Image robustness to adversarial attacks on no-reference image-quality metrics

Daniil Konstantinov*, Sergey Lavrushkin†‡, Dmitriy Vatolin†‡§
*Moscow Institute of Physics and Technology, Moscow, Russia
†MSU Institute for Artificial Intelligence, Moscow, Russia
‡ISP RAS Research Center for Trusted Artificial Intelligence, Moscow, Russia
§Lomonosov Moscow State University, Moscow, Russia
Emails: dkonstantinov0@gmail.com, sergey.lavrushkin@graphics.cs.msu.ru, dmitriy@graphics.cs.msu.ru

***Abstract*—Quality assessment is an essential part of image and video processing, such as super-resolution, compression, content generation, and similar tasks. Most modern quality metrics are learning-based methods, and although they exhibit a higher correlation with subjective scores than traditional methods do, they are vulnerable to adversarial attacks. Today, no-reference image-quality assessment (NR-IQA) metrics are becoming more popular, as they can assess images and videos without any additional information. This paper presents results of conducting various adversarial attacks on different NR-IQA metrics and introduces an Image Robustness to Adversarial Attacks model that estimates an image's vulnerability to attacks. Our analysis of adversarial attacks on NR-IQA metrics revealed an image class that is robust to these attacks and, conversely, an image class that is vulnerable to most of them. We analyzed several image datasets and found that distortions such as denoising and various types of blur reduce an image's robustness to adversarial attacks.**

***Index Terms*—quality assessment, image robustness, adversarial attacks, no-reference metrics**

I. Introduction

Image-quality assessment (IQA) plays a crucial role in computer vision applications, such as developing and comparing image-compression algorithms [1], super-resolution [2], and generation [3]. Most new IQA methods have neural-network-based architectures, as deep-learning methods outperform traditional approaches and better correlate with subjective scores [4]. The main issue with these new approaches is their vulnerability to adversarial attacks, which improve metric scores without a corresponding image-quality enhancement [5]. Developers of image-processing methods can thus exploit metric vulnerabilities to achieve better outcomes relative to competitors.

No-reference (NR) metrics are becoming more popular in real-world scenarios, as access to a pristine reference image may be impractical or impossible, but they typically correlate less well with subjective quality relative to full-reference (FR) and reduced-reference metrics. However, recent findings indicate that new NR metrics outperform many existing FR ones, so we chose NR metrics for this work.

Authors of [5] noticed that adversarial attacks on metrics behave differently depending on the image's content and distortions. Fig. 1 shows two instances of this behavior. We believe it is crucial to analyze image stability against adversarial attacks, as doing so will allow construction of datasets to develop new metrics that are resilient to attacks on unstable images while also guiding development of new attacks that work on stable ones.

(a) Weak attack (b) Strong attack

Fig. 1: Results of applying four adversarial attacks to the NR metric SPAQ on two different images. (a) shows insignificant metric-score increases. (b) shows drastic metric-score changes. The left parts are from the original images, the right parts — after the UAP attack.

We define an image as unstable if, for a specific metric, its score generally increases following application of most adversarial attacks. On the other hand, images are stable when their scores remain unchanged or decrease after the attacks. Evaluating image robustness for a particular metric involves applying numerous attacks, often a computationally demanding task. To address this challenge, we propose a model for analyzing image robustness to adversarial attacks. We tested several commonly-used public models and analyzed how distortions affect image robustness.

Our code is available at https://github.com/CI314X/iraa.

