

Saliency-based deep blind image quality assessment

Kamal Lamichhane¹, Marco Carli¹, and Federica Battisti²

¹Department of Engineering, University of Roma Tre, Roma, Italy (name.lastname@uniroma3.it)

²Università degli studi di Padova, Padova, Italy, federica.battisti@unipd.it

Abstract

Assessing the quality of images is a challenging task. To achieve this goal, the images must be evaluated by a pool of subjects following a well-defined assessment protocol or an objective quality metric must be defined. In this contribution, an objective metric based on neural networks is proposed. The model takes into account the human vision system by computing a saliency map of the image under test. The system is based on two modules: the first one is trained using normalized distorted images. It learns the features from the original and the distorted images and the estimated saliency map. Furthermore, an estimate of the prediction error is performed. The second module (non-linear regression module) is trained with the available subjective scores. The performances of the proposed metric have been evaluated by using state of the art quality assessment datasets. The achieved results show the effectiveness of the proposed system in matching the subjective quality score.

Introduction

Image quality is an important factor that may influence the use of information conveyed by an image. We are witnessing the proliferation of multimedia transmission services in which video and images are the main media for sharing information. Despite huge advances in both hardware and software, it is worth remembering that imaging devices, storage systems, encoding algorithms, or the transmission systems themselves can create artifacts that affect the image quality. The ability to reliably measure the final quality is a key element in the development of such systems.

In the literature, many authors have faced this problem, trying to define a multipurpose quality measure that match, as closely as possible, to the human judgment [1–5]. These methods can be classified according to the possibility of using original image information during the evaluation phase. Full Reference Image Quality Assessment (FRIQA), Reduced Reference Image Quality Assessment (RRIQA), or No Reference Image Quality Assessment (NRIQA) methods respectively exploit the availability of the original image [4, 6, 7], some of its features, or any information.

NRIQA is performed either using conventional methods or the Convolutional Neural Network (CNN). Conventional NRIQA [2, 8, 9] depends on the computation of handcrafted [10] or low level features from a distorted image and performing classification or regression operations [3]. Innovative approaches based on Deep Neural Network (DNN) have recently been proposed. A DNN is an artificial neural network inspired by human biology that seeks to mimic the way neurons in the human brain process inputs from human senses. In particular, DNN are machine learning algorithms representing neural networks with

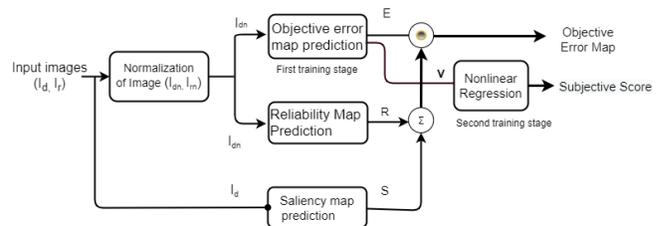


Figure 1: Block diagram of the proposed (DIQAS) model

many hidden layers. CNN, a class of DNN, automatically and adaptively learns spatial hierarchies of features through back propagation using multiple building blocks, such as convolution layers, pooling layers, and fully connected layers.

In this work, we propose a no reference image quality assessment technique by exploiting information on the saliency of the distorted image. In the following, the proposed model is denoted as Deep Image Quality assessment based on Saliency (DIQAS). This model can be considered as an extension of the DIQA model proposed in [5]. The saliency map of the distorted image is generated by extracting the foreground object in the image and it represents the most influential feature of the visually sensitive area of the distorted image. It is worth noticing that the available datasets used to evaluate the performance of blinded image quality assessment algorithms have a very limited number of distorted images. For this reason, deep learning models often suffer from overfitting during the training process.

To overcome this issue, the deep learning system is divided into primary and secondary training modules. The primary training module is called objective error map prediction (OEMAP) while the secondary training module is called non-linear regression module. OEMAP is used as an objective quality measure of the distorted image. The saliency map is used in the back-propagation phase to to modify the loss function parameters. This stage helps to minimize the limitation of small size dataset by minimizing the overfitting problem of the deep learning system. The secondary training module, designed using fully connected layers, is used to generate the estimate the subjective quality scores. The predicted scores are compared, by using SROCC and PLCC, to the subjective Mean Opinion Scores available in the dataset. LIVE_IQA, CSIQ and TID2013 are the three datasets used to train and test our model.

