# Perceptual Quality Assessment of Images and Videos In the Wild

Deepti Ghadiyaram<sup>\*</sup> Advisor: Dr. Alan Bovik

# 1 Introduction

The field of visual media has been witnessing explosive growth in recent years, driven by significant advances in technology made by camera and mobile device manufacturers, and by the synergistic development of very large media-centric social networking websites. All of these now allow consumers to efficiently capture, store, and share high-resolution images and videos with their friends or the community at large. Furthermore, given the ubiquitous availability of portable mobile devices for image and video capture, over-the-top (OTT) video streaming now dominates global mobile data traffic. The vast majority of the digital media being uploaded on the social media are taken by casual, inexpert users, where the capture process is affected by delicate variables such as lighting, exposure, aperture, noise sensitivity, and lens limitations, each of which could perturb an image's (or video's) perceived visual quality. This leads to large numbers of images and videos of unsatisfactory *perceptual quality* being captured and stored along with more desirable ones. Pertaining to OTT video streaming, network impairments or bandwidth limitations can cause volatile network conditions, resulting in rebuffering or stalling events, which interrupt a video's playback.

Accounting for an end user's quality of experience (QoE) is very important. The presence of poor quality images and videos on any multimedia service impacts a viewer's QoE with that service. With regards to OTT video streaming, network-induced stalling events can negatively impact a viewer's satisfaction with cellular network service quality. Being able to automatically identify and cull low quality images and videos, or to prevent their occurrence by suitable quality correction processes during capture are thus highly desirable goals that could be enabled by automatic quality prediction tools [1].

**Goal of my thesis:** The overarching goal of my thesis is to *objectively* predict the visual quality of real-world videos and images *in-the-wild*. The task of an objective **no-reference** (or blind) visual quality assessment (NR VQA) model is as follows: given an image or a video (possibly distorted) and no other additional information, automatically and accurately predict its perceptual quality. Both perceptual visual quality and an end user's QoE are highly subjective in nature and are the result of a combined effect produced by factors such as diverse distortions (spatial, temporal, or those induced due to network congestion), visual content, an individual's sensitivity to distortions, aesthetics, and so on. Therefore, designing a quality predictor that accounts for such subjective factors but still correlates well with human opinion scores is highly challenging.

## 2 Research Summary

Below, I describe my research on no-reference image and video quality assessment on images and videos afflicted with real world distortions and summarize my research contributions.

#### 2.1 No-Reference Image Quality Assessment

#### 2.1.1 A new image quality database of real world distortions

Most top-performing image quality assessment (IQA) models (full, reduced, and no-reference) have been designed and evaluated on legacy singly-distorted databases [2, 3]. These databases have been designed to contain images corrupted by <u>only one</u> of a few synthetically introduced distortions, which are introduced in a controlled manner. These single, synthetic distortion databases are especially problematic and hinder the development of *blind* IQA models which have great potential to be employed in large-scale, real-world, user-centric visual media applications, where the encountered images contain mixtures of authentic distortions.

To address this limitation, we created a challenging blind image quality database containing images that were captured using numerous individual mobile devices. Each image was collected without artificially introducing any distortions beyond those occurring during capture, processing, and storage by a user's device. These images are affected by unknown mixtures of single or more commonly occurring multiple interacting **authentic** distortions of diverse severities. The content and characteristics of the new LIVE In the Wild Image Quality Challenge Database, which contains 1162 *authentically* distorted images captured from many diverse mobile devices can be found in [4]

#### 2.1.2 A crowdsourcing framework for conducting subjective studies efficiently

The human opinion scores in most of the legacy datasets [2, 3] were collected by conducting subjective studies in laboratory settings with stringent controls on the experimental environments, where images were displayed on a single device with a fixed display resolution and which the subjects viewed from a fixed distance, involving small, nonrepresentative subject samples (typically graduate and undergraduate university students). However, the subjective

<sup>\*</sup>The University of Texas at Austin. Email: deepti@cs.utexas.edu.

image quality opinions gathered under artificially controlled settings do not necessarily mirror the picture quality perceived on widely used portable display devices having varied resolutions.