II. Related work

Deep learning models are vulnerable to adversarial examples: inputs deliberately crafted to appear normal to humans but having the potential to mislead and manipulate a machine learning model's predictions. The initial efforts to deceive machine learning models primarily focused on classification. Firstly, a box-constrained L-BFGS method [6] was employed to uncover the prevalence of adversarial examples in deep neural network-based classifiers. Next, the Fast Gradient Sign Method (FGSM) [7] was introduced — a simple yet effective

technique that perturbs the input image by adding a small amount of noise. Further advancements include an extension called iterative FGSM (IFGSM) [8], which uses an iterative approach with a small step size to generate adversarial examples. Moreover, the authors of [9] enhanced IFGSM attacks by incorporating momentum. In [10] adversarial training is conceptualized in the framework of robust optimization, casting adversarial-example generation as the solution to an inner-maximization problem through projected gradient descent. Most techniques discussed in the literature about adversarial-example generation fall into the category of white-box attacks, where the attacker possesses information about the trained neural network model, including its architecture and parameters. By contrast, a black-box attacker was introduced in [11].

In regression tasks, the absence of natural margins (like in classification) creates difficulties for adversarial learning. The definition of adversarial attacks, measurement of their success, and establishment of evaluation metrics become complicated in the context of regression. Nevertheless, for our specific task, we can employ modified classifier attacks. We focus on adversarial attacks that artificially increase estimated quality scores, as such attacks already occur in many real-life scenarios. A pioneering work in this area [12] proposes an iterative attack on NR metrics, incorporating the full-reference metrics into the loss function to regulate the attack quality. Also, the authors noticed that poor-quality or blurred images are more likely to experience a notable rise in metric scores following an adversarial attack and, conversely, that high-quality images demonstrate greater stability to such attacks. The authors of [13] applied the concept of Universal Adversarial Perturbation (UAP) to the regression problem, attacking differentiable NR metrics. In [14] they introduced a U-Net network that generates adversarial noise for each image. Note that although metric robustness to adversarial attacks is already under investigation, the influence of image content on susceptibility to these attacks has yet to be studied, so in this work we introduced this problem.

III. Proposed model

A. Image Robustness Index

To analyze the stability of an image for adversarial attack $g_j : \mathbb{R}^{H\times W\times 3} \to \mathbb{R}^{H\times W\times 3}$ and metric $f_i : \mathbb{R}^{H\times W\times 3} \to \mathbb{R}$ we consider a normalized relative difference between a metric score of the original image $\mathbf{x}$ and of the attacked image $\mathbf{x}^* = g_j(\mathbf{x})$, which we formulate as:

$$\Delta_{ij}(\mathbf{x}) = \frac{\hat{f}_i(g_j(\mathbf{x})) - \hat{f}_i(\mathbf{x})}{\hat{f}_i(\mathbf{x}) + \varepsilon}, \tag{1}$$

where $\hat{f}_i(\mathbf{x}) \in [0,1]$ is a normalized metric score and $\varepsilon = 10^{-5}$ is a constant to avoid division by zero. We calculate $\hat{f}_i(\mathbf{x})$ as follows:

$$\hat{f}_i(\mathbf{x}) = \frac{f_i(\mathbf{x}) - \min_{\hat{\mathbf{x}}\in X} f_i(\hat{\mathbf{x}})}{\max_{\hat{\mathbf{x}}\in X} f_i(\hat{\mathbf{x}}) - \min_{\hat{\mathbf{x}}\in X} f_i(\hat{\mathbf{x}})}, \tag{2}$$

where X is the set of images.

We categorize all normalized scores $\Delta_{ij}(\mathbf{x})$ of an image into four groups (3). Initially, negative scores have the designation $R_{ij}(\mathbf{x}) = 0$, indicating that the attack g_j failed to enhance the score of f_i for image $\mathbf{x}$. We then construct a frequency distribution for the remaining scores $\Delta_{ij}(\mathbf{x}) > 0$ and further divide these scores into three groups based on the basis of 25% and 75% quantiles of the resulting distribution: a strong group $(q_0^{ij}, q_{0.25}^{ij})$, a medium group $(q_{0.25}^{ij}, q_{0.75}^{ij})$, and a weak group $(q_{0.75}^{ij}, q_1^{ij})$.

$$R_{ij}(\mathbf{x}) = \begin{cases} 0, & \Delta_{ij}(\mathbf{x}) \le 0 \\ 1, & q_0^{ij} < \Delta_{ij}(\mathbf{x}) \le q_{0.25}^{ij} \\ 2, & q_{0.25}^{ij} < \Delta_{ij}(\mathbf{x}) \le q_{0.75}^{ij} \\ 3, & q_{0.75}^{ij} < \Delta_{ij}(\mathbf{x}) \le q_1^{ij} \end{cases} \tag{3}$$

As our primary focus is on identifying images that are either highly robust or highly susceptible to adversarial attacks, we chose unequal group sizes to separate these cases from those with medium attack results.

