Real-time No-Reference Image Quality Assessment based on Filter Learning: Supplementary Material

Peng Ye, Jayant Kumar, Le Kang, David Doermann
Institute for Advanced Computer Studies
University of Maryland, College Park, MD, USA
{pengye,jayant,lekang,doermann}@umiacs.umd.edu

1 Supervised Filter Learning

Suppose we have n training images and the k-th training image is denoted as X_k. The corresponding feature vector of X_k is $Z_k = \phi(X_k, B)$ where $B = [b_1, ..., b_K] \in R^{d\times K}$ represents a set of filters. The supervised filter learning method jointly optimizes the prediction model and the set of filters. The objective function of this optimization problem is defined as follows:

$$C(B, w, \{X_k\}_{k=1}^n) = \sum_{k=1}^n L(y_k, f(\phi(X_k, B), w)) + \lambda_1 ||w||_{l_2}^2 + \lambda_2 avecorr(B)$$
$$subject\ to\ ||b_i|| = 1, i = 1, ..., K \tag{1}$$

where f is the prediction function: $f(Z_k, w) = \sum_{i=1}^{2K} w_i Z_k(i) + w_0$, y_k is the true quality score of the k-th image, L is the ϵ-insensitive loss function and $avecorr(B) = \frac{1}{K-1}\sum_{i=1}^K \sum_{j:j\neq i} < b_i, b_j >$ is the average correlation of one filter with every other filters.

Optimal B and w is given by

$$(B^*, w^*) = argmin_{B,w} C(B, w, \{X_k\}_{k=1}^n)$$

This optimization problem can be solved by optimizing alternatively over B and w. The initial set of filters are obtained by performing k-means clustering on a set of local features. When B is fixed, optimal w can be found using a standard SVR program. Given w fixed, we can apply stochastic gradient descent (SGD) to find the optimal B.

The SGD process requires us to compute the gradient of the objective function C with respect to a filter $b_i, i = 1, ..., K$. Using chain rule, we can have:

$$\frac{\partial C}{\partial b_i} = \sum_{k=1}^n \frac{\partial L}{\partial f_k}\frac{\partial f_k}{\partial Z_k}\frac{\partial Z_k}{\partial b_i} + \lambda_2 \frac{\partial avecorr(B)}{\partial b_i} \tag{2}$$

where $f_k = f(Z_k, w)$ is the predicted quality score for the k-th training image.

The loss function used in ϵ-SVR is non-differentiable, therefore we use the following approximation (Huber loss) for computing gradient

$$L(y, \hat{y}) = \begin{cases} 0 & |y-\hat{y}| \le \epsilon - h \\ |y-\hat{y}| - \epsilon & |y-\hat{y}| \ge \epsilon + h \\ \frac{(y-\hat{y}-\epsilon+h)^2}{4h} & \epsilon - h < y - \hat{y} < \epsilon + h \\ \frac{(y-\hat{y}+\epsilon-h)^2}{4h} & -\epsilon - h < y - \hat{y} < -\epsilon + h \end{cases} \tag{3}$$

where $0 < h < \epsilon$. When $h \to 0$, Eq. 3 is equivalent to the ϵ-insensitive loss used in ϵ-SVR. The derivative of the above loss function is given by:

$$\frac{\partial L}{\partial f_k} = \begin{cases} 0 & |y_k - f_k| \le \epsilon - h \\ -1 & y_k - f_k \ge \epsilon + h \\ 1 & y_k - f_k \le -\epsilon - h \\ \frac{f_k - y_k + \epsilon - h}{2h} & \epsilon - h < y_k - f_k < \epsilon + h \\ \frac{f_k - y_k - \epsilon + h}{2h} & -\epsilon - h < y_k - f_k < -\epsilon + h \end{cases} \tag{4}$$

The derivative of the prediction function with respect to the global feature vector is given by [1]:

$$\frac{\partial f_k}{\partial Z_k} = [w_1, w_2, ..., w_{2K}] \tag{5}$$

The global feature vector $Z_k = [Z_k(1), ..., Z_k(2K)]^T$, where for $i = 1, ..., K$

$$Z_k(i) = b_i \cdot x^k_{max,i} \ , \ x^k_{max,i} = argmax_{x_l \in X_k}(b_i \cdot x_l)$$

$$Z_k(i+K) = b_i \cdot x^k_{min,i} \ , \ x^k_{min,i} = argmin_{x_l \in X_k}(b_i \cdot x_l)$$

where superscript k is the index of the training image, $x_l \in X_k$ means x_l is a local feature vector from image X_k and $\cdot$ represents inner product. We therefore have $\frac{\partial Z_k(i)}{\partial b_i}^T = x^k_{max,i}$, $\frac{\partial Z_k(i+K)}{\partial b_i}^T = x^k_{min,i}$ and the derivative of the global feature vector with respect to b_i is given by:

$$\begin{aligned} \frac{\partial Z_k}{\partial b_i} &= [0, ..., 0, \frac{\partial Z_k(i)}{\partial b_i}^T, 0, ..., 0, \frac{\partial Z_k(i+K)}{\partial b_i}^T, 0, ..., 0]^T \\ &= [0, ..., 0, x^k_{max,i}, 0, ..., 0, x^k_{min,i}, 0, ..., 0]^T \end{aligned} \tag{6}$$