Proposed Model

As mentioned in the previous section, a deep blind image quality assessment technique based on saliency information (DIQAS) is proposed. The block diagram of the proposed model is shown in Figure 1. In the following, each block of DIQAS is

described.

Deep CNN model architecture

The convolutional network [5, 11] used in the **DIQAS** is shown in Figure 2. This unit implements the objective error map prediction and nonlinear regression modules shown in Figure 1. The CNN can be divided in 3 parts. The first section consists of 8 different convolution layers designed using different parameters such as 3x3 kernel size, the unity batch size in the input layer, the 2X2 size of the max pooling strides in the second and fourth layer and different feature channels as shown in the Figure. Symbols $B : 1$, $S : 2 \times 2$, and Conv_3x3 represent the batch size, the stride size for max pooling, and the convolution kernel size. The rectifier linear unit (Relu) is the activation function adopted in each convolutional layer. This complete set of layers is symbolically represented as $F(\cdot)$. $F(\cdot)$ is also considered as the feature map and it is used also for second stage training. In the last layer of first stage training, the output of primary block $F(\cdot)$ is fed to a convolution layer $G(\cdot)$ having kernel size 1x1, feature channel 1, and a linear activation function. Thus, it performs the linear combinations of the feature maps to generate the error map (E). The size of the output of $G(\cdot)$ is 1/4 times the input image. The downscaling size of the model output is determined by the max-pooling layers (strides) and its size. Meanwhile, the ground truth error map of the distorted image is also scaled to the same size. Finally, the feature map $F(\cdot)$ is fed into the Global Average Pooling (GAP) layer, followed by the three fully connected layers with one hundred twenty-eight features channels in the first two and sixty-four features channels in the last layers. These fully connected layers are represented as $H(\cdot)$.

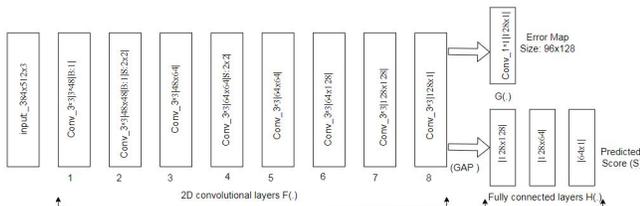


Figure 2: Architecture of CNN used as a sub-network to predict the objective error and subjective score of the input distorted image.

Image normalization

In this preprocessing stage the RGB input image is converted to the grayscale image and then subtracted the low pass filtered image obtained from the following three-stages function: to obtain a blurred version of the grayscale image, downsize it by 1/4 and then to upscale of the image to the original size. Let I_d be the distorted input image, I_{dg} the grayscale image obtained from I_d , and I_{dl} the corresponding low pass filtered image. The normalized form of the input image (I_{dn}) [3, 13] is defined as:

$$I_{dn} = |I_{dg} - I_{dl}| \quad (1)$$

The normalization phase is performed for two main reasons. First, distortion present in the image hardly impact on the lower frequency band. Similarly, the human visual system (HVS) is less sensitive at the low-frequency band of the image [13]. Hence, the image normalization process helps to remove the unnecessary information contained in the input image that may impact the CNN

learning process during the training stage. Likewise, it helps improve the efficiency of the model by reducing the overall computational cost and increasing the prediction accuracy. However, there is a drawback of losing some information contained in the input image. To compensate for this, the saliency map of the distorted image obtained from the foreground object detection method is combined with the reliability map, which is used to improve the objective error map obtained from the first training phase.