This limitation motivated us to design and implement an **online crowdsourcing system** which we used to gather more than 350,000 human ratings of image quality on the new LIVE Challenge Database. Using this framework, we were able to successfully conduct two large-scale image quality subjective studies [4, 5].

#### 2.1.3 A top-performing image quality predictor for real world images

As mentioned earlier, all state-of-the-art models are trained and evaluated on synthetic, and usually singly distorted images contained in benchmark databases [2, 3]. Thus their performance on images containing complex mixtures of authentic distortions such as those found in the real-world pictures that are captured using mobile devices is questionable. To address this limitation, we devised an approach that leverages the idea that different perceptual image representations may distinguish different aspects of the loss of perceived image quality. Thus, given an image, we first construct several *feature maps* in multiple color spaces and transform domains, then extract individual and collective scene statistics from each of these maps. We train an SVR on these statistical features and achieve a performance that is superior to all existing state-of-the-art models [6]

#### 2.2 No-Reference Video QoE Prediction

Here, the overarching goal is to study and understand the influence of the effect of network impairments on QoE and design generalizable QoE models for mobile videos. We describe our efforts towards this goal below:

#### 2.2.1 A subjective quality database for mobile streaming videos with stalling events

Over-the-top mobile video streaming is invariably influenced by volatile network conditions which cause playback interruptions (stalling events), thereby impairing users' quality of experience (QoE). We conducted a subjective study to thoroughly understand the specific factors regarding video stream quality that effect viewers' QoE and better understand how fluctuations in network video quality affect viewer behavior. The resulting **LIVE-Avvasi Mobile Video database** [7] consists of 180 distorted videos generated from 24 reference videos with 26 unique stalling events and 4830 human opinions obtained from 54 subjects who viewed the videos on mobile devices.

#### 2.2.2 A time-varying subjective quality model for videos with stalling events

Existing objective models that predict QoE are based on global video features, such as the number of stall events and their lengths. The model we proposed in [8] goes beyond previous models as it also accounts for the fundamental effect that a viewers recent level of satisfaction or dissatisfaction has on their overall viewing experience, in addition to accounting for the lengths, frequency of occurrence, and the positions of stall events - factors that interact in a complex way to affect a users QoE. We evaluated our model on the LIVE-Avvasi Video Database and achieve state-of-the-art prediction performance.

## References

- [1] A. C. Bovik. Automatic prediction of perceptual image and video quality. IEEE Proc., 101:2008–2024, 2013.
- [2] H. R. Sheikh, M. F. Sabir, and A. C. Bovik. A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Trans. Image Process.*, 15(11):3440–3451, 2006.
- [3] N. Ponomarenko, O. Ieremeiev, V. Lukin, K. Egiazarian, L. Jin, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, and C.-C. Jay Kuo. Color image database tid2013: Peculiarities and preliminary results. Proc. of 4th European Workshop on Visual Info. Proc., pages 106–111, 2013.
- [4] D. Ghadiyaram and A.C. Bovik. Massive online crowdsourced study of subjective and objective picture quality. *IEEE Trans. on Image Proc.*
- [5] D. Kundu, D. Ghadiyaram, A.C. Bovik, and B. L. Evans. Large-scale crowdsourced study for high dynamic range pictures. *IEEE Trans. on Image Proc.* (under review).
- [6] D. Ghadiyaram and A.C. Bovik. Perceptual quality prediction on authentically distorted images using a bag of features approach. J. of Vision. (in print) https://arxiv.org/abs/1609.04757.
- [7] D. Ghadiyaram, A. C. Bovik, H. Yeganeh, R. Kordasiewicz, and M. Gallant. Study of the effects of stalling events on the Quality of Experience of mobile streaming videos. *IEEE Global Conf. on Signal and Information Proc.*, December 2014.
- [8] D. Ghadiyaram, J. Pan, and A.C. Bovik. A time-varying subjective quality model for mobile streaming videos with stalling events. *Proc. SPIE Optical Engineering + Applications*, Aug. 12 15, 2015.