in the low resolution grid, in contrast to the high resolution grid $\mathbf{u}'$ in which $x\left[\mathbf{u}'\right]$ is represented. α is the subsampling factor, v_i represents the displacement due to camera shake, and ϵ_i represents the camera noise. We assume they are independent and identically distributed. The single image case corresponds to $n = 1$, which implies no misalignment, *i.e.* $y\left[\mathbf{u}\right] = x\left[\alpha\mathbf{u}\right] + \epsilon\left[\mathbf{u}\right]$.

Dataset. Describable Textures Dataset (DTD) [17] is a natural texture database consisting of 5640 images with 47 categories (120 images for each). Image sizes range between 300×300 and 640×640. The data is split into three equal parts for training, validation, and testing, with 40 images per class, for each split.

We generate a synthetic burst super-resolution dataset based on the DTD dataset. Each image is centre-cropped to 128×128 to get the high-resolution ground truth (HR), and the LR image is obtained by bilinear interpolation with a scaling factor of $\times 4$. Following the burst synthesizing process provided by [9], in each burst, we randomly add Poisson noise (shot noise) $n_p \sim P(\lambda_p)$ and Gaussian noise (readout noise) $n_g \sim N(0, \sigma_g)$ to each image. The first image in each burst is the reference frame aligned with HR. For the rest images in the burst, we add random translation on both vertical and horizontal axis $\Delta x_s \sim N(0, \sigma_s), \Delta y_s \sim N(0, \sigma_s)$.

Training details. We look at burst super-resolution methods to analyse the quality of the reconstructed image. In order to navigate the Perception-Distortion Tradeoff, we consider the ESRGAN [18] network trained with two stages, where the first stage is distortion-oriented and the second is perception-oriented. We linear interpolate the parameters of these two models by $\theta^{interp} = (1-\alpha)\theta^D + \alpha\theta^P$ to obtain a continuous P-D curve. We repeat this training for each different noise and shake levels given in Table 1.

More specifically, we first train the distortion-oriented model only with L1 loss. The learning rate is initialized as 2×10^{-4} and decayed by a factor of 2 every 50 epochs. Then this model is employed as initialization for the generator. We fine-tune the generator with adversarial loss and perceptual loss to optimize the perceptual quality. The learning rate is set to 1×10^{-4} and halved at every 25 epochs. Our model contains 23 residual blocks, and all the images in a burst are concatenated as input. We optimize using Adam with β_1 of 0.9 and β_2 of 0.99, batch size of 16. We train and test the models using PyTorch on an NVIDIA GeForce 3090Ti GPU.

Item	Value
Gaussian Noise (σ_g)	$[0, 10, 20, 30, 40]$
Poisson Noise (λ_p)	$[0, 1, 2, 3, 4]$
Shake (σ_s)	$[0, 1, 2, 3, 4]$ (pixel)
Burst Length (n)	$[1, 5, 10, 20, 30, 40, 50]$

Table 1: Experimental settings for degraded burst images. For $\sigma_s = 0$, only the single-frame models are trained.

4. ANALYSIS AND EVALUATION

Bursts are generally collected in a very short period. Therefore, the content and imaging conditions of a burst are basically the same. The main differences between each frame are degrees of noise and misalignment. Since image noise is unavoidable in the imaging process, we analyse two common cases: when all images in a burst are perfectly aligned and when the images are not aligned.

For evaluation, we measure perceptual quality using the no-reference metrics NIQE [19] and BRISQUE [20], and for distortion, we measure RMSE. We calculate AUR of NIQE-RMSE and BRISQUE-RMSE both with $[D_{min}, D_{max}] = [0, 0.3]$, $[P_{min}, P_{max}] = [0, 150]$.

4.1. Perfectly Aligned Bursts

This setting covers two cases: (1) *The burst is captured in a stable condition, i.e. there is no shake or motion during imaging.* (2) *The accurate motion parameters or flow between frames can be measured by equipment or estimated by algorithms.* In this case, we only consider the impact of noise on the quality of restoration results.

Foremost, the P-D curves in Fig. 3 prove that the P-D tradeoff still exists in burst restoration. As the noise level of the input image increases, for both Gaussian noise and Poisson noise, the P-D curve lies further from the origin, which indicates that both the perceptual quality of the restoration image and distortion are getting worse (See Fig. 3 column 1-2). At a certain noise level, perception and distortion improve as the input burst length increases. When the burst length reaches a certain number, the benefit of using more images for processing decreases, since most information has already been restored. As illustrated in Fig. 3 column 3, the two-frame P-D curve shows a wide margin over the single-frame curve, but the gap between 10 and 100 images is quite narrow. The AUR value (see Fig. 3 column 4) also indicates the same tendency, proving the AUR curve captures well the P-D plane. The analysis of a perfectly aligned burst proves the importance of alignment for multi-frame processing. When a burst is well aligned, the more images the input has, the higher the image quality obtained for both perception and fidelity.