Reliability map prediction

Most distortion methods like JPEG and JPEG2000 compression, Gaussian blur, etc. may cause image blurring. It is difficult to determine if the blurry region is due to distortion or not without knowing the reference image [5]. Similarly, in some images, severe distortion is applied during the creation of the dataset. When severe distortion is applied to an image, its error map (difference between the normalized distorted and the reference image) receives the higher frequency component. Meanwhile, the distorted image loses high frequency detail. Blurry and severely distorted images could generate a significant portion of image having the nearly same range of pixels and we may consider it as a homogeneous region. So the objective error map prediction model designed with multiple convolutional layers will likely fail to predict the homogeneous region of the input image. To avoid the aforementioned problem, the reliability of the predicted error map E_{pred} is estimated by measuring the texture strength of the distorted image. Thus, a reliability map R is defined by:

$$R = \frac{2}{1 + \exp(-\alpha(I_{dn}))} - 1, \quad (2)$$

where, α is a constant used to control the saturation property (whiteness) of the reliability map. Equation 2 represents a modified sigmoid function: only the positive part of it is considered to obtain the normalized form of the reliability map. To minimize the unnecessary influence of the reliability map on objective error map, it is further divided by its average value as defined in equation 3.

$$R_n = \frac{R}{\frac{1}{H_R W_R} \sum_{i,j} R_{i,j}}, \quad (3)$$

where, H_R and W_R are the shape parameters of the reliability map. The reliability map returns zero values when there is a zero pixel weight or no spatial information at I_{dn} image. Similarly, since the reliability map is obtained from the preprocessed distorted image I_{dn} , it can not represent complete information of the input image. The saliency map prediction method detects the most bright and significant area of the image with unnoticeable blur and distortion during human observation. Hence, the obtained saliency map in normalized form is used to adjust the pixel weight of the corresponding spatial location in R_n , resulting in a more reliable reliability map.

Saliency map prediction

The saliency map is used to identify the most influential features present in an image. In [14], researchers compared the performances of general IQA methods with Saliency based IQAs have been compared. The analysis shows that Saliency based IQAs offer the best match to the human judgement. To predict

the saliency map of the foreground object, in the first step, individual pixel similarity is measured to the background of the image using a Histogram back-projection [15] technique. Here, as a pre-processing stage, the input image is filtered by using multi-scale pyramidal [16] mean shift filter to generate a smooth image by grouping close pixels together. This stage is also considered as a initial segmentation stage or abstraction. In the next step, back-projection of hue and saturation histogram of the image, on the preprocessed image, is performed. Finally, backprojected image is normalized with mean shift operation; its contrast is enhanced using histogram equalization and inverted the image to obtain the initial stage of the saliency map. To refine the salient regions of the image, bounding boxes are used as initial estimate of the foreground object and the grabcut is applied [17]. Its output is represented as Saliency map S .

Learning process of objective error map

The training of the objective error map prediction module (OEMAP) is the first training of the **DIQAS** model. To achieve the effect of the training process, the objective error map image is represented as the objective quality metric. Regarding the regression of the model, the required loss function L_1 is defined by the mean squared error between the product of the difference between predicted and ground truth error map and the integrated image of reliability map with saliency map. Its general expression is defined in equation 4. θ_F and θ_G are the hyper-parameters of the CNN model for the respective convolutional layers and E_{gt} is a ground truth error map [5] defined by equation 5.

$$L_1(I_{dn}; \theta_F, \theta_G) = \| (G(F(I_{dn}; \theta_F), \theta_G) - E_{gt}) \odot (R+S) \|_2^2 \quad (4)$$

$$E_{gt} = |I_{rn} - I_{dn}|^p, \quad (5)$$

Here, p is a constant value set to $0 < p \leq 1$. For $p = 1$, E_{gt} could result in a very small value of pixel weights, even zero values. Therefore, it might not represent the exact ground truth error map. To properly tune the ground truth error map of a distorted image, the recommended value of p is 0.2 [5].

Nonlinear Regression and learning subjective score

The nonlinear regression module is a secondary training module and is used to predict the subjective score of the image quality. Subjective score defines the predicted score obtained from the nonlinear regression module. MOS or $DMOS$ available in the dataset is considered as a ground truth score. In non linear regression module, the output of the trained subnetwork $F(\cdot)$ is connected to the global average pooling (GAP) layers followed by the fully connected layers as shown in fig2. As the training of the *CNN* is a state-full operation, all the information contained in $F(\cdot)$ is reflected in input of GAP. The output of GAP function (V), hyperparameters of fully connected layers θ_H and ground truth subjective score S_{gt} are used to define the loss function required during training of the nonlinear regression module. The corresponding loss function used to minimize the error during regression of subjective score is defined as:

$$L(I_d; \theta_F, \theta_H) = \| (H(V); \theta_H) - S_{gt} \|_2^2, \quad (6)$$

In (6), V is an output of *GAP* and mathematically, it is a function of output of $F(\cdot)$ and model parameters Θ_F as represented in (7).

$$V = GAP(F(I_{dn}; \Theta_F)) \quad (7)$$

Experiments and Analysis

Datasets

To evaluate the performance of the **DIQAS** model Live_IQA [18], TID2013 [19], and CSIQ [20] datasets are used.

