Reviving Traditional Image Quality Metrics Using CNNs

Seyed Ali Amirshahi*†, Marius Pedersen*, Azeddine Beghdadi†
* The Norwegian Colour and Visual Computing Laboratory, Norwegian University of Science and Technology, Gjøvik, Norway
† L2TI, Universite Paris 13, Sorbonne Paris Cite, Villetaneuse France
s.ali.amirshahi@ntnu.no, marius.pedersen@ntnu.no, azeddine.beghdadi@univ-paris13.fr

Abstract

Objective Image Quality Metrics (IQMs) are introduced with the goal of modeling the perceptual quality scores given by observers to an image. In this study we use a pre-trained Convolutional Neural Network (CNN) model to extract feature maps at different convolutional layers of the test and reference image. We then compare the feature maps using traditional IQMs such as: SSIM, MSE, and PSNR. Experimental results on four benchmark datasets show that our proposed approach can increase the accuracy of these IQMs by an average of 23%. Compared to 11 other state-of-the-art IQMs, our proposed approach can either outperform or perform as good as the mentioned 11 metrics. We can show that by linking traditional IQMs and pre-trained CNN models we are able to evaluate image quality with a high accuracy.

Introduction

Over the years a great amount of attention has been paid to evaluating the quality of images both in the case of subjective and objective [1] assessment. Image Quality Metrics (IQMs) are objective methods which aim to evaluate the image quality with the highest correlation possible to perceived image quality. Although a high number of IQMs have been introduced in different studies, there still exists room for improvement mainly with regards to the fact that the performance of IQMs change when evaluated on different datasets and distortions [1, 2]. Other factors such as geometric changes, multiple distortions, run-time performance, and memory requirements have also been mentioned as challenges when introducing new IQMs. When it comes to evaluating the quality of images, subjective quality assessment seems to be the first option. Yet, compared to subjective evaluation, objective IQMs are superior when it comes to time consumption, being consistent, and the fact that they can be used for quality optimization [3].

Different approaches have been taken to objectively evaluate image quality. Among these, structural similarity, color difference, spatial extensions of color difference formulae, simulation of detail visibility, scene statistics, low- and mid-level visual properties, saliency, and machine learning are few of the many [4, 5]. Although IQMs are used to evaluate the general quality of images, some are designed specially to address specific tasks such as printing, displays, spectral imaging, image compression, biometrics, and medical imaging [6, 7].

In this work, we propose the use of traditional IQMs on feature maps of the reference and test images extracted from Convolutional Neural Networks (CNNs). In other words, IQMs are used to compare feature maps extracted from the images using CNN models. This work could be seen as an extension of our previous study in which we compared the strength of feature maps at different convolutional layers using a spatial pyramid approach [8]. In this work, we take a step further and we compare matching feature maps with each other. In recent years different tasks in the field of computer vision and image processing have taken advantage of CNNs resulting in dramatically better performance. Unlike these works, image and video quality metrics have mainly focused on using a limited number of handcrafted features [9]. Using CNNs in our proposed approach allows us to take into account low, mid, and high level features.

Due to the lack of large datasets in the field of image quality assessment we are currently not able to train any full reference IQMs using CNNs. For this reason, we have used pre-trained networks to extract feature maps from the image. Though we use simple IQMs such as MSE, PSNR, SSIM, etc., the proposed scheme outperforms some of the state-of-the-art IQMs. In other words, we will use already available methods to propose new IQMs which increase the accuracy of traditional ones.

This paper is organized as follows: The first section is dedicated to a short review on a number of previous IQMs. In the next Section we introduce our proposed approach, while the experimental results are presented thereafter. Finally, a conclusion of the work is given.

Previous Metrics

There exists a large number of IQMs in the literature. These metrics can be divided into three categories, Full Reference (FR) metrics which we have access to both the reference and test image, Reduced Reference (RR) metrics where we have access to the test image and only partial information on the reference image, and No Reference (NR) metrics that we only have access to the test image, but no information about the reference image. Since the proposed IQM in this work is a FR metric we will focus on introducing few FR IQMs used in this work to compare our proposed approach to. For an in depth review on FR IQMs we refer the reader to [1].

