ON THE USE OF A CONVOLUTIONAL NEURAL NETWORK TO PREDICT PERCEPTUAL QUALITY OF IMAGES WITHOUT REFERENCE FOR DIFFERENT VIEWING DISTANCES

Aladine Chetouani

PRISME Laboratory,
University of Orleans, Orleans
aladine.chetouani@univ-orleans.fr

Marius Pedersen

Department of Computer Science
Norwegian University of Science and Technology
marius.pedersen@ntnu.no

ABSTRACT

A plethora of image quality metrics have been proposed in the literature. These metrics aims to estimate the perceptual image quality automatically. One important key aspect that the perceived quality is dependent on is the viewing distance from the observer to the image. In this study, we propose to consider this information by estimating the quality of a given image without a reference image for different viewing distances. For that, a Convolutional Neural Network (CNN) model was used in this study. Relevant patches are first selected from the image and they are then used as inputs to the CNN. The selection is here based on saliency information. The used CNN is composed of two outputs that correspond to the predicted subjective scores for two viewing distances (50 cm and 100 cm). Our method was evaluated using the Colourlab Image Database: Image Quality (CID:IQ) that provides subjective scores at two different viewing distances. The obtained results show the efficiency of our method.

Index Terms— Image Quality, Convolutional Neural Network, Patch Selection, Viewing distances

1. INTRODUCTION

Image quality assessment is important in a number of applications such as photography, color printing, etc. The interest in image quality assessment has increase significantly in the last decade. Subjective assessment is still considered to be the "gold standard", but objective assessment is becoming more common. A number of objective assessment methods, commonly known as Image Quality Metrics (IQMs), have been introduced in the literature [1]. These metrics have also been extensively evaluated [2]. Depending on the availability of the reference image, IQMs can be divided into full-reference, reduced-reference, or no-reference. Full-reference requires access to the complete reference, while reduced-reference required partial information of the images, and no-reference does not require access to the reference image.

Traditionally IQMs only incorporated information on the intensity of the distortion, such as peak-signal-to-noise-ratio (PSNR) and Mean-Squared-Error (MSE). These have been used in many applications with success, but they have been shown not to correlate well with perceived image quality for natural images [3]. In the last decade IQMs based on structural similarity [4] have been become popular, and shown in many datasets to correlate better with perceived image quality than PSNR [3]. Other IQMs based on modelling the low-level vision have also been proposed, such the spatial CIELAB (S-CIELAB) [5]. In recent years the use of deep learning has attracted attention of many researchers [6, 7, 8, 9, 10, 11, 12].

One key aspect when observers are evaluating image quality is the viewing distance from the observer to the image [13]. This well-known issue has, however, been overlooked in many of the existing IQMs, and perhaps especially in those based on deep learning. Most existing datasets for estimating the performance of IQMs have only been evaluated by observers at a single viewing distance or the viewing distance has not been controlled. The Colourlab Image Database: Image Quality (CID:IQ) [14] is one of the publicly available datasets where observers have evaluated image quality at different viewing distances.

In this work we use a Convolutional Neural Network (CNN) to predict perceived image quality at different viewing distances. To our knowledge this is the first attempt where the viewing distance is incorporated in a CNN-based IQM.

We first present relevant background, before we introduce the proposed method. Then we present the experimental results, and at last we conclude.

2. BACKGROUND

A number of IQMs based on deep learning have been proposed in the literature. Kang et al. [7] proposed a no-reference metrics that calculated quality for patches in images using CNNs. Bianco et al.[6] proposed a no-reference approach using CNNs, where quality scores are predicted for sub-regions in the image and support vector regression is used on the CNN features. Li et al. [8] extracted simple features from the image using a Shearlet transform, and then treating image quality as a classification problem using deep neural

networks. Chetouani et al. [15] also treated image quality as a classifying problem using linear discriminant analysis. In [16], Chetouani extended the previous work by using a CNN model for degradation identification and quality prediction. Li et al. [17] combined CNNs and Prewitt magnitude on a segmented image to predict image quality. Kim et al. [9] used local quality maps as intermediate targets for CNNs. Lv et al. [10] used a multi-scale Difference of Gaussian to generates features, which were processed using a deep neural network in their no-reference IQM. Gao et al. [11] introduced a full-reference IQM that used deep neural networks to measure the local similarities between the features from the distorted and reference image. Amirshahi et al. [12] proposed a full-reference IQM based on self-similarity and a CNN model. This approach was improved in [18] where the feature maps were compared using traditional IQMs. In this work, we propose to go further than existing CNN-based IQMs by estimating image quality without reference for different viewing distances. To do so, a modified pre-trained CNN model was used. Relevant patches were selected based on saliency information.