Evaluation Metrics

To evaluate the performance of the proposed model, Spearman rank-order correlation coefficient (*SROCC*) and Pearson linear correlation coefficient (*PLCC*) are used. Let S_g be the ground truth score and S_p the predicted score. The general expression for *SROCC* and *PLCC* as described in [21] are represented in (8), (9). In equation 8, D represents the difference between the corresponding rank variables of the ground truth score and the respective predicted score considered to measure the value of the required metrics. Similarly n in (8) and (9) represents the number of array or tensor elements or image numbers considered for evaluating performance metrics.

$$SROCC(S_g, S_p) = 1 - \frac{6 \sum_i D_i^2}{n(n^2 - 1)} \quad (8)$$

$$PLCC(S_g, S_p) = \frac{\sum_i (S_{pi} - \mu_{Sp})(S_{gi} - \mu_{Sg})}{\sqrt{\sum_i (S_{pi} - \mu_{Sp})^2} \sqrt{\sum_i (S_{gi} - \mu_{Sg})^2}} \quad (9)$$

Ablation Studies

The proposed **DIQAS** model presents the best result once the variable parameters of different modules are perfectly tuned. Multiple testing of the algorithms with different parameters are carried out to fine-tune the network.

Pretraining with OEMAP

As a first step of testing of the **DIQAS** model we evaluated the effect of pretraining on OEMAP. TABLE 1 illustrates the effects of number of epochs on the performance of the model. To compare the result, *SROCC* and *PLCC* values of subjective score obtained from second stage training after training of first stage with five different number of epochs for all three datasets are shown. From the results, in LIVE_IQA 60 epochs of pretraining offers the best result. Whereas in TID2013, it offers the best result in 10 number of epochs and CSIQ dataset offers it in 40 epochs of pretraining in the first stage. We further evaluated the LIVE_IQA dataset up to 80 epochs of training in first stage and confirmed that the results are more stable after 60 epochs of training. *SROCC* and *PLCC* values for 80 epochs of training in first stage were 0.974 and 0.985 respectively.

Impacts of saliency map

The impact of saliency map module used in **DIQAS** model is measured by evaluating *SROCC* and *PLCC* value in all datasets. Overall result with and without considering saliency map is presented on TABLE 2. From the results, Saliency map helps to improve the performance of the model for all datasets.

Table 1: SROCC and PLCC comparison for different epochs at first stage on LIVE_IQA, CSIQ and TID2013 datasets

Epochs	LIVE_IQA		CSIQ		TID2013	
	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC
	1	0.965	0.972	0.951	0.952	0.872
10	0.970	0.979	0.960	0.960	0.874	0.894
20	0.972	0.982	0.961	0.960	0.873	0.891
40	0.974	0.982	0.962	0.961	0.872	0.890
60	0.975	0.985	0.960	0.957	0.871	0.890

Table 2: SROCC and PLCC comparison with and without using saliency map on LIVE_IQA, TID2013 and CSIQ datasets

Saliency map	LIVE_IQA		TID2013		CSIQ	
	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC
	Yes	0.975	0.985	0.874	0.894	0.962
No	0.971	0.979	0.873	0.892	0.956	0.954

Experimental setup and results

The implementation of DIQAS model and its performance evaluation is carried out in the python environment using tensorflow. Numpy, opencv and SciPy are additional library functions used to implement the algorithm of DIQAS model. To simplify the implementation of the model, some basic functions required for image processing are extracted from Image quality assessment (IQA) ¹ library.