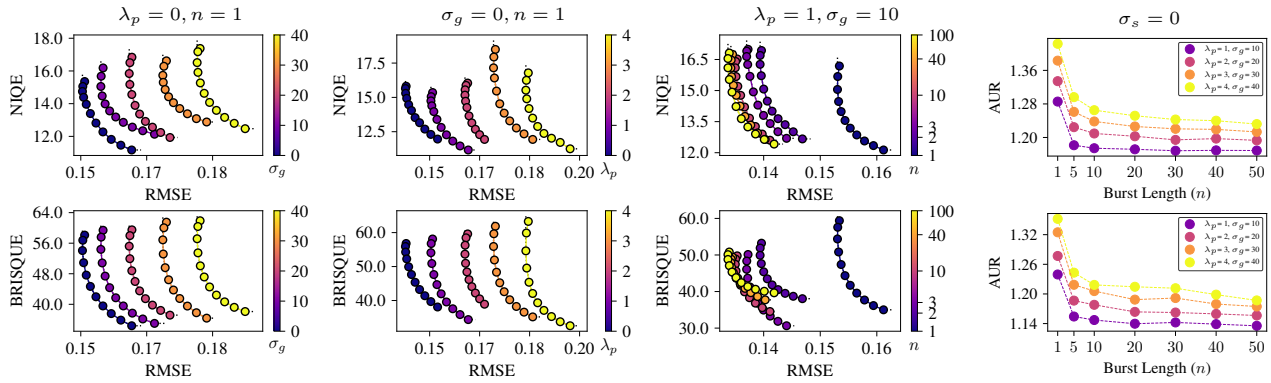

Fig. 3: P-D curves of perfectly aligned bursts. Columns 1-3 compare the P-D curves with different levels of Gaussian noise, Poisson noise, and burst length, respectively. Column 4 shows the AUR. When images are perfectly aligned, and the noise level in each image is lower than the signal itself, using more images for burst restoration leads to better restoration quality. Note that the black dash line in the P-D planes indicates the modified P-D curves.

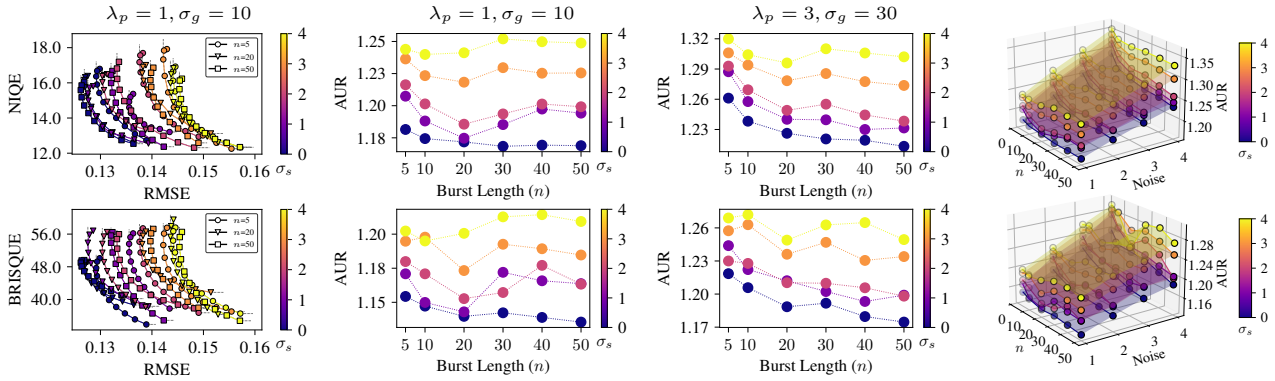

Fig. 4: P-D curves of misaligned bursts. Column 1 compares the P-D curves with different levels of shake. Column 2,3,4 shows the AUR under different levels of noise and shake. When the burst is imperfectly aligned, an optimal burst length for restoration exists, depending on the noise level and displacement level.

4.2. Misaligned Bursts

For bursts taken by a handheld camera, shake is almost an unavoidable problem. Misaligned bursts result from two possible conditions: (1) *Direct restoration without any alignment.* (2) *Misalignment resulting from inaccurate motion or flow estimation*. In our case, we consider the impact of both shake and noise. Here we assume that the entire burst is acquired in a very short period, so only the random shake is considered. Let us also note that this case can also be understood as a proxy for the error of alignment methods.

As shown in Fig. 4, as the burst length increases, the restoration results gradually get better at first, and after reaching the optimum quality at a certain burst length, the image quality gradually gets worse. When the displacement between images is relatively small, complementary information between different images helps recover more image details. However, when the displacement between images is too large, using more images to restore will worsen the quality. As illustrated in Fig. 4 column 2, when $\lambda_p = 1, \sigma_g = 10$, *i.e.* the AUR curves for the first image column, 20 is the optimal length for burst with shake $\sigma_s = 1, 2, 3$. As the noise level increases (column 3, $\lambda_p = 3, \sigma_g = 30$), the larger the length of the input burst is, the better the quality of the restored image is. Therefore, the optimal burst length is determined by both the shake and the noise level.