The Image Robustness Index (IRI) derives from an average over all attacks and metrics:

$$\mathrm{IRI}_{n,m}(\mathbf{x}) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m} R_{ij}(\mathbf{x})}{n\,m} \in [0,3], \tag{4}$$

where n is a number of metrics and m is a number of adversarial attacks. Hence, the higher this value is for an image, the more vulnerable it is to adversarial attacks.

B. Image robustness dataset

Gathering data to train a model to predict image robustness requires executing a series of adversarial attacks on a set of metrics. Any image dataset can serve this purpose: we selected the Microsoft Common Objects in COntext (MS COCO) dataset [15] for to its substantial size and content diversity. Our objective was to select contemporary metrics that strongly correlate with subjective scores and encompass diverse architectural characteristics. We therefore picked five NR-IQA metrics: MDTVSFA [16], LINEARITY [17], KonCept512 [18], SPAQ [19] and PaQ-2-PiQ [20]. For these metrics, we applied four adversarial attacks: FGSM [7], I-FGSM [8], UAP [13] and FACPA [14]. We used public source code for all NR-IQA metrics without additional pre-training, and we selected default parameters. We determined the amplitudes of the maximum change that an adversarial attack can induce for each metric to ensure that the score distributions for different attacks aligned as closely as possible.

Overall, we obtained 20 scores ($n = 5$, $m = 4$) for each image in MS COCO dataset. The training part comprises 118k images, the validation part comprises 5k images, and the test part comprises 40k images.

We also explored the convergence of the IRI depending on the number of attacks over which it is averaged. To do so, we calculated the Pearson correlation coefficient (PCC) between $\mathrm{IRI}_{n,m}$ and the final target, $\mathrm{IRI}_{5,4}$. Considering attacks on different metrics as distinct (the number of attacks $= nm$), we identified a limit on the number of attacks beyond which

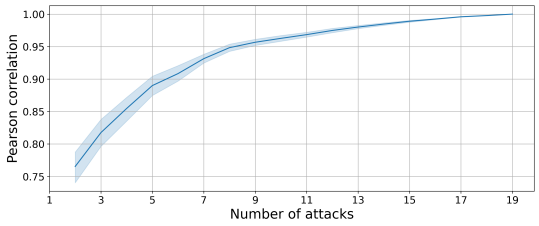

Fig. 2: PCC between $\text{IRI}_{n,m}$ and $\text{IRI}_{5,4}$ (30 shuffles of the attack sequences).

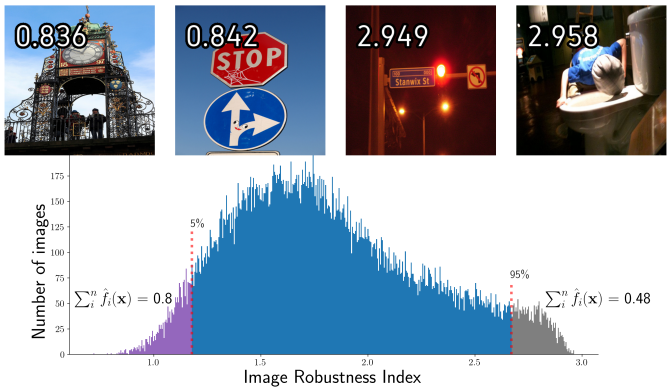

Fig. 3: Distribution of the Image Robustness Index on the test set of MS COCO. The left two images are the most stable, while the right two images are the most vulnerable. Each image includes its corresponding target value in the upper-left corner. The means of normalized scores for the 5th and 95th percentiles are provided.