Full Reference Image Quality Metrics for Color Images

A spatial extension of the ΔE^*_{ab} color difference equation is the S-CIELAB metric [10]. To model the human visual system, the images are filtered using contrast sensitivity functions. The ΔE^*_{ab} color difference equation is used to calculate the quality of the image.

Pedersen and Hardeberg [11, 12] introduced the Spatial Hue Angle MEtrics (SHAME and SHAME-II). The mentioned IQMs have a similar framework to S-CIELAB but uses a weighting function based on the hue channel. The images are first filtered using contrast sensitivity functions. A hue angle algorithm is then applied on the image to take into account the systematic errors over the entire image which might be noticeable and unacceptable. Another metric which is based on Adaptive Bilateral Filters (ABF) is also based on the human visual system [13] where they blur the image based on the viewing distance. The final quality score is based on the ΔE^*_{ab} color difference equation. Apart from the mentioned metrics, other IQMs have been proposed based on different color difference equations [14].

The iColor-Image-Difference (iCID) metric [15] is based on SSIM. iCID is designed to enhance the prediction of chromatic distortions such as what is created by gamut-mapping algorithms.

Deep neural networks and CNN based metrics

A big challenge when trying to use CNNs in IQMs is the lack of a dataset large enough to train and test a CNN model. To address this issue, a number of studies have used images uploaded and ranked or scored on photo-sharing websites such as Photo.net, and Flickr to design a NR IQM using a new CNN model.

Kang et al. [16] proposed a NR IQM by calculating the average score of CNN quality estimates for all patches in a given image. Their evaluation showed high correlation coefficients between their objective scores and perceptual scores.

DeepBIQ is another NR IQM which is based on calculating the average quality scores predicted for multiple sub-regions in the image [17]. A Support Vector Regression (SVR) machine is used over CNN features to calculate the quality scores of the mentioned sub-regions.

Li et al. [18] proposed a general-purpose NR IQM using shearlet transform and deep neural networks. Simple features are extracted by the shearlet transform, which is used to describe the behavior of natural images and distorted images. The features are then enhanced to make them more discriminative. By looking at image quality as a classification problem, as done in [19], a classifier is used to estimate image quality.

Li et al. [20] introduced a NR IQM using Prewitt magnitude based on CNNs. The image is first segmented using a graph-based technique. Then the Prewitt operator is used to get the gradient map. The CNN output is weighted by the gradient map.

A blind image evaluator based on a CNN (BIECON) was introduced in [21]. BIECON has two steps; the first is to use a CNN model that is regressed onto local metric score and in the second step pooled features are regressed to the final score.

Lv et al. [22] used local normalized multi-scale Difference of Gaussian (DoG) generated from the distorted images as features. Then by using a deep neural network as a pooling strategy, they introduced a blind IQM.

DeepSim [23] is a FR IQM that uses deep neural networks to measure the local similarities between the features of the reference image and the distorted image. Then local quality indices are gradually pooled to generate an overall image quality value.

We previously introduced a new FR IQM using a CNN model [8]. Our proposed approach was inspired by [24, 25, 26, 27, 28] which self-similarity was calculated for images using a Pyramid Histogram of Orientation Gradient (PHOG) approach [29]. Similar to the PHOG approach, a histogram is calculated for each convolutional layer in the CNN model. The bins in the histogram represents the strength of the feature maps extracted from the image. These bins are not only calculated for each feature map but it is also calculated for the feature maps at different spatial levels. The matching histograms are then compared to one another to first calculate quality scores for each convolutional layer and then the image itself.