3. PROPOSED METHOD

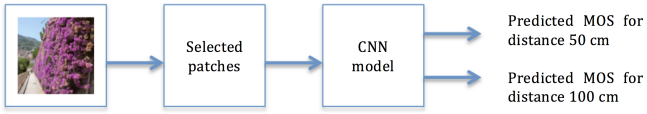

Fig. 1. Flowchart of the proposed method

We propose a CNN-based IQM that allows to predict the quality of a given image for two viewing distances. Fig. 1 presents the general framework of our method. From a given image, we select relevant patches of the image based on saliency information. Then, a CNN model is used to predict the subjective scores for the considering viewing distances.

A plethora of CNN models with different architectures were proposed in the literature. In this work, the model developed by Oxford Visual Geometry Group (VGG) was used, since it is widely used and provides good results in several applications [19]. More precisely, VGG16 was employed. The latter is initially composed of 13 convolutional layers and 3 Fully Connected layers (FC). VGG16 was here modified to match the size of our patches and to adapt the last layers to our context. For that, the three FC layers were replaced by two FC layers of size 128 and 2. The latter corresponds to the number of considered viewing distances. The SoftMax layer (classification task) was also substituted by a regression layer to predict "continuous values". The Mean Square Error function was used as loss function. These modifications allows us to adapt the model to our task and also considerably reduce the number of learnable parameters, since we have now around 14 M of learnable parameters against 138 M initially. It is worth noting that model pre-trained on ImageNet was used by applying fine-tuning. Further, as our visual system is sensitive to rotation [20], no data-augmentation was applied.

The final architecture of our model is presented by the Fig. 2. The input size was fixed to 32x32x3, since the dataset is not big enough to use directly the whole image and different studies highlighted this choice [7, 21].

During the training step, the learning rate and the momentum were fixed to 0.01 and 0.9, respectively. SGD was used as optimization function. The number of epochs and the batch size were set to 25 and 16, respectively. At the end of each epoch, the training data were shuffled and the model was saved. The model that provided the best performance was then retained.

Instead of using all patches, a saliency-based patch selection was applied. From the saliency map of a given image, fixation points are determined using a scanpath predictor [22]. The latter aims to mimic the behavior of human visual system when it faces a real image. The number of fixation points can be modified and any saliency map can be used. In this study, the Graph-Based Visual Saliency (GBVS) method [23] was used and the number of fixation points was fixed to 100. For more details about the patch selection, we refer the reader to [24]. For each determined fixation point, a patch of size 32x32x3 was extracted. Similar to previous studies [7], the overall quality scores for both viewing distances were calculated by averaging the predicted scores of each patch.

4. EXPERIMENTAL RESULTS

4.1. Dataset

As mentioned above, we propose a method that is able to estimate the quality of a given image for different viewing distances. CID:IQ (Colourlab Image Database: Image Quality) [14] is one of the publicly available datasets that considers this aspect. CID:IQ consists of 690 distorted images derived from 23 pristine images. For each distorted image, subjective scores were collected at two different distances (50cm and 100cm). Distorted images were obtained using six types of degradation at five levels: JPEG2000 (JP2K), JPEG, Gaussian Blur (GB), Poisson noise (PN), ΔE gamut mapping (DeltaE) and SGCK gamut mapping (SGCK). A sample of distorted images is presented in Fig 3. The subjective results for the two viewing distances are varying, as an example the different levels of JPEG2000 compression can be differentiated at 50 cm, but this is not the case at 100 cm.