To train the model the loss function defined by equation 6 is used and to perform the backpropagation step, automatic differentiation function² defined as a gradient in TensorFlow is used. Similarly, to optimize the weight parameters during the backpropagation phase, Nadam optimizer [22] is used. The use of linear activation function in the last layer of the 2D convolutional layer generates the required error map image for the input image. Stride having size $[2 \times 2]$ is a max-pooling filter used in layer 2 and layer 4 of the CNN as mentioned in 2. The output of $F(\cdot)$ is fed to the GAP and the obtained result is applied to the fully connected network $H(\cdot)$. *Usually, fullyconnectednetworkformedbyfullyconnectedlayersispronetoovertfitting.*

Performance on same dataset

To train and test the DIQAS model, we randomly divided the reference images and distorted images into two subsets having 80% for training purpose and 20% for testing purpose from original dataset without overlapping one subset to other subset. This is performed for all three datasets LIVE_IQA, TID2013 and CSIQ. We compared our model performance with 7 different NR_IQA methods (BIECON [23], CNN [24], DeepIQA [25], DIQA [5], DIQA_BASE [5] MGDNN [26] and SESANIA [27]). The performance comparison is reported in TABLE 3.

Performance on crossdataset test

To further evaluate the performance of the DIQAS model, we used one dataset for training and the other two for testing. The results, evaluated in terms of SROCC and PLCC, are shown in TABLE 4.

In [28] cross-dataset performance of two NRIQA methods BRISQUE and BIECON is presented in terms of PLCC. Based on those results, a comparative results including DIQAS model

¹<https://github.com/ocampor/image-quality>

²<https://www.tensorflow.org/guide/autodiff>

Table 3: SRCC and PLCC comparison on three datasets

Model	LIVE_IQA		TID2013		CSIQ	
	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC
BIECON	0.958	0.962	0.721	0.765	0.825	0.838
CNN	0.956	0.963	–	–	–	–
DeepIQA	0.960	0.972	–	–	–	–
DIQA_BASE	0.963	0.964	0.800	0.803	0.812	0.791
DIQA	0.975	0.977	0.825	0.850	0.884	0.915
MGDNN	0.951	0.949	–	–	–	–
SESANIA	0.934	0.948	–	–	–	–
DIQAS	0.975	0.985	0.874	0.894	0.962	0.961

Table 4: SRCC and PLCC on cross datasets

Train	Test	SROCC	PLCC
LIVE_IQA	TID2013	0.49	0.515
	CSIQ	0.69	0.705
TID2013	LIVE_IQA	0.825	0.81
	CSIQ	0.675	0.735
CSIQ	LIVE_IQA	0.64	0.825
	TID2013	0.45	0.529

is presented in 5. The two best results given by three different models are shown in bold. From the results it can be seen that the DIQAS model was one of the best two models.

Conclusion

The DIQAS model is an efficient model designed using DNN-based 2-stage training modules. The first stage training module is an objective error map prediction module based on normalized image obtained by preprocessing a distorted image. Its performance is measured using an objective metric and is also used as a proxy regression target. The use of saliency map helps improve the learning of the first training phase. The secondary training module, designed using fully connected layers, is used to predict subjective scores. Performance of the DIQAS model is measured in terms of SROCC and PLCC on LIVE_IQA, TID2013 and CSIQ datasets. The achieved result show the effectiveness of the proposed model.

References

- [1] Prachi R Rajarapolu and Vijay R Mankar, “Bicubic interpolation algorithm implementation for image appearance enhancement,” *International Journal of Computer Science And Technology*, vol. 8, no. 2, 2017.
- [2] Michele A Saad, Alan C Bovik, and Christophe Charrier, “Blind image quality assessment: A natural scene statistics approach in the det domain,” *Transactions on Image Processing*, vol. 21, no. 8, pp. 3339–3352, 2012.
- [3] Jongyoo Kim and Sanghoon Lee, “Fully deep blind image quality predictor,” *Journal of selected topics in signal processing*, vol. 11, no. 1, pp. 206–220, 2016.
- [4] Jongyoo Kim and Sanghoon Lee, “Deep learning of human visual