Proposed Approach

The proposed approach is based on comparing the feature maps extracted at different convolutional layers of a pre-trained CNN model. In our experiments the AlexNet model [30], which is pre-trained on the ImageNet [31] dataset and implemented in the MatConvNet toolbox [32], was used. We compare feature maps between the test and reference images using IQMs. The assumption is that feature maps of a test image would look similar to the corresponding feature maps in the reference image. Due to the nature of feature maps, it is clear that simple grayscale IQMs which specifically deal with the structure seen in the feature map would provide a better result compared to using complex metrics that try to take into account the content of the image and take advantage of complex handcrafted features. Below you can find a step by step guide on how the proposed approach is calculated:

1. Feature maps from the reference ($\mathscr{I}_R$) and the test ($\mathscr{I}_T$) image are extracted at different convolutional layers.
2. For an arbitrary feature map ($\mathscr{F}(\mathscr{I}_T, n, m)$) in the test image, we calculate
$$m_{IQM}(\mathscr{I}_T, n, m) = \mathrm{IQM}(\mathscr{F}(\mathscr{I}_R, n, m), \mathscr{F}(\mathscr{I}_T, n, m)) \quad (1)$$
which represents the quality score of that feature map in $\mathscr{I}_T$. In Eq. (1), m represents the feature map number in the convolutional layer n (Figure 1). IQM corresponds to the IQM used in our calculations.
3. We calculate the average quality score of all the feature maps in each convolutional layer,
$$m_{IQM}(\mathscr{I}_T, n) = \frac{1}{M}\sum_{m=1}^{M} m_{IQM}(\mathscr{I}_T, n, m) \quad (2)$$
where M corresponds to the total number of feature maps in convolutional layer M (Figure 1).
4. Finally, we calculate the quality score of the test image $\mathscr{I}_T$ using geometrical mean similar to what was done in [8]
$$IQ(\mathscr{I}_T) = \sqrt[N]{\prod_{n=1}^{N} m_{IQM}(\mathscr{I}_T, n)} \quad (3)$$
where N corresponds to the total number of convolutional layers in our CNN model.

As mentioned earlier, due to the nature of feature maps we aim to use traditional gray-scale IQMs. Our aim is to investigate whether the accuracy of simple IQMs could improve by combining them with modern computer vision techniques. In the following we provide a list of the metrics used.

- Structural Similarity Index (SSIM) [33]
- Mean Square Error (MSE)
- Peak Signal to Noise Ratio (PSNR)
- Mean Average Error (MAE)
- Laplacian Mean Square Error (LMSE)
- Normalized Absolute Error (NAE)
- Maximum difference (MD)
- Structural Content (SC)