5. CONCLUSION

In this paper, we extend the theory of perception-distortion tradeoff to multiple images, in particular to bursts. We analyse the impact of noise and shake on multi-frame restoration from the perspective of the P-D tradeoff and determine the importance of inter-frame alignment for burst restoration. On this basis, we propose a new metric for evaluating and comparing P-D curves. We believe our work provides reference for the design of multi-frame restoration methods. In the case of estimable image noise and camera shake, the analysis results can also be used as a reference for selecting the optimal burst length.

6. REFERENCES

[1] Bartlomiej Wronski, Ignacio Garcia-Dorado, Manfred Ernst, Damien Kelly, Michael Krainin, Chia-Kai Liang, Marc Levoy, and Peyman Milanfar, "Handheld multi-frame super-resolution," *ACM Transactions on Graphics (TOG)*, vol. 38, no. 4, pp. 1–18, 2019.

[2] Akshay Dudhane, Syed Waqas Zamir, Salman Khan, Fahad Shahbaz Khan, and Ming-Hsuan Yang, "Burst image restoration and enhancement," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 5759–5768.

[3] Goutam Bhat, Martin Danelljan, Luc Van Gool, and Radu Timofte, "Deep burst super-resolution," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 9209–9218.

[4] Ziwei Luo, Lei Yu, Xuan Mo, Youwei Li, Lanpeng Jia, Haoqiang Fan, Jian Sun, and Shuaicheng Liu, "Ebsr: Feature enhanced burst super-resolution with deformable alignment," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 471–478.

[5] Goutam Bhat, Martin Danelljan, Fisher Yu, Luc Van Gool, and Radu Timofte, "Deep reparametrization of multi-frame super-resolution and denoising," in *IEEE/CVF International Conference on Computer Vision*, 2021, pp. 2460–2470.

[6] Ziwei Luo, Youwei Li, Shen Cheng, Lei Yu, Qi Wu, Zhihong Wen, Haoqiang Fan, Jian Sun, and Shuaicheng Liu, "Bsrt: Improving burst super-resolution with swin transformer and flow-guided deformable alignment," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 998–1008.

[7] Ben Mildenhall, Jonathan T Barron, Jiawen Chen, Dillon Sharlet, Ren Ng, and Robert Carroll, "Burst denoising with kernel prediction networks," in *IEEE conference on computer vision and pattern recognition*, 2018, pp. 2502–2510.

[8] Samuel W Hasinoff, Dillon Sharlet, Ryan Geiss, Andrew Adams, Jonathan T Barron, Florian Kainz, Jiawen Chen, and Marc Levoy, "Burst photography for high dynamic range and low-light imaging on mobile cameras," *ACM Transactions on Graphics (ToG)*, vol. 35, no. 6, pp. 1–12, 2016.

[9] Goutam Bhat, Martin Danelljan, Radu Timofte, Yizhen Cao, Yuntian Cao, Meiya Chen, Xihao Chen, Shen Cheng, Akshay Dudhane, Haoqiang Fan, et al., "Ntire 2022 burst super-resolution challenge," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2022, pp. 1041–1061.

[10] Qingsen Yan, Dong Gong, Qinfeng Shi, Anton van den Hengel, Chunhua Shen, Ian Reid, and Yanning Zhang, "Attention-guided network for ghost-free high dynamic range imaging," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 1751–1760.

[11] Deqing Sun, Xiaodong Yang, Ming-Yu Liu, and Jan Kautz, "Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume," in *IEEE conference on computer vision and pattern recognition*, 2018, pp. 8934–8943.

[12] Yochai Blau and Tomer Michaeli, "The perception-distortion tradeoff," in *IEEE conference on computer vision and pattern recognition*, 2018, pp. 6228–6237.

[13] Dror Freirich, Tomer Michaeli, and Ron Meir, "A theory of the distortion-perception tradeoff in wasserstein space," *Advances in Neural Information Processing Systems*, vol. 34, pp. 25661–25672, 2021.

[14] Zeyu Yan, Fei Wen, and Peilin Liu, "Optimally controllable perceptual lossy compression," *Proceedings of Machine Learning Research*, 2022.

[15] Dong Liu, Haochen Zhang, and Zhiwei Xiong, "On the classification-distortion-perception tradeoff," *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[16] Yochai Blau and Tomer Michaeli, "Rethinking lossy compression: The rate-distortion-perception tradeoff," in *International Conference on Machine Learning*. PMLR, 2019, pp. 675–685.

[17] M. Cimpoi, S. Maji, I. Kokkinos, S. Mohamed, , and A. Vedaldi, "Describing textures in the wild," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2014.

[18] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy, "Esrgan: Enhanced super-resolution generative adversarial networks," in *European conference on computer vision (ECCV) workshops*, 2018, pp. 0–0.

[19] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal processing letters*, vol. 20, no. 3, pp. 209–212, 2012.

[20] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Transactions on image processing*, vol. 21, no. 12, pp. 4695–4708, 2012.