Table 5: PLCC comparison of cross dataset test

Train	Test	BRISQUE	BIECON	DIQAS
LIVE_IQA	TID2013	0.494	0.506	0.515
	CSIQ	0.689	0.744	0.705
TID2013	LIVE_IQA	0.798	0.859	0.81
	CSIQ	0.692	0.660	0.735
CSIQ	LIVE_IQA	0.889	0.761	0.825
	TID2013	0.528	0.482	0.529

- sensitivity in image quality assessment framework,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1676–1684.
- [5] Jongyoo Kim, Anh-Duc Nguyen, and Sanghoon Lee, “Deep cnn-based blind image quality predictor,” *Transactions on neural networks and learning systems*, vol. 30, no. 1, pp. 11–24, 2018.
- [6] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli, “Image quality assessment: from error visibility to structural similarity,” *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [7] Lin Zhang, Lei Zhang, Xuanqin Mou, and David Zhang, “Fsim: A feature similarity index for image quality assessment,” *IEEE transactions on Image Processing*, vol. 20, no. 8, pp. 2378–2386, 2011.
- [8] Anush Krishna Moorthy and Alan Conrad Bovik, “Blind image quality assessment: From natural scene statistics to perceptual quality,” *IEEE transactions on Image Processing*, vol. 20, no. 12, pp. 3350–3364, 2011.
- [9] Wufeng Xue, Xuanqin Mou, Lei Zhang, Alan C Bovik, and Xiangchu Feng, “Blind image quality assessment using joint statistics of gradient magnitude and laplacian features,” *IEEE Transactions on Image Processing*, vol. 23, no. 11, pp. 4850–4862, 2014.
- [10] Loris Nanni, Stefano Ghidoni, and Sheryl Brahmam, “Handcrafted vs. non-handcrafted features for computer vision classification,” *Pattern Recognition*, vol. 71, pp. 158–172, 2017.
- [11] Karen Simonyan and Andrew Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.
- [12] Richard A Haddad, Ali N Akansu, et al., “A class of fast gaussian binomial filters for speech and image processing,” *IEEE Transactions on Signal Processing*, vol. 39, no. 3, pp. 723–727, 1991.
- [13] Scott J Daly, “Visible differences predictor: an algorithm for the assessment of image fidelity,” in *Human Vision, Visual Processing, and Digital Display III*. International Society for Optics and Photonics, 1992, vol. 1666, pp. 2–15.
- [14] Edoardo Ardizzone and Alessandro Bruno, “Image quality assessment by saliency maps..” in *Proceedings of the International Conference on Computer Vision Theory and Applications, VISAPP*, 2012.
- [15] Michael J Swain and Dana H Ballard, “Color indexing,” *International Journal of Computer Vision*, vol. 7, no. 1, pp. 11–32, 1991.
- [16] PJ Burt, “Fast filter transforms for image processing,” *Multiresolution Image Processing and Analysis (A. Rosenfeld, Ed.)*, Springer-Verlag, Berlin/New York, 1983.
- [17] Carsten Rother, Vladimir Kolmogorov, and Andrew Blake, ““Grab-cut” interactive foreground extraction using iterated graph cuts,” *ACM Transactions on Graphics*, vol. 23, no. 3, pp. 309–314, 2004.
- [18] HR Sheikh, “Live image quality assessment database release 2,” <http://live.ece.utexas.edu/research/quality>, 2005.
- [19] Nikolay Ponomarenko, Lina Jin, Oleg Ieremeiev, Vladimir Lukin, Karen Egiazarian, Jaakko Astola, Benoit Vozel, Kacem Chehdi, Marco Carli, Federica Battisti, et al., “Image database tid2013: Peculiarities, results and perspectives,” *Signal Processing: Image Communication*, vol. 30, pp. 57–77, 2015.
- [20] Eric Cooper Larson and Damon Michael Chandler, “Most apparent distortion: full-reference image quality assessment and the role of strategy,” *Journal of electronic imaging*, vol. 19, no. 1, pp. 011006.1–21, 2010.
- [21] ITU, “Final report from the video quality experts group on the validation of objective models of video quality assessment, phase ii (fr-tv2),” <https://www.itu.int/md/T01-SG09-C-0060>, Accessed: 2020-05-29.
- [22] Timothy Dozat, “Incorporating Nesterov Momentum into Adam,” *International Conference on Learning Representations*, 2016.
- [23] Heeseok Oh, Sewoong Ahn, Jongyoo Kim, and Sanghoon Lee, “Blind deep s3d image quality evaluation via local to global feature aggregation,” *IEEE Transactions on Image Processing*, vol. 26, no. 10, pp. 4923–4936, 2017.
- [24] Le Kang, Peng Ye, Yi Li, and David Doermann, “Convolutional neural networks for no-reference image quality assessment,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014.
- [25] Sebastian Bosse, Dominique Maniry, Thomas Wiegand, and Wojciech Samek, “A deep neural network for image quality assessment,” in *IEEE International Conference on Image Processing (ICIP)*. IEEE, 2016.
- [26] Yaqi Lv, Gangyi Jiang, Mei Yu, Haiyong Xu, Feng Shao, and Shanshan Liu, “Difference of gaussian statistical features based blind image quality assessment: A deep learning approach,” in *IEEE International Conference on Image Processing (ICIP)*. IEEE, 2015.
- [27] Yuming Li, Lai-Man Po, Xuyuan Xu, Litong Feng, Fang Yuan, Chun-Ho Cheung, and Kwok-Wai Cheung, “No-reference image quality assessment with shearlet transform and deep neural networks,” *Neurocomputing*, vol. 154, pp. 94–109, 2015.
- [28] Dan Yang, Veli-Tapani Peltoketo, and Joni-Kristian Kamarainen, “Cnn-based cross-dataset no-reference image quality assessment,” in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2019.