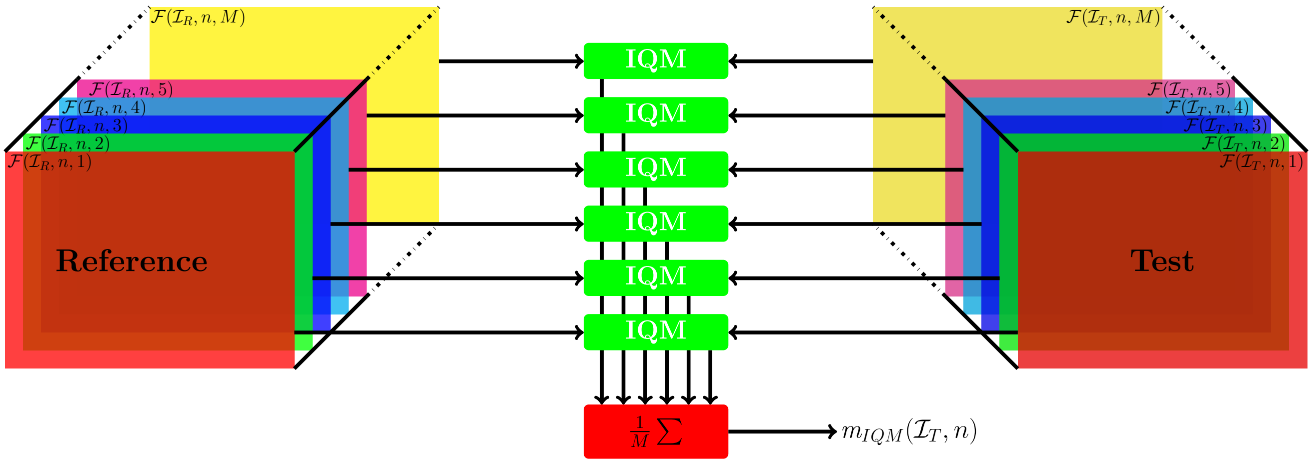


Figure 1. *Pipeline used to calculate image quality at a given convolutional layer n. Feature maps of the test and reference image extracted from the convolutional layers are compared to each other using different IQMs.*

Table 1: Summary of the datasets used in our experiments.

	# reference image	# test image	# distortions	# observers
CID:IQ	23	690	6	17
CSIQ	30	866	6	35
TID2013	25	3000	24	971
LIVE2	29	982	5	–

Experimental Setup and Results

In this section we first describe the datasets used in our experiments and then present the results of our new IQM.

Datasets Used

Over the years different subjective datasets have been collected with the aim to evaluate IQMs [2]. We evaluate the IQMs on four commonly used datasets. Table 1 provides information about the datasets and key information as the number of reference and test images, the number of distortions and observers.

- Tampere Image Database (TID2013) [34]
- Computational and Subjective Image Quality (CSIQ) [35]
- Colourlab Image Database: Image Quality (CID:IQ) [36]
- LIVE Image Quality Assessment Database release 2 (LIVE2) [37, 38]

Results and Discussions

To measure the performance of the proposed approach we calculate the Spearman correlation between the subjective scores provided in each dataset and the objective scores calculated using different IQMs. Using Fishers Z-transform, confidence intervals are calculated giving us a 95% confidence interval for the correlation values. In addition to the Spearman correlation, we also calculated the Pearson and Kendall correlation coefficients. Results showed similar correlation rates close in value and order of performance to that of Spearman, therefore we will only report on the Spearman coefficients.

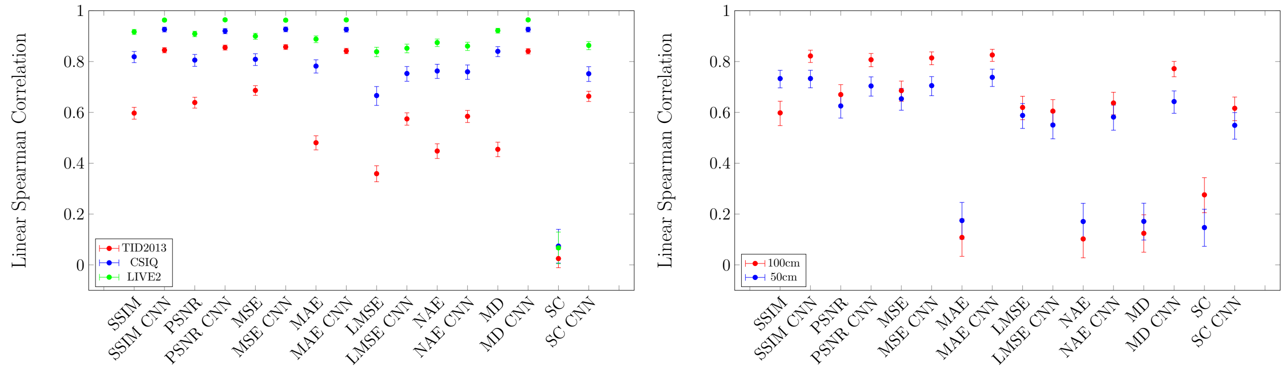
Overall, compared to traditional IQMs the proposed CNN enhanced metrics have a very good performance among different datasets and distortions (Table 2). As mentioned earlier, a drawback of most IQMs is that while they perform well in one particular dataset and/or distortion type this performance is not consistent over all datasets and distortions. This is not the case in our proposed approach where we see a high correlation rate across different datasets and distortions. In TID2013 and CSIQ the performance increases for all IQMs, for LIVE2 all expect NAE has a higher correlation value, and for CID:IQ all except LMSE has a higher correlation value. In the rest of this section we first show how combining traditional IQMs with CNNs improves the performance of the IQMs. We then compare our approach with a few state-of-the-art IQMs and finally assess the performance of our approach with regards to different types of distortions.

