https://doi.org/10.2352/EI.2022.34.9.IQSP-334

Assessing the Impact of Image Quality on Object-Detection Algorithms

Abhinau K. Venkataramanan*, Marius Facktor*, Praful Gupta, and Alan C. Bovik;
Department of Electrical and Computer Engineering,
The University of Texas at Austin

Abstract

The field of image and video quality assessment has enjoyed rapid development over the last two decades. Several datasets and algorithms have been designed to understand the effects of common distortions on the subjective experiences of human observers. The distortions present in these datasets may be synthetic (applying artificially computed blur, compression, noise, etc.) or authentic (in-capture lens flare, motion blur, under/overexposure, etc.). The goal of quality assessment is often to quantify the loss of visual "naturalness" caused by the distortion(s). We have recently created a new resource called LIVE-RoadImpairs, which is a novel image quality dataset consisting of authentically distorted images of roadways. We use the dataset to develop a no-reference quality assessment algorithm that is able to predict the failure rates of object-detection algorithms. This work was among the overall winners of the PSCR Enhancing Computer Vision for Safety Challenge.

Introduction

After AlexNet [1] achieved state-of-the-art (SOTA) performance on the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [2], Deep Neural Networks (DNNs) have emerged as dominant computing architectures for various conducting computer vision tasks such as detection and semantic segmentation of objects [3] [4], faces [5] [6], medical images [7] [8], etc. Such networks are often pre-trained on large datasets of annotated images, such as ImageNet, or adapted to smaller datasets using transfer learning. Most of the images in these datasets are of reasonably good quality.

With regards to image quality, the success of the Structural Similarity index (SSIM) [9] and other image quality prediction models has led to the creation of a variety of subjective quality datasets directed towards different aspects of image integrity, including natural images impaired by synthetic [10] [11] and authentic distortions [12] [13], 3D images [14] [15], VR images [16] [17], and HDR images [18] [19]. These datasets are particularly valuable since they embody the subjective quality experiences of human observers.

While the experience of subjective quality is often regarded as "pre-cognitive," the role of "quality" in task-specific imaging modalities, such as medical imaging, is tied to the usefulness of the image for completing the task, which may be strongly affected by its quality. The role of image quality as it affects the performances of object detection DNNs is less clear, since they are not designed to mimic human observers, and so, may not be affected by quality degradations in the same way. For example, while DNNs can surpass human performance on object-detection benchmarks, human observers have been shown to be more robust to image degradations than DNNs [20]. Further, recent evidence has also suggested that using high-quality images to train DNNs on visual recognition tasks may not result in optimal performance on images encountered in practice [21].

In other words, for object-detection DNNs trained on high-quality images, distorted images constitute a domain shift, which often leads to a loss in performance. This is particularly true of the kinds of real world "authentic" distortions that typically arise during the image capture process, which can include complex combinations of blur, noise, shake, poor exposure, compression and mode. This is especially true of images taken in disaster-response situations, like those that were the focus of the PSCR Enhancing Computer Vision for Safety Challenge. Images captured in such situations may suffer from the effects of bad weather, low-light, potentially rapid camera/object motions, and the limitations of the camera. The ways these distortions may combine and interact to create new distortions makes this task especially challenging.

The first contribution of our work is a novel dataset called LIVE-RoadImpairs, which consists of images of roadways distorted by authentic capture distortions. We used the YOLOv3 DNN [4] as a representative SOTA object-detection system, using which we obtained ground-truth "failure rates." The second contribution of our work is a no-reference quality model that is able to predict the failure rate of the SOTA object-detection algorithm. Such a model has the potential to be used to guide improvements to the quality of the images input to object-detection algorithms, or to aid training of the detector to be "distortion-aware", or simply to ascribe confidence scores to the detection outcomes.

The rest of the paper is organized as follows. In Section , we describe previous work on understanding the effects of distortions on the performances of computer vision algorithms. In Section , we describe the key features of the LIVE-RoadImpairs dataset. In Section we describe a new failure-rate assessment algorithm. Finally, in Section , we study and discuss the performance of our algorithm.