the target's distribution remains practically unchanged. Acknowledging the significance of the order of including attacks in the $\text{IRI}_{n,m}$, we conducted 30 (as this number increases, the confidence interval will narrows) random shuffles of the order in which attacks are taken into the calculation of $\text{IRI}_{n,m}$. Fig. 2 shows the result of this procedure. So, after using nine attacks, the PCC with the final target exceeded 0.95, indicating we incorporated a sufficient number of them.

Fig. 3 shows the images from the IRI distribution tails. The first two images are the most stable examples, while the last two images are the most unstable ones. Visually, unstable images appear darker and blurrier, and they exhibit lower quality metrics, whereas stable images are clear, bright, and higher in quality. Therefore, a certain correlation is observed: the lower the image's visual quality, the more vulnerable it is to attack. We described this in more detail in Sec. IV-A.

C. Image Robustness to Adversarial Attacks model

We want to train a small and efficient model that lets us quickly and accurately predict the IRI. Therefore, we designed the Image Robustness to Adversarial Attacks (IRAA) model using several convolutional blocks. Fig. 4 shows the network's architecture. The IRAA model's backbone comprises three sequentially connected ConvBlocks, shown on the right in Fig. 4. "ConvBlock, X" consists of two convolutional layers with X out channels, two ReLU activations with a 3×3 filter

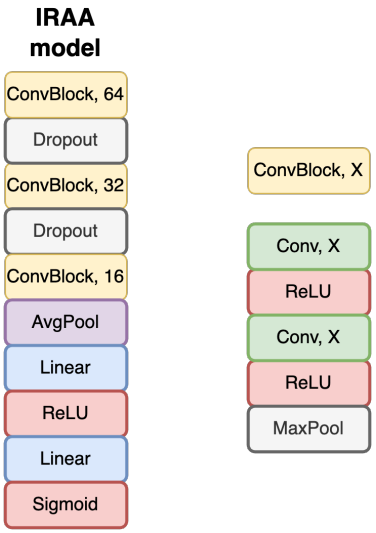

Fig. 4: The architecture of the IRAA model.

and one maximum pooling (MaxPool) layer with a 3×3 kernel and stride of 3. To mitigate overfitting, we employed a dropout layer with $p = 0.1$ between ConvBlocks. The network's final layers include a global average pooling (AvgPool) layer with a 3×3 kernel and two fully connected (Linear) layers with 144 and 32 neurons, respectively, ending with a sigmoid activation function. We multiply the network's output by 3 to adjust the target distribution (4).

D. Training details

We conducted all experiments in Python using the PyTorch framework. Our selection of the best results during neural network training was based on the validation set. We used the AdamW optimizer with default settings and a learning rate of 0.0003.

All model training employed input images of original size, only padding it with zeroes if necessary to ensure the models can accommodate different resolutions.

IV. Experiments

In this section, we evaluate the accuracy of our proposed method, investigate the relationship between image robustness and visual quality, and assess the robustness of images from various datasets.

A. PCC between IRI and image quality before attack

To assess our model's performance we employed two evaluation metrics: mean squared error (MSE) and PCC. We chose PCC to gauge the relative linear relationships between IRI and predicted values.

If we assume that lower image quality correlates with higher susceptibility to adversarial attack, we can evaluate image robustness using the metric value itself. As IRI demonstrates, images with various distortions are generally less stable than those without. To explore this concept, we experimented on the training part of our dataset, assessing the correlation between the mean of normalized scores $-\hat{f}_i(\mathbf{x})$, collected before applying adversarial attacks, and the IRI. The resulting correlation

TABLE I: Comparison of IRAA and public models on the test set of MS COCO.