The derivative of the correlation penalty term with respect to b_i is given by:

$$\frac{\partial avecorr(B)}{\partial b_i} = \frac{1}{K-1} \sum_{j:j \neq i} b_j^T \tag{7}$$

To sum up, when linear ϵ-SVR is used, we can compute the derivative of the objective function as follows [2]:

$$\frac{\partial C}{\partial b_i} = (\sum_{k=1}^{n} \frac{\partial L}{\partial f_k}(w_i x^k_{max,i} + w_{i+K} x^k_{min,i}) + \lambda_2 \frac{1}{K-1} \sum_{j:j \neq i} b_j)^T \tag{8}$$

2 Examples of filter responses

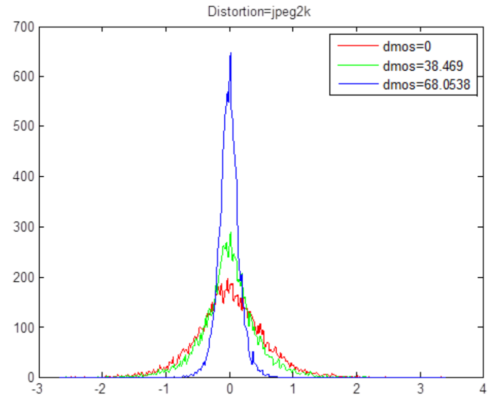
To demonstrate that with properly learned filters, the distribution of filter responses from distorted and non-distorted images are very different, we show filter responses of images with five different types of distortions at three different distortion levels in Fig. 1. The non-distorted reference image is also shown in Fig. 1.

3 Doc-IQA Dataset

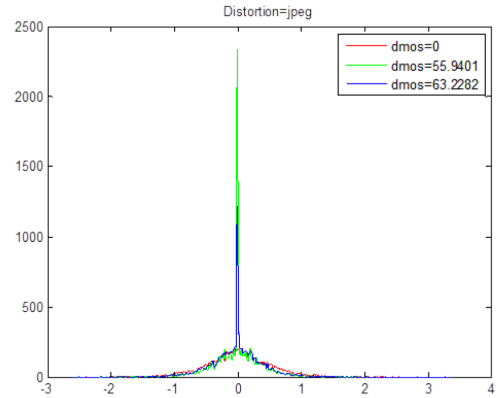
Examples of images from the SOC and the Newspaper dataset are shown in Fig. 2 and Fig. 3 respectively. The histogram of OCR accuracy of images from these two Doc-IQA dataset are shown in Fig. 4. It can be seen that the distribution of the OCR accuracy of images from both Doc-IQA datasets are non-uniform and the SOC dataset is highly imbalanced.

[1] If $y \in R^m$, $x \in R^n$, then $\frac{\partial y}{\partial x} \in R^{m \times n}$

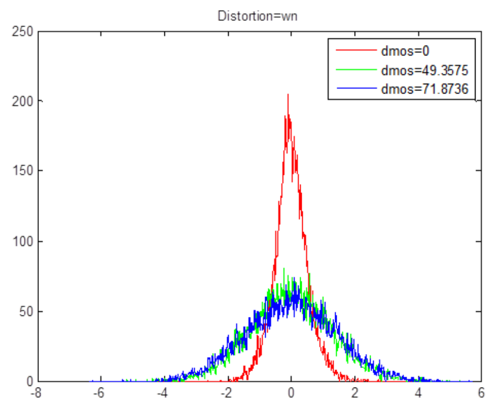
[2] For ease of representation, the transpose sign in Eq. 8 is omitted in the paper.