Related Work

The effect of distortions on the performances of computer vision algorithms has been a topic of interest in recent years. For example, Quality Labeled Faces in the Wild (QLFW) [22] dataset consists of faces of images subjected to five different distortions. Similarly, the noisy MNIST (n-MNIST) [23] dataset consists of images of handwritten digits subjected to noise, motion blur, and reduced contrast. Dodge et al. [24] conducted an evaluation of pre-trained object detection models on distorted images, analyzing the impact of synthetically generated noise, blur, compression,

and reduced contrast on performance. Similarly, Vasiljevic et al. [25] evaluated the effects of blur on the performance of DNNs, noting that by fine-tuning networks on blurry images, their robustness to blur was improved. The benefits of fine-tuning and of retraining models on synthetically distorted images has been studied in [26].

ImageNet-C [27] was the first benchmark dedicated to analyzing the robustness of image classification networks against corruptions of input images. It contains ImageNet images corrupted by noise, blur, motion, weather, and brightness variations, all of which were also synthetically applied. ImageNet-C inspired the development of the Robust Object Detection Benchmark, consisting of similarly constructed PASCAL-C, COCO-C, and Cityscape-C datasets [28]. A similar resource for face recognition is also available, which consists of images from the Labelled Faces in the Wild (LFW) [29] dataset that have been subjected to several distortions, including noise, blur, contrast, etc., and occlusions on parts of the face, such as the eyes, nose, and mouth. [30].

Early work on task-specific quality models either analyzed linear detectors of objects against noisy backgrounds [31] [32], or conducted contrived tasks such as detecting nylon beads in radiographic images [33]. A comprehensive study of human detection performance on a non-trivial task in the presence of image degradations was conducted in [20]. This work compared the object detection performances of humans and deep networks. The authors found that the performance of human observers was significantly more robust to image degradations than of deep networks. In a similar vein, the NIST-LIVE X-Ray IED image-quality dataset [34] is a basic tool for evaluating the threat detection performance of trained bomb technicians when viewing security images of varying quality. This dataset was used to develop QUIX [35], which is a suite of algorithms that can predict bomb technician performance based on image quality.

A key feature of all the above work is that the available datasets consist of pristine images that have been artificially subjected to idealized, synthetic degradations. While synthetic distortions allow for systematic analyses of detection performance by varying the "strengths" of distortions, neither these analyses, nor models built on them, can be effectively applied on images with authentic distortions, which typically occur at the capture stage and often consist of complex combinations of multiple coincident distortions that are difficult or impossible to effectively model. What is needed are datasets of task-specific images that have been naturally distorted as they are normally acquired, processed, and stored. Here, we attempt to bridge this gap by building a dataset of authentically distorted images containing objects to be detected. Using this resource, we also develop an algorithm that is able to predict the performance of a SOTA deep object detection network on distorted image data.

Dataset

LIVE-RoadImpairs is an image dataset that includes common impairments affecting images capture by cameras on the roadway. The images are in JPEG format and have a resolution of 3840 horizontal pixels by 2160 vertical pixels. The images were obtained by first recording video with an iPhone 8 in 4K mode at 24 FPS. Then, frames were extracted from the video which exemplify various impairments occurring while driving.

The dataset consists of 789 images and the composition by impairment type is listed in Table 1, where the "High" and "Low" columns list the number of images that exhibit high and low levels of severity of each impairment respectively. This classification was performed by human visual inspection.

Table 1: Number of images of each distortion category in the RoadImpairs dataset

Category	Quantity	High	Low
Pristine	172	-	-
Out of Focus	91	-	-
Night	164	-	-
Oncoming Headlights	65	-	-
Oily	111	-	-
Motion Blur	113	45	68
Direct Sunlight	38	13	25
Rain	108	19	89
Snow	102	1	101

Each image in the dataset suffers from at least one impairment type, except for the "pristine" labeled images which were deemed to be without impairments. However, images are often affected by multiple simultaneous impairments, that combine to create complex composite distortions that are difficult to model. For example, a single image may contain "rain" and also be "out of focus." The "oily" distortion was created by applying a thin layer of plant oil to the camera lens (a blend of golden Jojoba and Moroccan Argan oils). This was used to simulate the effects of residual oil left by fingerprints on improperly handled camera lenses.