Table 2: Difference between the Spearman correlation calculated for the CNN enhanced approach and the original IQMs in different datasets. A positive value indicates that the CNN enhanced approach performs better, while a negative value indicates the opposite.

	TID2013	CSIQ	LIVE2	CID:IQ 50	CID:IQ 100
SSIM	0.25	0.11	0.05	0.00	0.22
PSNR	0.22	0.11	0.05	0.08	0.14
MSE	0.17	0.12	0.06	0.05	0.13
MAE	0.36	0.14	0.07	0.56	0.71
LMSE	0.22	0.09	0.01	-0.04	-0.01
NAE	0.14	0.00	-0.01	0.41	0.53
MD	0.39	0.09	0.04	0.47	0.65
SC	0.64	0.68	0.80	0.40	0.34

In the case of the **TID2013** dataset, the proposed approach outperforms traditional metrics significantly (Figure 2(a)). It is interesting to see the improvement in correlation values when calculating the enhanced CNN metrics. It is also worth noticing that the performance for the CNN enhanced metrics are approximatly similar (above 0.8), and the largest improvement is seen for SC.

For the **CSIQ** dataset (Figure 2(a)), the enhanced metrics are



(a) (b)

Figure 2. *Spearman correlation values for different traditional IQMs and their CNN enhanced approach calculated for the (a) TID2013, CSIQ, and LIVE2 datasets and (b) CID:IQ dataset viewed at a distance of 50 and 100 cm shown with 95% confidence intervals.*

statistically significantly better in all except for NAE. The correlation values are for many of the enhanced metrics higher than 0.9. Compared to the other datasets, the highest correlation rates are observed in the **LIVE2** dataset. Similar to the previous cases the CNN enhanced IQMs show a better performance compared to the original IQMs, except for LMSE and NAE (Figure 2(a)). Similar to the other datasets, the traditional IQMs show mostly lower correlation rates compared to their CNN enhanced metrics in the **CID:IQ** dataset at both a distance of 50 and 100cm (Figure 2(b)).

When comparing our methods to 11 different state-of-the-art metrics (Figure 3(a)), our approach outperforms most metrics for **TID2013**, **CSIQ**, and **LIVE2** datasets. For **LIVE2** many of the CNN enhanced metrics is statistically significantly better than metrics such as SSIM, MSSIM, and more. When comparing the CNN enhanced metrics to the state-of-the-art for **CID:IQ**, our proposed approach outperforms them significantly in both observation distances for most metrics (Figure 3(b)).

In general, in the majority of the datasets, our CNN enhanced IQMs outperform or match state-of-the-art IQMs (Figure 3). From the figures it is clear that the CNN enhanced SSIM, PSNR, MSE, MAE, MD and SC IQMs perform better than the other introduced approaches. Generally, these methods show a better improvement compared to the other methods (Table 2). It is interesting to observe that the same IQMs are also the "go to" metrics when a speedy calculation is needed. From the results we observed that the proposed approach is not only stable across different datasets, but it is rather stable when calculated on images with different types of distortions.

To further investigate the results we also studied the Spearman correlation at different convolutional layers for the CNN enhanced IQMs for different datasets. From the results it is interesting to observe that in the case of the CNN enhanced SC method significantly lower values is seen in the first convolutional layer compared to other layers. Keeping in mind the nature of the feature maps in the first convolutional layer this finding is not surprising. While no other specific patterns can be observed when all the CNN enhanced methods and datasets are taken into account, Amirshahi et. al. [8] showed that as we go deeper in the convolutional layers, in their method, the correlation increased.