Author Biography

Kamal Lamichhane is a Ph.D. student at Roma TRE University, Rome, Italy. From 2004 to 2007, he worked as a teaching assistant at the Institute of Engineering, western regional Campus, and assistant lecturer at Pokhara Engineering College, Pokhara, Nepal. Since 2007 he has been a Telecom Engineer at Nepal Telecom, Incumbent Telecom Operator of Nepal. His current research fields include multimedia communication and image quality assessment.

Marco Carli received the Ph.D. degree from the Tampere University of Technology, Tampere, Finland. He is an Associate Professor with the Department of Engineering, Università degli Studi Roma TRE. He was a Visiting Researcher with the Image Processing Laboratory, University of California, Santa Barbara, USA, from 2000 to 2004. His research interests are in the area of digital signal and image processing with applications to multimedia communications, digital watermarking, multimedia quality evaluation, and information security.

Federica Battisti received the Ph.D. degree from Roma Tre University, Rome, Italy, in 2010. She is currently assistant professor at the Department of Information Engineering at University of Padua. Her research interests are in multimedia quality assessment and security. She is IEEE Senior member, and she serves as Associate Editor of *IEEE Transactions on Multimedia*, of *EURASIP Journal on Image and Video Processing*, and of *ELSEVIER Signal Processing: Image Communication*.

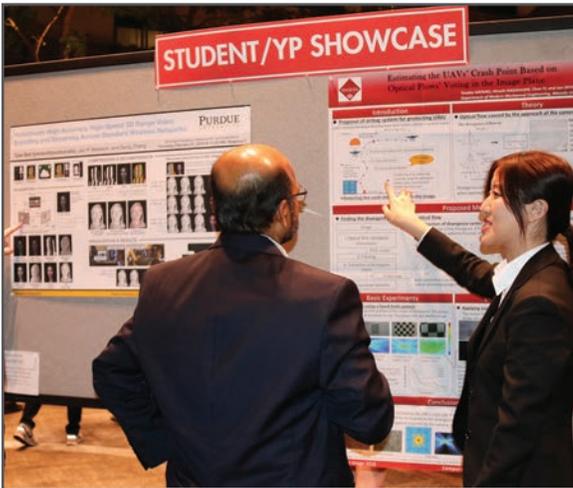
JOIN US AT THE NEXT EI!

IS&T International Symposium on

Electronic Imaging

SCIENCE AND TECHNOLOGY

Imaging across applications . . . Where industry and academia meet!



- **SHORT COURSES • EXHIBITS • DEMONSTRATION SESSION • PLENARY TALKS •**
- **INTERACTIVE PAPER SESSION • SPECIAL EVENTS • TECHNICAL SESSIONS •**

www.electronicimaging.org