Since the impairment categories were assigned manually, there was some subjectivity in the process. We labelled an image as having an impairment if that impairment was visually noticeable on the image. For example, a car facing the camera may have had its headlights on, but we only reported "oncoming headlights" if the image recorded an artifact because of these headlights, such as a halo effect or lens flare.

In addition to the impairments, we also assigned one of five metadata "types" to the image: "clear," "night," "oily," "rainy," and "snowy." These metadata types describe the conditions under which the images were taken and are not necessarily the same as their impairment classifications. For example, the image, "rainy_11," was captured on a rainy day, but it is classified as "pristine" since there were no observable impairments.

Computer Vision Task

The object detection task embodied by the LIVE-RoadImpairs dataset is to identify roadway related objects under various conditions and distortions. The object detection algorithm that we used was the popular YOLOv3 network with fixed weights pretrained on ImageNet. The inferenced outputs of YOLOv3 are bounding boxes around putative detected objects, and predicted labels denoting the most likely classes the detected objects fall into, as quantified by a generated vector of class probabilities.

To combine object detection and object classification performance, we created a scalar "accuracy score" to describe YOLOv3's performance on a particular image. This score mea-

sures the Intersection over Union (IoU) between the ground-truth bounding boxes and the bounding boxes proposed by YOLOv3. It also includes the class likelihood for the true class, regardless of whether that class is predicted as the most likely. The contributions from each object are weighted by the relative size of the true bounding boxes. The final accuracy score of an image x with N objects, each of class c_i is as follows:

$$S = \sum_{i=1}^{N} w_i (IoU_i^{\rho} \times p(y_i = c_i \mid x)^{1-\rho}), \quad (1)$$

where the weight factor is given by

$$w_i = \frac{Area_i}{\sum_{n=1}^{N} Area_n}, \quad (2)$$

where $Area_i$ is the area of the true bounding box of object i.

For simplicity, we used $\rho = 0.5$. The accuracy score is a number between 0 and 1, where 0 means zero overlap and/or zero predicted likelihood, and 1 means the predictions were exactly correct. This number indicates how well the object detector performed on an image. The failure rate on the image is then

$$FR = 1 - S. \quad (3)$$

To facilitate learning a quality model, we annotated two non-overlapping subsets of RoadImpairs - a training set of 172 images, and a test set of 50 images. The distributions of images by impairment in the train and test sets are listed in Table 2. "H" and "L" denote images that exhibit high and low severity of some of the impairments.

Table 2: Number of images of each category in the Training and Test sets

	Training			Test		
Category	**Quantity**	**H**	**L**	**Quantity**	**H**	**L**
Pristine	35	-	-	7	-	-
Out of Focus	26	-	-	6	-	-
Night	26	-	-	10	-	-
Oncoming Headlights	11	-	-	3	-	-
Oily	26	-	-	8	-	-
Motion Blur	37	17	20	6	4	2
Direct Sunlight	10	3	7	4	1	3
Rain	26	10	16	11	3	8
Snow	17	1	16	6	0	6

When constructing the training and test sets, we ensured that the images contained at least one object. We annotated the images using the online resource MakeSense [36], and only labeled road-related objects, viz., "car," "truck," "motor bike," "bicycle," "stop sign," and "fire hydrant." We generated a ground-truth accuracy score on every image by performing inference using YOLOv3. These two sets were then used to train and test a model that maps a RoadImpairs image to an estimated accuracy score.

Failure Rate Assessment

Log-Gabor Features

No-Reference (NR) Image Quality Assessment (IQA) typically involves measuring the deviation of a given test image from "natural" statistical behaviour. This is achieved by transforming the image such that the transform coefficients of natural images are expected to exhibit regular statistical properties, except when they are distorted. Measuring the loss of statistical naturalness is the bases of the most successful NR IQA models that predict human impressions of visual quality. A divisive normalization transform (DNT) applied to bandpass (wavelet) image coefficients is the basic processing flow for these kinds of NR IQA models.