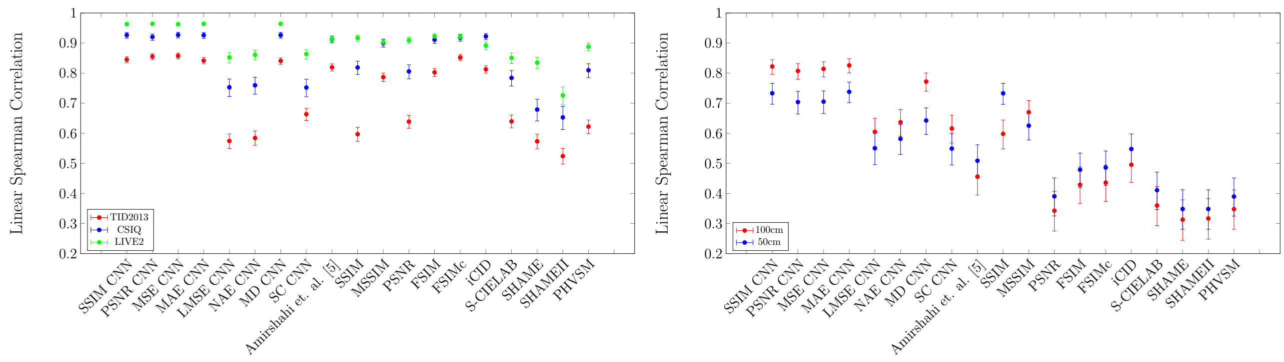
Finally, we study the role of different CNN models on the proposed approach, we also calculated the enhanced CNN IQMs using the VGG model [39] both in the case of VGG16 and VGG19. From the results, it is clear that while the CNN enhanced metrics show a higher correlation compared to the original IQMs, the change of the CNN models does not affect the results in a dramatic manner. This finding is proof on how using CNNs could significantly improve the performance of IQMs while the results are stable across different models. Keeping in mind that the VGG19 model is a deeper network than VGG16 the change in the results show that deeper networks do not necessarily improve the performance of the approach. It is also interesting that no pattern can be observed between the depth of the convolutional layer and its calculated image quality. Keeping in mind the computational costs needed to perform the metric on deeper networks, it is our assessment that the AlexNet model would be a better choice for calculating the proposed approach.

Conclusion

A new approach for calculating FR IQMs using CNNs is introduced in this study. The proposed approach is based on using traditional IQMs such as the MSE, PNSR, and SSIM IQMs to compare the feature maps on a pre-trained CNN model. Results show that by combining classical IQMs with modern computer vision techniques we are able to dramatically improve the performance of the IQMs (on average an increase of 23%). Our proposed approach not only works as good or better than state-of-the-art IQMs, but its performance is similar across different datasets and distortion types. For future work we aim to extend our investigation to other types of CNN models and IQMs. Finding relevant feature maps in the CNN models and giving a higher weight to those features are also another direction we are aiming for.

Acknowledgments

This project has received funding from The Research Council of Norway through a FRIPRO Mobility Grant, contract no 250653/F20. The FRIPRO Mobility grant scheme (FRICON) is co-funded by the European Unions Seventh Framework Programme for research, technological development and demonstration under Marie Curie grant agreement no 608695.



(a) (b)

Figure 3. *Spearman correlation values the enhanced CNN based IQMs along with other state-of-the-art IQMs calculated for the (a) TID2013, CSIQ, and LIVE2 datasets and (b) CDI:IQ dataset viewed at a distance of 50 and 100 cm shown with* 95% *confidence intervals.*

References

[1] Pedersen, M. and Hardeberg, J. Y., "Full-reference image quality metrics: Classification and evaluation," *Foundations and Trends® in Computer Graphics and Vision* **7**(1), 1–80 (2012).

[2] Winkler, S., "Analysis of public image and video databases for quality assessment," *IEEE Journal of Selected Topics in Signal Processing* **6**, 616–625 (2012).

