Discriminative Convolutional Neural Network for Image Quality Assessment with Fixed Convolution Filters

Motohiro TAKAGI††, Akito SAKURAI††, Members, and Masafumi HAGIWARA†, Senior Member

SUMMARY Current image quality assessment (IQA) methods require the original images for evaluation. However, recently, IQA methods that use machine learning have been proposed. These methods learn the relationship between the distorted image and the image quality automatically. In this paper, we propose an IQA method based on deep learning that does not require a reference image. We show that a convolutional neural network with distortion prediction and fixed filters improves the IQA accuracy.

Key words: image quality assessment, subjective image quality, convolutional neural network

1. Introduction

In image applications, image quality assessment (IQA) is important for improving the quality of service. Measuring the subjective image quality enables improving the quality of the image viewing experience. However, the subjective IQA is difficult to evaluate precisely. The need for simpler methods to measure the subjective image quality is increasing. Objective IQA methods such as peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) [1] have been conventionally used as substitutes for subjective IQA. The PSNR and SSIM are full reference IQA (FR IQA) methods that do not depend upon the data. Recently, data-driven FR IQA methods based on unsupervised learning have been proposed [2], [3], which perform better than the PSNR and SSIM. However, these FR IQA methods require the original images and the distorted image to evaluate the image quality. In certain mobile-device image applications, FR IQA methods cannot be applied because the original image is unavailable. Therefore, a no reference IQA (NR IQA) that does not require the original image has become crucial.

2. No Reference Image Quality Assessment

NR IQA methods using machine learning have been proposed. CORNIA generates feature vectors based on a codebook and uses them as inputs for regression to estimate the image quality [4]. Kang et al. [5] proposed a method using a convolutional neural network (CNN). They constructed a five-layer CNN model that can learn parameters from divided images directly. In addition, they proposed the IQA model that can estimate the type of distortion and image quality simultaneously [6]. However, their neural network is shallow and has a few parameters to estimate the subjective image quality.

These IQA methods using the CNN achieved high performance. Therefore, we proposed a model that incorporates the distortion type prediction with more parameters [7]. The model used the L2 norm as an evaluation metric. In this letter, we propose a new CNN model that incorporates DCT-based fixed filters, and show that it improves the image quality estimation accuracy.

3. Discriminative CNN with Fixed Convolution Filters

3.1 Network Structure and Loss Function

The network structure for this study was developed based on the model in [7]. It consists of five layers of units, 32 × 32 − 26 × 26 − 2 × 50 − 800 − 800 − 7. We use the same normalization as in [5] for the input. In the first convolution layer 98 kernels are applied to the input image. Each kernel produces a 26 × 26 image to be used as an input to the second layer. This image goes through max and min pooling before reaching the second layer. Forty-nine kernels of the first-layer filters are learned by network optimization. The other 49 kernels are constructed based on the DCT. In the third and fourth layers, the units are fully connected with rectified linear units (ReLUs) as activation functions. The fifth layer features elements that output the predicted image quality scores and the distortion type probabilities. Elements are prepared for each distortion type. By adding such elements, it is assumed that each distortion signal can be captured more precisely and the image quality can be estimated more accurately. To predict the subjective score and distortion label, our model minimizes the following loss function [7]:

\[ L_{\text{total}} = w_1 L_r + w_2 L_c, \]  

(1)

where \( L_r \) is the regression error between the subjective and estimated scores. We use the L2 norm as the regression error. \( L_c \) is the softmax cross entropy required for minimizing the classification error to classify the distortion label. \( w_1 \) and \( w_2 \) are weights. By optimizing the regression and classification errors by Eq. (1), the accuracy can be improved.
3.2 Fixed Convolution Filters Based on DCT

We propose a new method that incorporates DCT filters to the model. By including a DCT basis, the input image can be separated linearly, and noise can be expressed more precisely. It is assumed that the performance is improved by separating the signal in advance because certain distortions vary considerably in the low frequency region, while others vary considerably in the high frequency region. The generated DCT bases are of the same size as the other kernels, i.e., 7×7. Therefore, 49 DCT bases are added to the convolution layer.