While the bandpass-DNT approach has perceptual relevance, the "viewer" in our case is a machine, and not a human. Therefore, we posited that using a processing flow that reflects how deep networks represent images would lead to better performance. While the exact parameters of deep neural networks vary greatly, it has been commonly observed that filters in early layers resemble bar- and edge-detectors, or more explicitly, bandpass filters resembling Gabor functions. The log-Gabor function whereby the Gabor filtering is applied in the log-frequency domain is also a commonly-used model. Log-Gabor filters have also been successfully used for NR IQA, as for example in IL-NIQE [37].

Since the source images were captured at 4K resolution, we downsampled them to 720p. This substantially reduces the computational burden of our algorithm.

Following the notation from [37], consider a filter bank of NJ filters, corresponding to N center frequencies and J orientations. A 2D log-Gabor filter can be expressed in the frequency domain as

$$G_{nj}(\omega, \theta) = \exp\left(-\frac{\log^2(\omega/\omega_n)}{2\sigma_r^2}\right) \exp\left(-\frac{(\theta-\theta_j)^2}{2\sigma_\theta^2}\right), \quad (4)$$

where ω_n is the n-th central frequency and θ_j is the j-th orientation. We constructed a log-Gabor filter bank using the same parameters as in [37]:

- $N = 3$ center frequencies, $J = 4$ orientations
- Minimum wavelength $\lambda_{min} = 2.4$
- $\sigma_r = -\log(0.55)$
- Multiplication factor $\mu = 1.31$
- Center frequencies $\omega_n = 1/(\lambda_{min} \times \mu^n)$ for $n = 0, 1, \ldots N-1$.
- Orientations $\theta_j = j\pi/J$ for $j = 0, 1, \ldots J-1$

Each test image was convolved with this filter bank to obtain a set of subband log-Gabor coefficients. Each subband was partitioned into non-overlapping blocks of size 50×50, then the real and imaginary parts of the coefficients were collected into two-dimensional vectors. A parametric Bivariate Generalized Gaussian Distribution (BGGD) model was fit to these vectors to obtain best-fit mean μ, covariance matrix Σ, and shape parameters α, using the moment-matching method [38]. The probability density function of a BGGD is given by

$$f(x; \mu, \Sigma, \alpha) = K \exp\left(-\frac{1}{2}\left((x-\mu)^T C^{-1}(x-\mu)\right)^{\alpha}\right), \quad (5)$$

where

$$K = \frac{\alpha}{\pi\Gamma(1/\alpha) * 2^{1/\alpha}\sqrt{\det(C)}}, \quad (6)$$

and

$$C = \frac{2^{1-1/\alpha}\Gamma(1/\alpha)}{\Gamma(2/\alpha)}\Sigma. \tag{7}$$

In practice, we observed that the measurements of the correlation between the real and imaginary parts were very low, so we did not include it as a feature. In all, we collected five quality aware features - means, variances, and shape parameter, from each patch, i.e.

$$\mathbf{f} = [\mu_0, \mu_1, \sigma_0^2, \sigma_1^2, \alpha]^T. \tag{8}$$

The features from all of the subbands of each 50×50 patch were concatenated, to obtain a 60-dimensional quality-aware feature vector for each patch.

Object Occurrence-Based Weighting

Incorporating visual attention information in the form of saliency maps appears to slightly improve the performance of image quality algorithms that predict human judgments of image quality [39]. In this vein, when viewing distorted images, observers' quality judgments tend to be more heavily affected by the worst-quality regions in an image [40], especially when they degrade salient objects. Hence, spatial pooling methods that assign higher weights to more distorted regions tend to perform better [41]. However, when machine vision algorithms process images to detect objects, the worst distorted region may not be the most salient. For example, an image of a car may suffer lens flare high above in the sky from the sun. As a result, the sky would be the worst-distorted region, but the car's image may remain unaffected, likely leading to reliable detection. Therefore, when evaluating machine vision performance against distortion, it is natural to consider assigning larger distortion weights on regions that are more likely to contain objects.