4.2. Protocol and Evaluation criteria

The performance evaluation was done by computing the Pearson (PCC) and Spearman (SROCC) correlation coefficients. These criteria are commonly used in this area and were here

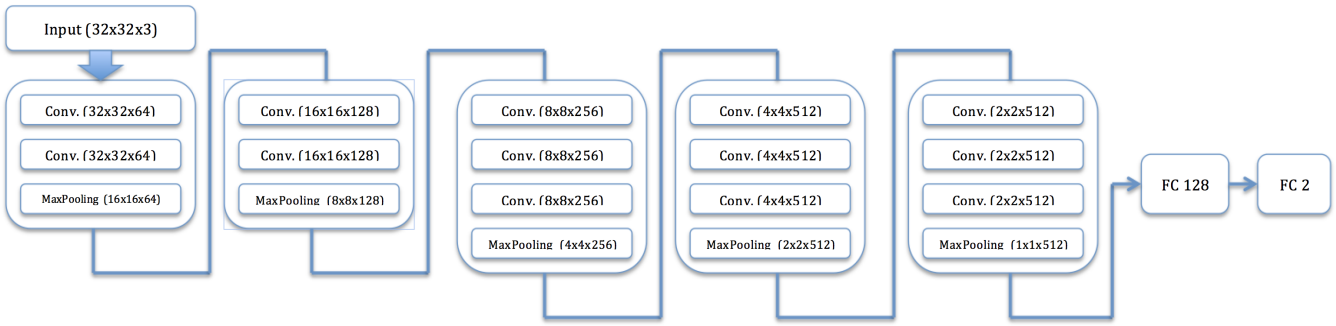

Fig. 2. Architecture of the modified VGG16 model employed in this study

Fig. 3. Samples of the CID:IQ dataset

computed after a nonlinear mapping [25]. The correlation coefficients were computed between subjective scores for both viewing distances and the corresponding predicted values. A correlation equals to 1 means a perfect prediction, while a correlation equals to 0 indicates no correlation.

To train our model, we split the dataset into training-validation and test sets. The test set is formed by one reference image and its degraded versions, while the training-validation set is composed of the rest. The latter was split randomly without overlapping (80% for the training and 20% for the validation). This procedure was repeated 23 times, since the used dataset is composed of 23 original images. This protocol ensures non-overlap or redundancy (in terms of image content) between sets. The correlation coefficients were computed by concatenating the predicted scores.

4.3. Performance Evaluation

The obtained results on the CID:IQ dataset are shown in Tables 1 and 2 for 50 cm and 100 cm, respectively. For both viewing distances, results of our method were compared to the state-of-the-art IQMs (PSNR, SSIM [4], VIF [26], FSIM [27], GMSD [28]). In addition we compare against two no-reference IQMs (BRISQUE [29] and DIVINE [30]). The latter were retrained on CID:IQ using the protocol described in Section 4.2. In [2], an exhaustive evaluation was performed using 60 full-reference metrics. Due to lack of space, we did not incorporate all the results in this paper. The well-know metric Visual Difference Predictor (VDP) [31], that exploits the contrast sensitivity function that integrates the viewing distance, was also compared to our method.

The proposed method outperformed all the compared metrics by more than 20% for 50 cm and 11% for 100cm. It is important to remind that the compared metrics are full-reference IQMs and thus assume that the pristine image is available. All the compared handcrafted metrics obtained low PCC correlations, since the high PCC is equal to 0.713 for 50 cm and 0.773 for 100 cm. The best performances for 50 cm and 100 cm were obtained VIF and FSIM, respectively. No-reference IQMs failed to estimate quality for both viewing distances.

Table 1. Pearson (PCC) and Spearman (SROCC) correlation coefficients between the quality estimation and the subjective scores for 50 cm viewing distance. Highest values in bold.

Metric	PCC	SROCC
PSNR	0.625	0.625
SSIM	0.707	0.761
VIF	0.713	0.711
FSIM	0.678	0.744
GMSD	0.709	0.743
VDP(50)	0.481	0.476
DIVINE	0.227	0.259
BRISQUE	0.499	0.520
Our Method	**0.858**	**0.855**

In Table 3 we show the correlations obtained for each distortion. As excepted, the performances vary according to the distortion type. The best values were obtained for GB, SCGK and DeltaE, while the worst ones were obtained for JP2K and PN. For PN distortion, the results may be due to our visual system because, depending on the viewing distance, the noise is averaged and its perceptual quality seems high, more than the other distortions. For JP2K, the observers are able to dif-

Table 2. Pearson (PCC) and Spearman (SROCC) correlation coefficients between the quality estimation and the subjective scores for 100 cm viewing distance. Highest values in bold.