Model	PCC	MSE	#Params
VGG11 [21]	0.896	0.053	132.9M
EfficientNet-B2 [22]	**0.936**	**0.033**	9.1M
Inception-V3 [23]	0.927	0.041	27.2M
ResNet-50 [24]	0.882	0.072	25.6M
ResNet-18 [24]	0.838	0.081	11.7M
IRAA	0.906	0.050	**78k**

TABLE II: Characteristics of public datasets.

Dataset	#Samples	Resolution	Mean IRI
CID2013 [25]	474	1600 × 1200	2.48
LIVE in the Wild [26]	1162	500 × 500	2.42
TID2013 [27]	3000	512 × 384	2.16
PIPAL [28]	23200	288 × 288	2.11
DIV2K [29]	800	2048 × 1080	1.95
KADID-10k [30]	10206	512 × 384	1.83
NIPS2017 [31]	1000	299 × 299	**1.64**
MS COCO (test2017)	41000	640 × 480	1.80

turned out to be 0.76. The drawback of this model, however, is its low correlation and the computational inefficiency of launching multiple metrics through various attacks.

B. Comparison with public models on the test set of MS COCO

Table I presents the results for common classification models we trained on our dataset generated from the train part of MS COCO dataset. Notice that all these models performed well, except for ResNet-18 and ResnNet-50. Also, attempts to add skip connections into the IRAA model reduced the model's quality. But the results for the public models are either worse or just slightly better than those of our model even though ours is 100 times smaller. This observation further proves that increasing our model's size failed to significantly improve the metrics for such architectures.

The PCC on the test part of MS COCO is 0.906 for the IRAA model — marginally lower than the best PCC of 0.936 achieved by EfficientNet-B2. So, our model exhibits the same accuracy as public models but takes up much less memory space.

C. Datasets robustness

To analyze dataset robustness to adversarial attacks, we selected the popular public datasets in Table II and applied the IRAA model to them. TID2013 provides information about the type and level of distortion, allowing us to analyze which ones have the greatest impact on image vulnerability to attacks (Sec. IV-D). NIPS2017 was used to develop and test adversarial attacks.

Table II shows the mean IRI scores for these datasets. NIPS2017 is the most impervious, since this dataset was designed to make adversarial image attacks more difficult. CID2013 and LIVE in the Wild are the most unstable. Images from LIVE in the Wild were captured using typical real-world mobile-device cameras, so many of them contain blurring and other distortions, explaining the observed result.

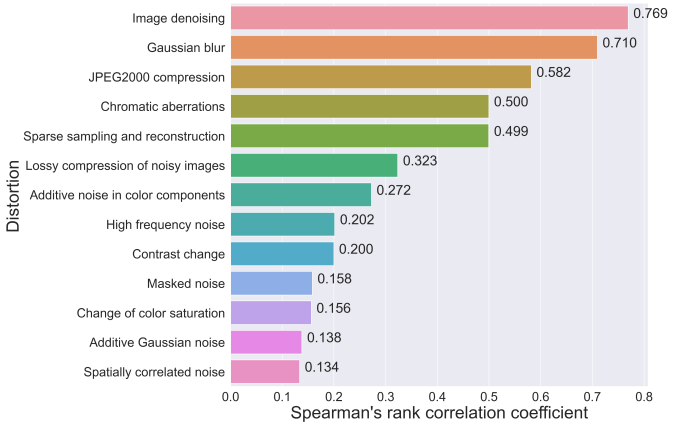

Fig. 5: Spearman's rank correlation coefficient (SRCC) between distortion level and image vulnerability to adversarial attacks for TID2013.

D. Distortion analysis

In this section we explore the types of distortions that have a positive or negative effect on image robustness.