Since object saliency is closely-related to object detection, SOTA deep-learning based saliency models are susceptible to similar losses in performance due to distortions as object detection models. Indeed, we have found that classical saliency methods that do not use deep-learning usually fail in the presence of distortions. So, we instead utilized a simple data-driven algorithm to estimate an "object-occurrence prior" (OOP), which quantifies the likelihood of a region in the image falling within an object. We term this a "prior" since it is estimated once from the training set and then is applied on all the images in the training and test sets. More formally, the value of the OOP at each pixel location i, j is

$$OOP(i,j) = P\left(pixel(i,j) \in object\right). \tag{9}$$

To obtain the OOP, we used Kernel Density Estimation (KDE). At each pixel, we computed the proportion of times that it lie within a bounding box. Then, we used a Gaussian kernel with $\sigma = 50$ to smooth these "empirical" estimates, and normalized them to sum to 1. The map was then resized to obtain the OOP of each 50×50 patch in the 720p image as the sum of the OOP values of all pixels within that patch. Hence, the size of the patch in the OOP map must be chosen appropriately. Let s_p denote the OOP of a patch p in an image, and $\mathbf{f}_p$ denote its 60-dimensional feature vector. Then, the object occurrence-weighted feature vector of the image is obtained as a weighted average of the patch-level feature vectors:

$$\mathbf{f} = \sum_p s_p \mathbf{f}_p. \tag{10}$$

Table 3: Mean Ground Truth and Predicted Accuracy Scores on RoadImpairs

Impairment Type	Accuracy Score	
	Mean Ground Truth	Mean Prediction
Pristine	0.84	0.75
Out of Focus	0.39	0.43
Night	0.45	0.44
Oncoming Headlights	0.43	0.41
Oily	0.78	0.60
High Motion Blur	0.67	0.52
Low Motion Blur	0.90	0.59
High Direct Sunlight	0.56	0.40
Low Direct Sunlight	0.87	0.60
High Rain	0.06	0.56
Low Rain	0.59	0.64
High Snow	-	0.28
Low Snow	0.71	0.58

Support Vector Regressor

We used a Support Vector Regressor (SVR) to model the relationship between the object occurrence-weighted feature vectors and the accuracy scores. The kernel used was the radial basis function (RBF), and L2 regularization was applied to avoid overfitting. The regularization parameter was obtained by optimizing performance on the test set. Therefore, this SVR functions as a mapping between an image's object occurrence-weighted log-Gabor features, and a number that represents how well YOLOv3 is expected to perform on that image.

On the test set, the SVR achieved a mean error of 0.26, with 38% of images scoring within ± 0.2. We then used the SVR to obtain an estimated accuracy score for every image in the LIVE-RoadImpairs dataset. Since the output of an SVR is an unbounded real number, we clipped all predicted values to the range $[0, 1]$.

Analysis of Results

We examined two ways to evaluate the efficacy of our quality model. First, we analyzed the test set results by impairment type. Second, we provide the average predicted accuracy score for each impairment type on RoadImpairs as a whole, and demonstrate that the results conformed to our understanding of the distortion types.

We examined the test set by measuring the differences between a predicted accuracy score and the mean true accuracy score for each impairment type. We preface these calculations involving the test set with an explanation. There are more factors than just impairment type and severity that affect how well

an object detection system performs. These include the distances of objects from the camera, and the complexity and commonness of each object's form. When navigating a roadway, the presences and locations of objects can be arbitrary, and a lack of any objects does not necessarily mean that an object detection algorithm has failed. Our method of analyzing distortion resilience focuses on the image impairments and without accounting for the positioning, form, or other natures of the objects in the image.

Yet, this aspect of the problem becomes important when analyzing the test set because the true accuracy scores heavily depend on the positioning of objects in the test images. Averaging the true accuracy scores over each impairment type yields a stable estimate of how well the object detector performed.