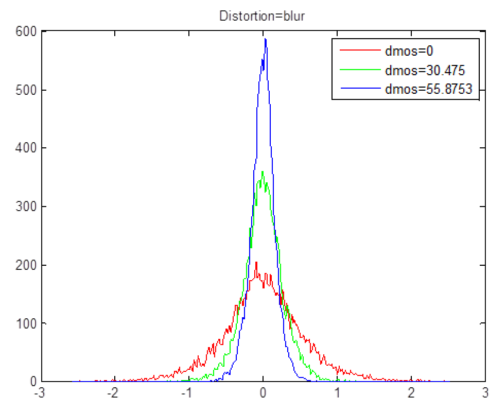
(a) JPGE2K Compression



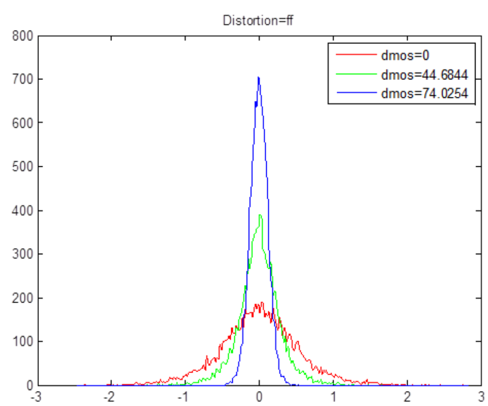
(b) JPEG Compression



(c) White Gaussian Noise



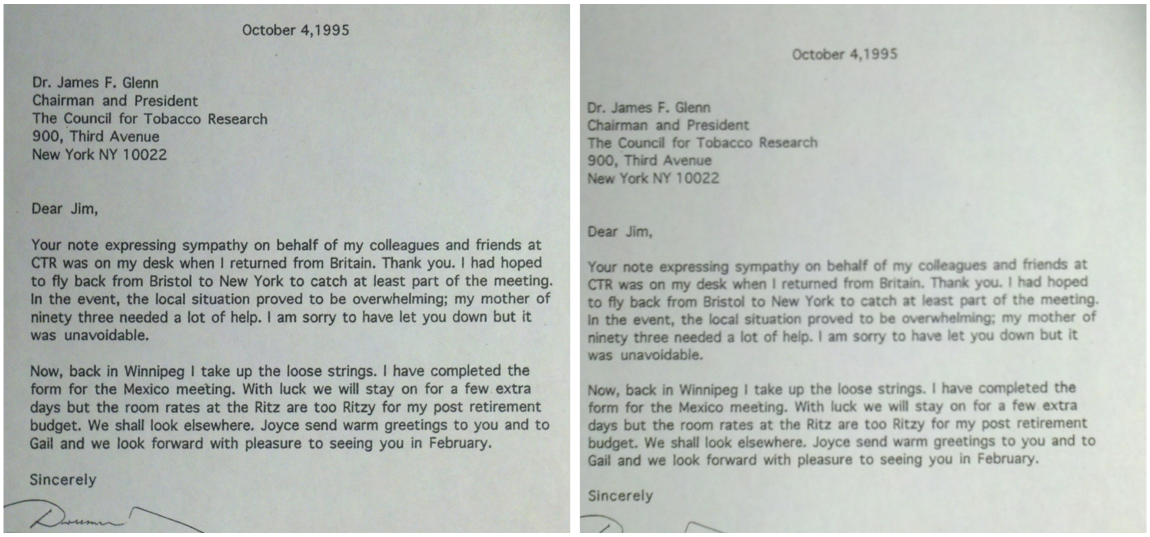
(d) Gaussian Blurring



(e) Fast Fading

(f) Reference Image

Figure 1: Examples of filter responses for different types and levels of distortions. (High DMOS indicates low quality)


October 4,1995

Dr. James F. Glenn
Chairman and President
The Council for Tobacco Research
900, Third Avenue
New York NY 10022

Dear Jim,

Your note expressing sympathy on behalf of my colleagues and friends at CTR was on my desk when I returned from Britain. Thank you. I had hoped to fly back from Bristol to New York to catch at least part of the meeting. In the event, the local situation proved to be overwhelming; my mother of ninety three needed a lot of help. I am sorry to have let you down but it was unavoidable.

Now, back in Winnipeg I take up the loose strings. I have completed the form for the Mexico meeting. With luck we will stay on for a few extra days but the room rates at the Ritz are too Ritzy for my post retirement budget. We shall look elsewhere. Joyce send warm greetings to you and to Gail and we look forward with pleasure to seeing you in February.

Sincerely

October 4,1995

Dr. James F. Glenn
Chairman and President
The Council for Tobacco Research
900, Third Avenue
New York NY 10022

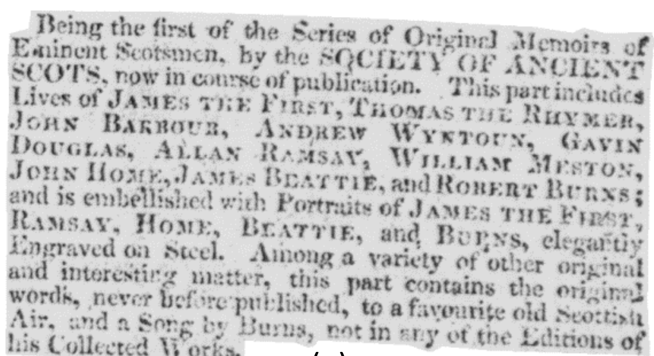
Dear Jim,

Your note expressing sympathy on behalf of my colleagues and friends at CTR was on my desk when I returned from Britain. Thank you. I had hoped to fly back from Bristol to New York to catch at least part of the meeting. In the event, the local situation proved to be overwhelming; my mother of ninety three needed a lot of help. I am sorry to have let you down but it was unavoidable.