4. Results

4.1 Experimental Setting

The proposed method was evaluated using the CSIQ database, which comprises 30 types of reference images [8]. In the database, six types of distortion were evaluated in five stages by subjects. The six types of distortion included in the JPEG2000 encoding distortion (JP2K), JPEG encoding distortion (JPEG), changes in the contrast (CONTRAST), additive Gaussian white noise (AWGN), Gaussian pink noise (FNOISE), and blur (BLUR). A total of 866 distorted images were used. A degradation mean opinion score (DMOS) was calculated for each distorted image. We divided the CSIQ dataset into training and validation data. Validation was calculated for each distorted image. We divided the dataset into training and validation data. Validation data used “1600”, “aerial city”, and “Boston”, categories. The remaining images were used for the training data. The linear correlation coefficient (LCC) and Spearman’s rank-order correlation coefficient (SROCC) were used as the assessment metrics. We use \( w_1 = 0.9 \) and \( w_2 = 0.1 \) in Eq. (1).

4.2 Estimation Results

Table 1 shows the SROCC and LCC between the predicted and true scores. The CNN [5] and the Discriminative CNN (DCNN) [7] are the baseline methods. The DCT-CNN has fixed DCT convolution filters, however, it does not include distortion discriminating elements in the network. The discriminated DCT-CNN (DDCT-CNN) implements our proposed structure with fixed DCT convolution filters and distortion discriminating elements. In terms of the SROCC, the DDCT-CNN and DCT-CNN outperform the CNN for all the distortion types, except JPEG2000, and the DDCT-CNN is comparable with the DCT-CNN. In terms of the LCC, the DDCT-CNN and DCT-CNN outperform the CNN for all the distortion types, and that the DDCT-CNN performs better than the DCT-CNN. AWGN is the most increase type of distortion in SROCC and LCC. It is expected that all the filters based on DCT will express Gaussian noise. The DCT-CNN and DDCT-CNN also perform better than the DCNN. The separation of signals in the early layer improves the accuracy. This result shows that distortions that generally cannot be learnt by CNNs were learnt by our model.

5. Conclusion

We proposed a CNN-based IQA model that incorporates the distortion type prediction and DCT-based fixed filters to the CNN. By incorporating the aforementioned features, correlation coefficients, image quality, and estimated image quality, were improved compared to those of the conventional CNNs.

References

[1] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” IEEE Trans. Image Process., vol.13, no.4, pp.600–612, April 2004.
[2] D. Temel, M. Prabhushankar, and G. AlRegib, “UNIQUE: Unsupervised image quality estimation,” IEEE Signal Process. Lett., vol.23, no.10, pp.1414–1418, Oct. 2016.
[3] M. Prabhushankar, D. Temel and G. AlRegib, “MS-UNIQUE: Multi-model and sharpness-weighted unsupervised image quality estimation,” IS&T Int. Symp. Electron. Imaging, vol.2017, no.12, pp.30–35, 2017.
[4] P. Ye, J. Kumar, L. Kang, and D. Doermann, “Unsupervised feature learning framework for no-reference image quality assessment,” Proc. IEEE Conf. Comput. Vision Pattern Recognit. (CVPR), pp.1098–1105, 2012.
[5] L. Kang, P. Ye, Y. Li, and D. Doermann, “Convolutional neural networks for no-reference image quality assessment,” Proc. IEEE Conf. Comput. Vision Pattern Recognit. (CVPR), pp.1733–1740, 2014.
[6] L. Kang, P. Ye, Y. Li, and D. Doermann, “Simultaneous estimation of image quality and distortion via multi-task convolutional neural networks,” Proc. IEEE Int. Conf. Image Process. (ICIP), pp.2791–2795, 2015.
[7] M. Takagi, D. Higurashi and A. Sakurai, “No-reference image quality assessment based on deep learning with distortion type prediction,” Journal of Japan Society for Fuzzy Theory and Intelligent Informatics, vol.30, no.6, pp.823–831, 2018 (in Japanese).
[8] E.C. Larson and D.M. Chandler, “Most apparent distortion: Full-reference image quality assessment and the role of strategy,” J. Electron. Imaging, vol.19, no.1, 011006, 2010.