We calculated the mean true accuracy score of each of the 13 impairment types, then computed the absolute differences between the predicted accuracy scores and the mean true score, for its associated impairment. For images with multiple impairments, we found the differences between the predictions and each true mean score. Using these differences, we computed percentage differences corresponding to each of these 61 absolute differences. Finally, we aggregated over the test set to obtain a Median Percentage Difference (MPD) score of 18.8%. We used the median to mitigate the adverse effect of a small number of outliers (3 out of 61 data points).

In addition, we observed that almost 48 of the predicted scores, or 78.7%, were within 30% of the mean true score. We believe this to be a fairer assessment of the test set results than simple mean error, since "accuracy" and "failure rates" are aggregate phenomena, i.e., at the scale of sets of images, rather than of individual images.

We also examined the efficacy of the predictions over the *entire* RoadImpairs dataset by comparing the average predicted scores for each impairment, as tabulated in Table 3. Intuitively, we should expect pristine images to earn higher average accuracy scores than images distorted any impairment type. This was the case, and the average pristine score was 0.106 higher than the second-highest impairment type (low direct sunlight). Further, high severity impairments should lead to lower accuracy scores than low severity impairments, on average. This was also satisfied by every impairment type that contains a High/Low distinction.

Conclusion

We constructed LIVE-RoadImpairs, a publicly available dataset of authentically distorted images of roadways, with annotations describing the imaging conditions and impairments. We also labelled a subset of the image data with failure rates. Further, we created an object occurrence-aware no-reference quality model that predicts the failure rate of YOLOv3, which is a popular SOTA object detection algorithm. In the future, we expect that the use of better saliency prediction models and more sophisticated DNNs may improve the performance of task-specific image quality models.

Acknowledgments

This research was sponsored by a grant from Facebook Video Infrastructure, prize money from the Public Safety Communications (PSCR) Division of the National Institute of Science and Technology (NIST) as part of the 2020 Enhancing Computer Vision for Public Safety Challenge, and by grant number 2019844 for the National Science Foundation AI Institute for Foundations of Machine Learning (IFML).

References

[1] A. Krizhevsky, I. Sutskever, and G.E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, Red Hook, NY, USA, 2012, NIPS'12, p. 1097–1105, Curran Associates Inc.

[2] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A.C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision (IJCV)*, vol. 115, no. 3, pp. 211–252, 2015.

[3] S. Ren, K. He, R.B. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *CoRR*, vol. abs/1506.01497, 2015.

[4] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," *arXiv*, 2018.

[5] Y. Sun, D. Liang, X. Wang, and X. Tang, "DeepID3: Face Recognition with Very Deep Neural Networks," *CoRR*, vol. abs/1502.00873, 2015.

[6] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A Unified Embedding for Face Recognition and Clustering," *CoRR*, vol. abs/1503.03832, 2015.

[7] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *CoRR*, vol. abs/1505.04597, 2015.

[8] M. Arvinte, S. Vishwanath, A.H. Tewfik, and J.I. Tamir, "Deep J-Sense: Accelerated MRI Reconstruction via Unrolled Alternating Optimization," 2021.

[9] Wang Z, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[10] H.R. Sheikh, "LIVE image quality assessment database release 2," *http://live. ece. utexas. edu/research/quality*, 2005.

[11] H. Lin, V. Hosu, and D. Saupe, "KADID-10k: A Large-scale Artificially Distorted IQA Database," in *2019 Tenth International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2019, pp. 1–3.

[12] V. Hosu, H. Lin, T. Sziranyi, and D. Saupe, "KonIQ-10k: An Ecologically Valid Database for Deep Learning of Blind Image Quality Assessment," *IEEE Transactions on Image Processing*, vol. 29, pp. 4041–4056, 2020.

[13] D. Ghadiyaram and A.C. Bovik, "Massive Online Crowdsourced Study of Subjective and Objective Picture Quality," *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 372–387, 2016.