In our initial observation, as Fig. 3 highlights, we suggested that the most unstable examples from MS COCO exhibit a blur effect, whereas the most stable ones are visually pleasing. To validate this observation, we thoroughly analyzed datasets with controlled distortions to determine which distortions have the greatest effect. We selected the artificially distorted image-quality dataset TID2013, which contains 3000 distorted images (24 distortion types and 5 distortion levels applyied to the 25 reference images). Subsequently, we applied the IRAA model to this dataset and calculated the Spearman's rank correlation coefficient (SRCC) the resulting scores and the distortion levels for each distortion (Fig. 5). We chose SRCC because we only need image robustness rankings, as we calculated correlation for integer distortion levels.

Our analysis confirmed that Gaussian blur, motion blur, lens blur, chromatic aberrations, JPEG2000 compression, image denoising, and sparse sampling decrease image robustness to adversarial attacks. It is worth noting that simply adding Gaussian noise has no effect, indicating that merely degrading the image quality is insufficient to alter image robustness. This finding aligns with expectations, as such cases are typically addressed when developing training datasets for quality metrics.

V. Conclusion

In this paper, we introduced the Image Robustness Index (IRI) and presented the Image Robustness to Adversarial Attacks (IRAA) model, which estimates the IRI. We demonstrated that image robustness depends on visual quality by analyzing the stability of publicly available datasets. Additionally, we explored how different distortions affect image stability to adversarial attacks: Gaussian blur, motion blur, lens blur, chromatic aberrations, JPEG2000 compression, image denoising, and sparse sampling are factors that decrease image robustness to adversarial attacks.

Our proposed IRAA model can help in preparing datasets for training new metrics or for applying adversarial attacks to

achieve more stable results; robust metrics should exhibit stability to different adversarial attacks, and effective adversarial attacks should break as many metrics as possible.

A. Limitations

Our initial training set only included images from MS COCO. After testing the IRAA model on KADID-10k and TID2013, we observed that distortions such as quantization noise and color quantization exhibit a high negative correlation with image robustness. This result is attributable to the target not correlating with such distortions.

B. Future work

Further research will extend the algorithm's applicability to video, prepare additional datasets for training metrics and attacks, and train metrics or attacks using the proposed method.

REFERENCES

[1] Anastasia Antsiferova, Alexander Yakovenko, Nickolay Safonov, Dmitriy Kulikov, Alexander Gushin, and Dmitriy Vatolin. Objective video quality metrics application to video codecs comparisons: choosing the best for subjective quality estimation. In *31th International Conference on Computer Graphics and Vision, Nizhny Novgorod, Russia*, 2021.

[2] Chao Ma, Chih-Yuan Yang, Xiaokang Yang, and Ming-Hsuan Yang. Learning a no-reference quality metric for single-image super-resolution. *Computer Vision and Image Understanding*, 158:1–16, 2017.

[3] Shuyang Gu, Jianmin Bao, Dong Chen, and Fang Wen. Giqa: Generated image quality assessment. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XI 16*, pages 369–385. Springer, 2020.

[4] Anastasia Antsiferova, Sergey Lavrushkin, Maksim Smirnov, Aleksandr Gushchin, Dmitriy Vatolin, and Dmitriy Kulikov. Video compression dataset and benchmark of learning-based video-quality metrics. *Advances in Neural Information Processing Systems*, 35:13814–13825, 2022.

[5] Anastasia Antsiferova, Khaled Abud, Aleksandr Gushchin, Ekaterina Shumitskaya, Sergey Lavrushkin, and Dmitriy Vatolin. Comparing the robustness of modern no-reference image-and video-quality metrics to adversarial attacks, 2024.

[6] Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, and Rob Fergus. Intriguing properties of neural networks. *ICLR*, 2014.

[7] Ian J Goodfellow, Jonathon Shlens, and Christian Szegedy. Explaining and harnessing adversarial examples. *ICLR*, 2015.

[8] Alexey Kurakin, Ian J Goodfellow, and Samy Bengio. Adversarial examples in the physical world. In *Artificial intelligence safety and security*, pages 99–112. Chapman and Hall/CRC, 2018.

