SOM: Semantic Obviousness Metric for Image Quality Assessment

Peng Zhang, Wengang Zhou, Lei Wu, Houqiang Li
Department of Electronic Engineering and Information Science, University of Science and Technology of China

Image quality assessment (IQA) tries to estimate human perceptual based image visual quality in an objective manner. Existing approaches target this problem with or without reference images. For no-reference image quality assessment, there is no given reference image or any knowledge of the distortion type of the image. Previous approaches measure the image quality from signal level rather than semantic analysis. They typically depend on various features to represent local characteristic of an image.

In this paper we propose a new no-reference (NR) image quality assessment (IQA) framework based on semantic obviousness. We discover that semantic-level factors affect human perception of image quality. With such observation, we explore semantic obviousness as a metric to perceive objects of an image. We propose to extract two types of features, one to measure the semantic obviousness of the image and the other to discover local characteristic. Then the two kinds of features are combined for image quality estimation. We evaluate our approach on the LIVE dataset. Our approach is demonstrated to be superior to the existing NR-IQA algorithms and comparable to the state-of-the-art full-reference IQA (FR-IQA) methods. Cross-dataset experiments show the generalization ability of our approach.

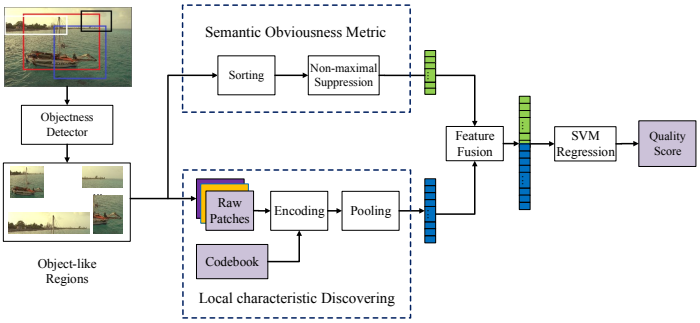

Figure 1: Framework of image quality assessment with semantic obviousness metric

The framework of our proposed image quality estimation method SOM is illustrated in Figure 1. As shown in the figure, we first extract all the object-like regions with an objectness detector. From these regions, we compute two kinds of features: one to measure the global semantic obviousness of the image and one to discover local characteristic.

Object-like Region Detection: We choose the object detector BING [1], which is extremely fast and shows high object detection rate with good generalization ability. Not all the detected regions are actually objects, but they are distinct from the neighbor regions so that they can attract more attention.

Semantic Obviousness Feature Extraction: For each image in the dataset, we extract all its object-like regions, each with a detection score. Typically BING extracts 2000 to 3000 object-like regions for a single image. We sort these regions in descending order based on the corresponding detection score. We define semantic obviousness feature based on the detection score of the top K object-like regions.

$$X = [x_1, x_2, \ldots, x_K]^T \tag{1}$$

Local Feature Extraction: We use a codebook based method [5] for local characteristic discovering. Unlike previous approaches [2, 3, 4, 5], we extract local features only from object-like regions instead of the who image. We denote the set of N top-scored regions as S. M raw image patches are uniformly and randomly sampled from S. Each patch is normalized and its columns are concatenated to generate a vector as its descriptor. As a result, for S we obtain a local descriptor $Y = [y_1, y_2, \ldots, y_M]$, where $y_i \in R^d$, $d = B \times B$. We use images from an unrelated dataset to construct a codebook $D_{d \times W} = [d_1, d_2, \ldots, d_W]$, where $d_i (d_i \in R^d, d = B \times B)$ are normalized cluster centroids of normalized raw image patches. Local features are then quantized by performing soft-assignment coding on the codebook D. The similarity between the i^{th} local feature y_i and the j^{th} codeword d_j is computed by their dot-product as: $s_{i,j} = \langle y_i, d_j \rangle$. Local feature y_i is encoded as follows:

$$\begin{aligned} c_i = &[max(s_{i,0}, 0), \ldots, max(s_{i,W}, 0), \\ &max(-s_{i,1}, 0), \ldots, max(-s_{i,W}, 0)]^T \end{aligned} \tag{2}$$

In the encoding step, we get a coefficient matrix $C_{2W \times M} = [c_1, c_2, \ldots, c_M]$, where $c_i = [c_{i,1}, c_{i,2}, \ldots, c_{i,2W}]^T$ is obtained by Equation (2). We perform max-pooling on each row of C to convert it to a vector. After pooling, we get a feature in the form of:

$$Z = [z_1, z_2, \ldots, z_{2W}]^T \tag{3}$$

where z_i is the maximum of the i^{th} row in coefficient matrix $C_{2W \times M}$. The final feature Z represents the local characteristic of the image.

Feature Fusion: We get two kinds of features for an input image: X measures the semantic obviousness and Z represents the local characteristic. We combine the two kinds of features to measure image quality both on semantic and pixel level. The final descriptor F is in the form as follows:

$$F_{(K+2W) \times 1} = [x_1, \ldots, x_K, z_1, \ldots, z_{2W}]^T \tag{4}$$

Regression: We use support vector machine regression (SVR) to map the image feature F to image quality score.

Various experiments are performed on the LIVE dataset. Our method outperforms the state-of-the-art NR-IQA methods and is comparable to the FR-IQA methods. The idea of introducing object detection into image quality assessment can be incorporated with existing FR-IQA and NR-IQA methods. For the full-reference algorithms, we extract N top-scored object-like regions of each reference image. For each region, the corresponding region in the distorted image is extracted to obtain a predicted quality score. For the non-reference algorithm, we extract top N object-like regions for each distorted image. Then we average the predicted scores of these regions to obtain the quality score of the whole distorted image. The performance of these methods get obvious improvement in this experimental setting.

[1] Ming-Ming Cheng, Ziming Zhang, Wen-Yan Lin, and Philip Torr. Bing: Binarized normed gradients for objectness estimation at 300fps. In *CVPR*, pages 3286–3293. IEEE, 2014.

[2] Dong-O Kim, Ho-Sung Han, and Rae-Hong Park. Gradient information-based image quality metric. *IEEE Transactions on Consumer Electronics*, 56(2):930–936, 2010.

[3] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. No-reference image quality assessment in the spatial domain. *TIP*, 21(12): 4695–4708, 2012.

[4] Anush K Moorthy and Alan C Bovik. A two-step framework for constructing blind image quality indices. *Signal Processing Letters*, 17(5): 513–516, 2010.

[5] Peng Ye, Jayant Kumar, Le Kang, and David Doermann. Unsupervised feature learning framework for no-reference image quality assessment. In *CVPR*, pages 1098–1105. IEEE, 2012.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.