[14] A.K. Moorthy, C. Su, A. Mittal, and A.C. Bovik, "Subjective evaluation of stereoscopic image quality," *Signal Processing: Image Communication*, vol. 28, no. 8, pp. 870–883, 2013, Special Issue On Biologically Inspired Approaches For Visual Information Processing And Analysis.

[15] J. Wang, A. Rehman, K. Zeng, S. Wang, and Z. Wang, "Quality Prediction of Asymmetrically Distorted Stereoscopic 3D Images," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3400–3414, 2015.

[16] M. Chen, Y. Jin, T. Goodall, X. Yu, and A.C. Bovik, "Study of 3D Virtual Reality Picture Quality," *IEEE Journal of Selected Topics in*

Signal Processing, vol. 14, no. 1, pp. 89–102, 2020.

[17] H. Duan, G. Zhai, X. Min, Y. Zhu, Y. Fang, and X. Yang, "Perceptual Quality Assessment of Omnidirectional Images," in *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2018, pp. 1–5.

[18] D. Kundu, D. Ghadiyaram, A.C. Bovik, and B.L. Evans, "Large-Scale Crowdsourced Study for Tone-Mapped HDR Pictures," *IEEE Transactions on Image Processing*, vol. 26, no. 10, pp. 4725–4740, 2017.

[19] P. Korshunov, P. Hanhart, T. Richter, A. Artusi, R. Mantiuk, and T. Ebrahimi, "Subjective quality assessment database of HDR images compressed with JPEG XT," in *2015 Seventh International Workshop on Quality of Multimedia Experience (QoMEX)*, 2015, pp. 1–6.

[20] S. Dodge and L. Karam, "A Study and Comparison of Human and Deep Learning Recognition Performance under Visual Distortions," in *2017 26th International Conference on Computer Communication and Networks (ICCCN)*, 2017, pp. 1–7.

[21] C. Wu, L.F. Isikdogan, S. Rao, B. Nayak, T. Gerasimow, A. Sutic, L. Ain-kedem, and G. Michael, "VisionISP: Repurposing the Image Signal Processor for Computer Vision Applications," in *2019 IEEE International Conference on Image Processing (ICIP)*, 2019, pp. 4624–4628.

[22] L.J. Karam and T. Zhu, "Quality labeled faces in the wild (QLFW): a database for studying face recognition in real-world environments," in *Human Vision and Electronic Imaging XX*. International Society for Optics and Photonics, 2015, vol. 9394, p. 93940B.

[23] S. Basu, M. Karki, S. Ganguly, R. DiBiano, S. Mukhopadhyay, and R.R. Nemani, "Learning Sparse Feature Representations using Probabilistic Quadtrees and Deep Belief Nets," *CoRR*, vol. abs/1509.03413, 2015.

[24] S.F. Dodge and L.J. Karam, "Understanding How Image Quality Affects Deep Neural Networks," *CoRR*, vol. abs/1604.04004, 2016.

[25] I. Vasiljevic, A. Chakrabarti, and G. Shakhnarovich, "Examining the Impact of Blur on Recognition by Convolutional Networks," *CoRR*, vol. abs/1611.05760, 2016.

[26] Y. Zhou, S. Song, and N. Cheung, "On Classification of Distorted Images with Deep Convolutional Neural Networks," *CoRR*, vol. abs/1701.01924, 2017.

[27] D. Hendrycks and T.G. Dietterich, "Benchmarking Neural Network Robustness to Common Corruptions and Perturbations," *CoRR*, vol. abs/1903.12261, 2019.

[28] C. Michaelis, B. Mitzkus, R. Geirhos, E. Rusak, O. Bringmann, A.S. Ecker, M. Bethge, and W. Brendel, "Benchmarking Robustness in Object Detection: Autonomous Driving when Winter is Coming," *CoRR*, vol. abs/1907.07484, 2019.

[29] G.B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments," Tech. Rep. 07-49, University of Massachusetts, Amherst, October 2007.

