A User-Centric Toolbox for Visualising and Evaluating Image Quality Metrics

Matt Clifford[1], Alexander Hepburn[1], Ricardo Kleinlein[2], Jorge Vila Tomás[3], Pablo Hernández Cámara[3], Paula Dauden[3], Jesús Malo[3], Nathan Lepora[1], Valero Laparra[3], and Raúl Santos-Rodríguez[1]

[1]University of Bristol, [2]Universidad Politécnica de Madrid, [3]Universitat de Valencia

Image Quality Metrics

Image quality metrics (IQMs) serve as an objective evaluation of the quality of an image and attempt to capture how humans perceive differences between images. Reference IQMs compare a reference image, and the same image with a distortion applied to it. The goal is to recreate how humans perceive this difference according to human psychophysical experiments. They are utilised in scenarios such as loss functions for deep learning models and benchmarking the performance of image processing algorithms where human evaluation is too expensive to obtain.

There are a plethora of IQMs to choose from, the simplest being a metric in Euclidean space such as the mean squared error. In traditional perceptual literature, IQMs can be categorised into two groups. The first group consists of metrics, like the structural similarity (SSIM) index, which operate on the premise that the image's structure remains unchanged despite the presence of distortion. These metrics adhere to the principle of *structural similarity*. The second group aims to measure how the *visibility of error* i.e. how much distortion is visible to humans. Recent literature utilises deep learning models in an attempt to mimic human perception by correlating the model's response with image quality ratings from human experiments.

The process of evaluating IQMs is both qualitative and quantitative. It is necessary to gather empirical data on the response profiles of different IQMs with respect to specific image distortions and parameter ranges. It is also important for a human to oversee the images produced from the distortion process to give context to the image data. The practitioner may also require specific behaviour from an IQM, which is not observable in just measuring correlation between human ratings and distances given by the IQM. For example, invariances are often extremely important. This type of evaluation and experimentation lends itself best to interactive and graphical software.

Our toolbox (IQM-Vis) enables quick access to visualising image distortions and evaluating IQMs through a simple and convenient python graphical interface. It provides many standard distortions and IQMs out of the box whilst adding custom distortions, IQMs and image datasets is also straightforward. Comparison graphs of IQMs are automatically generated as well as the option to compare with human scores to understand failures of the metric with particular distortions or images. This improves the quality and timescale of the IQM evaluation process. IQM-Vis manages the data storage from the experiments and shows the results with correlation plots against desired IQMs. Any image distortions which do not conform to the correlation can be selected for further analysis of the image properties.

IQM-Vis - Demo Information

In our demo we will exhibit our open-source toolbox (IQM-Vis). We will illustrate why picking the correct IQM is important and motivate the need for a user-centric approach since evaluation of image quality and IQMs needs to be integrated with human opinion. We will explain how to customise the toolbox with user defined image datasets, distortions and IQMs. We will illustrate how to evaluate the claims of an IQM for a specific image distribution and set of distortions, exploring where limitations of an IQM may arise.

One of the key features of IQM-Vis is its ability to provide a comprehensive analysis of IQMs through various quantitative graphs. We will show how these graphs offer practitioners a well-rounded understanding of how different IQMs perform under different scenarios.

Additionally, we will showcase IQM-Vis's functionality for conducting human perception experiments. This feature enables practitioners to analyse custom distortions and image datasets, which is crucial for real-world design and evaluation of IQMs. Experiment participants provide a ranking of the perceived quality of an image exposed to specific distortions. We will demonstrate how IQM-Vis facilitates the management and analysis of the collected data.

IQM-Vis is implemented in Python, the graphical user interface (GUI) is developed using PyQt6, the image handling and processing is covered by OpenCV. Numpy is used for image distortions and other backend tasks and many of the IQMs provided take advantage of the pytorch framework to enable GPU hardware acceleration. To accompany the demo, we have prepared online documentation, videos and tutorials that provide step-by-step instructions on installing, running and customising IQM-Vis.

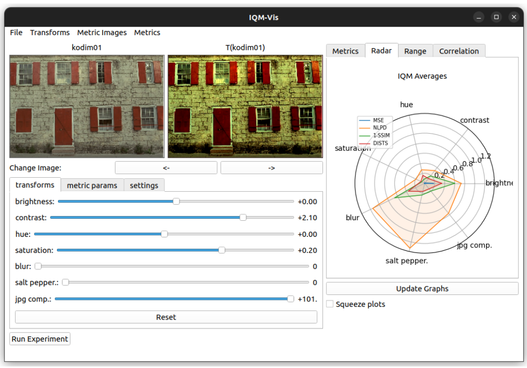

Figure 1: The IQM-Vis interface. The radar plot of the right shows the average IQM score for each distortion over its parameter range. This forms a quick comparison of their sensitivities. The distorted image for a specific set of distortions is shown on the left.