[3] Pedersen, M., Cherepkova, O., and Mohammed, A., "Image quality metrics for the evaluation and optimization of capsule video endoscopy enhancement techniques," *Journal of Imaging Science and Technology* **61**(4), 40402–1 (2017).

[4] Torkamani-Azar, F. and Amirshahi, S. A., "A new approach for image quality assessment using svd," in [*Signal Processing and Its Applications, 2007. ISSPA 2007. 9th International Symposium on*], 1–4, IEEE (2007).

[5] Chetouani, A., Beghdadi, A., and Deriche, M., "Image distortion analysis and classification scheme using a neural approach," in [*Visual Information Processing (EUVIP), 2010 2nd European Workshop on*], 183–186, IEEE (2010).

[6] Pedersen, M., *Image quality metrics for the evaluation of printing workflows*, PhD thesis, University of Oslo (2011).

[7] Amirshahi, S. A. and Torkamani-Azar, F., "Human optic sensitivity computation based on singular value decomposition," *Optica Applicata* **42**(1), 137–146 (2012).

[8] Amirshahi, S. A., Pedersen, M., and Yu, S. X., "Image quality assessment by comparing cnn features between images," *Journal of Imaging Science and Technology* **60**(6), 60410–1 (2016).

[9] Amirshahi, S. A. and Larabi, M.-C., "Spatial-temporal video quality metric based on an estimation of qoe," in [*Quality of Multimedia Experience (QoMEX), 2011 Third International Workshop on*], 84–89, IEEE (2011).

[10] Zhang, X. and Wandell, B. A., "A spatial extension of cielab for digital color-image reproduction," *Journal of the Society for Information Display* **5**(1), 61–63 (1997).

[11] Pedersen, M. and Hardeberg, J. Y., "A new spatial hue angle metric for perceptual image difference," in [*International Workshop on Computational Color Imaging*], 81–90, Springer (2009).

[12] Pedersen, M. and Hardeberg, J. Y., "A new spatial filtering based image difference metric based on hue angle weighting," *Journal of Imaging Science and Technology* **56**(5), 50501–1 (2012).

[13] Wang, Z. and Hardeberg, J. Y., "Development of an adaptive bilateral filter for evaluating color image difference," *Journal of Electronic Imaging* **21**(2), 023021–1 (2012).

[14] Chen, S., Beghdadi, A., and Chetouani, A., "Color image assessment using spatial extension to cie de2000," in [*Consumer Electronics, 2008. ICCE 2008. Digest of Technical Papers. International Conference on*], 1–2, IEEE (2008).

[15] Preiss, J., Fernandes, F., and Urban, P., "Color-image quality assessment: From prediction to optimization," *IEEE Trans. on Image Processing* **23**, 1366–1378 (Mar 2014).

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

[17] Bianco, S., Celona, L., Napoletano, P., and Schettini, R., "On the use of deep learning for blind image quality assessment," *arXiv preprint arXiv:1602.05531* (2016).

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

[19] Chetouani, A., Beghdadi, A., and Deriche, M. A., "A hybrid system for distortion classification and image quality evaluation," *Sig. Proc.: Image Comm.* **27**(9), 948–960 (2012).

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

[21] Kim, J. and Lee, S., "Fully deep blind image quality predictor," *IEEE Journal of selected topics in signal processing* **11**(1), 206–220 (2017).