Metric	PCC	SROCC
PSNR	0.676	0.670
SSIM	0.576	0.638
VIF	0.626	0.622
FSIM	0.773	0.816
GMSD	0.733	0.767
VDP(100)	0.376	0.397
DIVINE	0.225	0.247
BRISQUE	0.444	0.491
Our Method	**0.858**	**0.826**

ferentiate between the level of compression at 50 cm, but not to the same degree at 100 cm, which is a challenge for IQMs.

Table 3. Pearson (PCC) and Spearman (SROCC) correlation coefficients between the quality estimation of our method and the subjective scores for each distortion.

PCC		
Distortion type	**50 cm**	**100 cm**
JP2K	0.748	0.633
JPEG	0.822	0.822
PN	0.786	0.764
GB	0.886	0.905
SCGK	0.915	0.899
DeltaE	0.906	0.873
SROCC		
Distortion type	**50 cm**	**100 cm**
JP2K	0.756	0.582
JPEG	0.800	0.795
PN	0.792	0.747
GB	0.886	0.879
SCGK	0.893	0.879
DeltaE	0.864	0.847

We also computed the correlation between subjective scores of both viewing distances and we compared it to the one obtained by our method (Table 4). We added the correlation for traditional IQMs. The latter obtained a perfect correlation (PCC=1 and SROCC=1), since this kind of metrics provides same quality values whatever the viewing distance. As expected, a high correlation exists between subjective scores (MOS) for both viewing distances. Our method and VDP obtained also a high correlation with no perfect correlation, which indicates that those methods can well integrate this information. However, VDP failed to predict the quality, since it obtained very low correlations (see Tables 1 and 2).

Table 4. Correlations between scores from IQMs of both viewing distances.

	PCC	SROCC
MOS	0.895	0.890
VDP	0.888	0.903
Other traditional metrics	1	1
Our method	0.945	0.935

5. CONCLUSION

In this paper, a new CNN-based blind image quality method that predict subjective scores for two different viewing distances (50 cm and 100 cm) was introduced. To the best of our knowledge, there is no CNN-based method that permits to estimate the quality for different viewing distances. The obtained results were compared to the state-of-the-art methods and it showed its consistency with the subjective judgments. According to the obtained correlations, this work opens interesting perspectives as the performance are not very high (less than 0.90) and can thus be improved.

6. REFERENCES

[1] M. Pedersen and J.Y. Hardeberg, "Full-reference image quality metrics: Classification and evaluation," *Foundations and Trends® in Computer Graphics and Vision*, vol. 7, no. 1, pp. 1–80, 2012.

[2] M. Pedersen, "Evaluation of 60 full-reference image quality metrics on the CID:IQ," in *IEEE International Conference on Image Processing*, 2015, pp. 1588–1592.

[3] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "A comprehensive evaluation of full reference image quality assessment algorithms," in *IEEE International Conference on Image Processing*, 2012, pp. 1477–1480.

[4] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[5] X. Zhang and B.A. Wandell, "A spatial extension of cielab for digital color-image reproduction," *Journal of the Society for Information Display*, vol. 5, no. 1, pp. 61–63, 1997.

[6] S. Bianco, L. Celona, P. Napoletano, and R. Schettini, "On the use of deep learning for blind image quality assessment," *Signal, Image and Video Processing*, vol. 12, no. 2, pp. 355–362, 2018.

[7] L. Kang, P. Ye, Y. Li, and D. Doermann, "Convolutional neural networks for no-reference image quality assessment," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1733–1740.

[8] Y. Li, L-M. Po, X. Xu, L. Feng, F. Yuan, C-H. Cheung, and K-W. Cheung, "No-reference image quality assessment with shearlet transform and deep neural networks," *Neurocomputing*, vol. 154, pp. 94–109, 2015.

[9] J. Kim and S. Lee, "Fully deep blind image quality predictor," *IEEE Journal of selected topics in signal processing*, vol. 11, no. 1, pp. 206–220, 2017.