[30] S. Karahan, M.K. Yildirum, K. Kirtac, F.S. Rende, G. Butun, and H. Kemal H.K. Ekenel, "How Image Degradations Affect Deep CNN-Based Face Recognition?," in *2016 International Conference of the Biometrics Special Interest Group (BIOSIG)*, 2016, pp. 1–5.

[31] H.H. Barrett, "Objective assessment of image quality: effects of quantum noise and object variability," *J. Opt. Soc. Am. A*, vol. 7, no. 7, pp. 1266–1278, Jul 1990.

[32] H.H. Barrett, J.L. Denny, R.F. Wagner, and K.J. Myers, "Objective assessment of image quality. II. Fisher information, Fourier crosstalk, and figures of merit for task performance," *J. Opt. Soc. Am. A*, vol. 12, no. 5, pp. 834–852, May 1995.

[33] L. Loo, K. Doi, and C.E. Metz, "A comparison of physical image quality indices and observer performance in the radiographic detection of nylon beads," *Physics in Medicine & Biology*, vol. 29, no. 7, pp. 837, 1984.

[34] J. Glover, P. Gupta, N.G. Paulter, and A.C. Bovik, "Study of Bomb Technician Threat Identification Performance on Degraded X-ray Images," *Journal of Perceptual Imaging*, vol. 4, 05 2021.

[35] P. Gupta, Z. Sinno, J.L. Glover, N.G. Paulter, and A.C. Bovik, "Predicting Detection Performance on Security X-Ray Images as a Function of Image Quality," *IEEE Transactions on Image Processing*, vol. 28, no. 7, pp. 3328–3342, 2019.

[36] "MakeSense," https://www.makesense.ai/.

[37] L. Zhang, L. Zhang, and A. C. Bovik, "A Feature-Enriched Completely Blind Image Quality Evaluator," *IEEE Transactions on Image Processing*, vol. 24, no. 8, pp. 2579–2591, 2015.

[38] E. Gómez, M.A. Gomez-Viilegas, and J.M. Marín, "A multivariate generalization of the power exponential family of distributions," *Communications in Statistics - Theory and Methods*, vol. 27, no. 3, pp. 589–600, 1998.

[39] X. Min, G. Zhai, Z. Gao, and K. Gu, "Visual attention data for image quality assessment databases," in *2014 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2014, pp. 894–897.

[40] A.K. Moorthy and A.C Bovik, "Visual Importance Pooling for Image Quality Assessment," *IEEE Journal of Selected Topics in Signal Processing*, vol. 3, no. 2, pp. 193–201, 2009.

[41] A. K. Venkataramanan, C. Wu, A. C. Bovik, I. Katsavounidis, and Z. Shahid, "A Hitchhiker's Guide to Structural Similarity," *IEEE Access*, vol. 9, pp. 28872–28896, 2021.

Author Biography

Abhinau K. Venkataramanan received his B.Tech. degree in Electrical Engineering from the Indian Institute of Technology, Hyderabad, India, in 2019. He is currently pursuing his M.S. and Ph.D. degrees in Electrical and Computer Engineering at the University of Texas at Austin, TX, USA. In the past, he has worked as a research intern at Carnegie Mellon University and as a summer intern at Facebook, Inc.

Marius Facktor received his BS in computer engineering from the University of Wisconsin (2019) and his MS from the University of Texas (2021). He now works as a computer vision engineer at a medical technology company.

Praful Gupta received his M.S. degree followed by Ph.D. in electrical and computer engineering from The University of Texas at Austin, Austin, in 2017 and 2021, respectively. He is currently working as an Applied Scientist with Camera software team at Amazon Lab 126. His research interests include image and video processing, machine learning, and computer vision.

Al Bovik (HonFRPS) is the Cockrell Family Regents Endowed Chair Professor at The University of Texas at Austin. He will receive the 2022 IEEE Edison Medal "for pioneering high-impact scientific and engineering contributions leading to the perceptually optimized global streaming and sharing of visual media." Previously he received the a Technology and Engineering Emmy® Award, the RPS Progress Medal, the IEEE Fourier Award, the OSA Edwin Land Medal, and the 2015 Primetime Emmy® Award.