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

[23] Gao, F., Wang, Y., Li, P., Tan, M., Yu, J., and Zhu, Y., "Deepsim: Deep similarity for image quality assessment,"

Neurocomputing **257**, 104–114 (2017).
[24] Amirshahi, S. A., [*Aesthetic Quality Assessment of Paintings*], Verlag Dr. Hut (2015).
[25] Amirshahi, S. A., Koch, M., Denzler, J., and Redies, C., "PHOG analysis of self-similarity in aesthetic images," in [*IS&T/SPIE Electronic Imaging*], 82911J–82911J, International Society for Optics and Photonics (2012).
[26] Amirshahi, S. A., Redies, C., and Denzler, J., "How self-similar are artworks at different levels of spatial resolution?," in [*Symposium on Computational Aesthetics*], 93–100, ACM (2013).
[27] Braun, J., Amirshahi, S. A., Denzler, J., and Redies, C., "Statistical image properties of print advertisements, visual artworks and images of architecture," *Frontiers in Psychology* **4**(808) (2013).
[28] Redies, C., Amirshahi, S. A., Koch, M., and Denzler, J., "PHOG-derived aesthetic measures applied to color photographs of artworks, natural scenes and objects," in [*Computer Vision–ECCV 2012. Workshops and Demonstrations*], 522–531, Springer (2012).
[29] Bosch, A., Zisserman, A., and Munoz, X., "Representing shape with a spatial pyramid kernel," in [*Proceedings of the 6th ACM international conference on Image and video retrieval*], 401–408, ACM (2007).
[30] Krizhevsky, A., Sutskever, I., and Hinton, G. E., "Imagenet classification with deep convolutional neural networks," in [*Advances in neural information processing systems*], 1097–1105 (2012).
[31] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L., "Imagenet: A large-scale hierarchical image database," in [*Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*], 248–255, IEEE (2009).
[32] Vedaldi, A. and Lenc, K., "Matconvnet: Convolutional neural networks for matlab," in [*Proceedings of the 23rd ACM international conference on Multimedia*], 689–692, ACM (2015).
[33] Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P., "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing* **13**(4), 600–612 (2004).
[34] Ponomarenko, N., Jin, L., Ieremeiev, O., Lukin, V., Egiazarian, K., Astola, J., Vozel, B., Chehdi, K., Carli, M., Battisti, F., et al., "Image database tid2013: Peculiarities, results and perspectives," *Signal Processing: Image Communication* **30**, 57–77 (2015).
[35] Larson, E. C. and Chandler, D. M., "Most apparent distortion: full-reference image quality assessment and the role of strategy," *Journal of Electronic Imaging* **19**(1), 011006–011006 (2010).
[36] Liu, X., Pedersen, M., and Hardeberg, J. Y., "CID:IQ–a new image quality database," in [*International Conference on Image and Signal Processing*], 193–202, Springer International Publishing (2014).
[37] Sheikh, H. R., Sabir, M. F., and Bovik, A. C., "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Transactions on image processing* **15**(11), 3440–3451 (2006).
[38] Sheikh, H. R., Wang, Z., Cormack, L., and Bovik, A. C., "Live image quality assessment database release 2," (2005).
[39] Simonyan, K. and Zisserman, A., "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556* (2014).

Author Biography

Seyed Ali Amirshahi is a Marie Curie post-doctoral Fellow in the at the Norwegian University of Science and Technology (NTNU) and a visiting researcher at the L2TI lab at University Paris 13. His research is focused on image quality assessment and computational aesthetics. He received his PhD from the Friedrich Schiller University of Jena in Germany (2015). Prior to joining NTNU, he was a post-doctoral Fellow at the International Computer Science Institute in Berkeley.

Marius Pedersen is professor at the Norwegian University of Science and Technology. His work is centered on image quality assessment; he has more than 60 publications in this field. He received his PhD in color imaging (2011) from the University of Oslo. He is currently the head of the computer science group in Gjøvik in the department of computer science, and the head of the Norwegian Colour and Visual Computing Laboratory, both at NTNU.

Azeddine Beghdadi is Professor at University Paris 13, Sorbonne Paris Cit. He is a founding member of the Laboratory of Information Processing and Transmission (L2TI) and was its director from 2010 to 2016. His research interests include image quality enhancement and assessment, image and video compression, bio-inspired models for image analysis and processing, and physics-based image analysis. He received his PhD from University Pierre et Marie Curie (Paris 6). He published more than 260 international refereed scientific papers.