[10] Y. Lv, G. Jiang, M. Yu, H. Xu, F. Shao, and S. Liu, "Difference of gaussian statistical features based blind image quality assessment: A deep learning approach," in *IEEE International Conference on Image Processing*, 2015, pp. 2344–2348.

[11] F. Gao, Y. Wang, P. Li, M. Tan, J. Yu, and Y. Zhu, "Deepsim: Deep similarity for image quality assessment," *Neurocomputing*, vol. 257, pp. 104–114, 2017.

[12] S. A. Amirshahi, M. Pedersen, and S. X. Yu, "Image quality assessment by comparing CNN features between images," *Journal of Imaging Science and Technology*, vol. 60, no. 6, pp. 60410–1, 2016.

[13] K. Gu, M. Liu, G. Zhai, X. Yang, and W. Zhang, "Quality assessment considering viewing distance and image resolution," *IEEE Transactions on Broadcasting*, vol. 61, no. 3, pp. 520–531, Sep. 2015.

[14] X. Liu, M. Pedersen, and J.Y. Hardeberg, "CID:IQ - A New Image Quality Database," in *Image and Signal Processing*, pp. 193–202. Springer, 2014.

[15] A. Chetouani, A. Beghdadi, and M. A. Deriche, "A hybrid system for distortion classification and image quality evaluation," *Sig. Proc.: Image Comm.*, vol. 27, no. 9, pp. 948–960, 2012.

[16] A. Chetouani, "Convolutional neural network and saliency selection for blind image quality assessment," in *IEEE International Conference on Image Processing*, 2018, pp. 2835–2839.

[17] J. Li, L. Zou, J. Yan, D. Deng, T. Qu, and G. Xie, "No-reference image quality assessment using prewitt magnitude based on convolutional neural networks," *Signal, Image and Video Processing*, vol. 10, no. 4, pp. 609–616, 2016.

[18] S.A. Amirshahi, M. Pedersen, and A. Beghdadi, "Reviving traditional image quality metrics using CNNs," in *Color and Imaging Conference*, 2018, pp. 241–246.

[19] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR, abs/1409.1556*, 2014.

[20] C. S. Furmanski and S. A. Engel, "An oblique effect in human primary visual cortex," *Nature neuroscience*, vol. 3, no. 6, pp. 535, 2000.

[21] S. Bosse, D. Maniry, K. Mller, T. Wiegand, and W. Samek, "Deep neural networks for no-reference and full-reference image quality assessment," *IEEE Transactions on Image Processing*, vol. 27, no. 1, pp. 206–219, Jan 2018.

[22] 0. Le Meur and Liu Z., "Saccadic model of eye movements for free- viewing condition," *Vision Research*, 2015.

[23] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Advances in neural information processing systems*, 2007, pp. 545–552.

[24] A. Chetouani, "A blind image quality metric using a selection of relevant patches based on convolutional neural network," in *European Signal Processing Conference*. IEEE, 2018, pp. 1452–1456.

[25] VQEG, "Final report from the video quality experts group on the validation of objective models of video quality assessment," in *ftp://vqeg.its.bldrdoc.gov/Documents/*, 2000.

[26] Hamid R. Sheikh and Alan C. Bovik, "Image information and visual quality," in *IEEE Transactions on Image Processing*. IEEE, 2006, p. 430444.

[27] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "Fsim: A feature similarity index for image quality assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 2378–2386, Aug 2011.

[28] W. Xue, L. Zhang, X. Mou, and A. C. Bovik, "Gradient magnitude similarity deviation: A highly efficient perceptual image quality index," *IEEE Transactions on Image Processing*, vol. 23, no. 2, pp. 684–695, Feb 2014.

[29] Anush K. Moorthy Anish Mittal and Alan C. Bovik, "No-reference image quality assessment in the spatial domain," in *IEEE Transactions on Image Processing*. IEEE, 2012, vol. 21(12), p. 46954708.

[30] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," in *IEEE Transactions on Image Processing*. IEEE, 2011.

[31] S. J. Daly, "Visible differences predictor: an algorithm for the assessment of image fidelity," *Proc. SPIE 1666, Human Vision, Visual Processing, and Digital Display III*, 1992.