[9] Yinpeng Dong, Fangzhou Liao, Tianyu Pang, Hang Su, Jun Zhu, Xiaolin Hu, and Jianguo Li. Boosting adversarial attacks with momentum. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 9185–9193, 2018.

[10] Aleksander Madry, Aleksandar Makelov, Ludwig Schmidt, Dimitris Tsipras, and Adrian Vladu. Towards deep learning models resistant to adversarial attacks. *ICLR*, 2018.

[11] Jiawei Su, Danilo Vasconcellos Vargas, and Kouichi Sakurai. One pixel attack for fooling deep neural networks. *IEEE Transactions on Evolutionary Computation*, 23(5):828–841, October 2019.

[12] Weixia Zhang, Dingquan Li, Xiongkuo Min, Guangtao Zhai, Guodong Guo, Xiaokang Yang, and Kede Ma. Perceptual attacks of no-reference image quality models with human-in-the-loop. *Advances in Neural Information Processing Systems*, 35:2916–2929, 2022.

[13] Ekaterina Shumitskaya, Anastasia Antsiferova, and Dmitriy Vatolin. Towards adversarial robustness verification of no-reference image-and video-quality metrics. *Computer Vision and Image Understanding*, 240:103913, 2024.

[14] Ekaterina Shumitskaya, Anastasia Antsiferova, and Dmitriy Vatolin. Fast adversarial cnn-based perturbation attack on no-reference image-and video-quality metrics. *ICLR 2023 Tiny Papers*, 2023.

[15] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, Proceedings, Part V 13*, pages 740–755. Springer, 2014.

[16] Dingquan Li, Tingting Jiang, and Ming Jiang. Unified quality assessment of in-the-wild videos with mixed datasets training. *International Journal of Computer Vision*, 129(4):1238–1257, jan 2021.

[17] Dingquan Li, Tingting Jiang, and Ming Jiang. Norm-in-norm loss with faster convergence and better performance for image quality assessment. In *Proceedings of the 28th ACM International Conference on Multimedia*. ACM, oct 2020.

[18] 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.

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

[20] Zhenqiang Ying, Haoran Niu, Praful Gupta, Dhruv Mahajan, Deepti Ghadiyaram, and Alan Bovik. From patches to pictures (paq-2-piq): Mapping the perceptual space of picture quality. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3575–3585, 2020.

[21] K Simonyan and A Zisserman. Very deep convolutional networks for large-scale image recognition. In *3rd International Conference on Learning Representations (ICLR 2015)*. Computational and Biological Learning Society, 2015.

[22] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pages 6105–6114. PMLR, 2019.

[23] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016.

[24] 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.

[25] Toni Virtanen, Mikko Nuutinen, Mikko Vaahteranoksa, Pirkko Oittinen, and Jukka Häkkinen. Cid2013: A database for evaluating no-reference image quality assessment algorithms. *IEEE Transactions on Image Processing*, 24(1):390–402, 2014.

[26] Deepti Ghadiyaram and Alan C Bovik. Massive online crowdsourced study of subjective and objective picture quality. *IEEE Transactions on Image Processing*, 25(1):372–387, 2015.

[27] Nikolay Ponomarenko, Lina Jin, O. Ieremeiev, Vladimir Lukin, Karen Egiazarian, J. Astola, Benoit Vozel, Kacem Chehdi, Marco Carli, Federica Battisti, and C.-C. Jay Kuo. Image database tid2013: Peculiarities, results and perspectives. *Signal Processing: Image Communication*, 30:57–77, 01 2015.

[28] Gu Jinjin, Cai Haoming, Chen Haoyu, Ye Xiaoxing, Jimmy S Ren, and Dong Chao. Pipal: a large-scale image quality assessment dataset for perceptual image restoration. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XI 16*, pages 633–651. Springer, 2020.

[29] 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.

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

[31] Google Brain. Nips 2017: Adversarial learning development set. https://www.kaggle.com/datasets/google-brain/nips-2017-adversarial-learning-development-set, 2017.