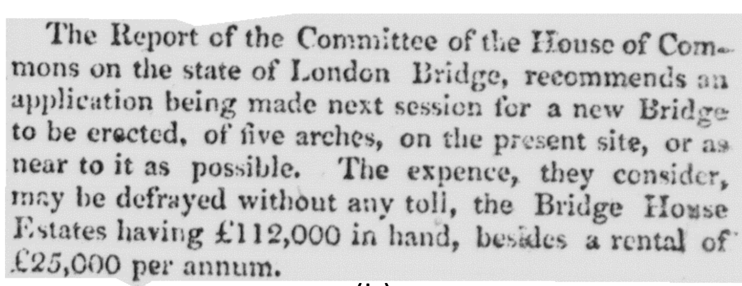
Now, back in Winnipeg I take up the loose strings. I have completed the form for the Mexico meeting. With luck we will stay on for a few extra days but the room rates at the Ritz are too Ritzy for my post retirement budget. We shall look elsewhere. Joyce send warm greetings to you and to Gail and we look forward with pleasure to seeing you in February.

Sincerely

Figure 2: Examples of images from the SOC dataset.

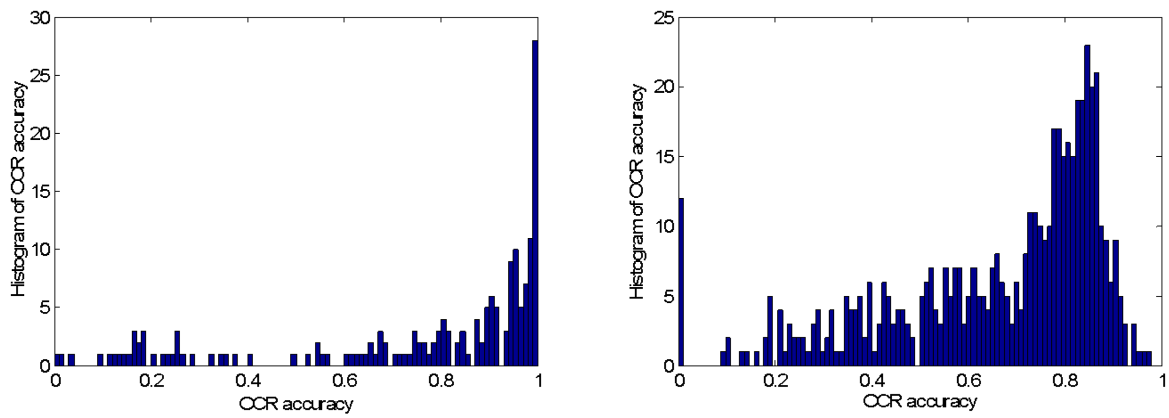

Being the first of the Series of Original Memoirs of Eminent Scotsmen, by the SOCIETY OF ANCIENT SCOTS, now in course of publication. This part includes Lives of JAMES THE FIRST, THOMAS THE RHYMER, JOHN BARBOUR, ANDREW WYNTOUN, GAVIN DOUGLAS, ALLAN RAMSAY, WILLIAM MESTON, JOHN HOME, JAMES BEATTIE, and ROBERT BURNS; and is embellished with Portraits of JAMES THE FIRST, RAMSAY, HOME, BEATTIE, and BURNS, elegantly Engraved on Steel. Among a variety of other original and interesting matter, this part contains the original words, never before published, to a favourite old Scottish Air, and a Song by Burns, not in any of the Editions of his Collected Works.

(a)


The Report of the Committee of the House of Commons on the state of London Bridge, recommends an application being made next session for a new Bridge to be erected, of five arches, on the present site, or as near to it as possible. The expence, they consider, may be defrayed without any toll, the Bridge House Estates having £112,000 in hand, besides a rental of £25,000 per annum.

(b)

Figure 3: Examples of images from the newspaper dataset. Character level OCR accuracy (a) 0.097 (b) 0.958.



(a) SOC dataset.

(b) Newspaper dataset.

Figure 4: Histogram of OCR accuracy of images from the SOC and the Newspaper dataset.