

## Circular-ELM for the reduced-reference assessment of perceived image quality

Sergio Decherchi <sup>a</sup>, Paolo Gastaldo <sup>b,\*</sup>, Rodolfo Zunino <sup>b</sup>, Erik Cambria <sup>d</sup>, Judith Redi <sup>c</sup>

<sup>a</sup> Department of Drug Discovery and Development—Fondazione Istituto Italiano di Tecnologia (IIT), Morego, Genova, Italy

<sup>b</sup> Department of Naval, Electric, Electronic and Telecommunications Engineering, DITEN, University of Genoa, Via Opera Pia 11a, 16145, Genova, Italy

<sup>c</sup> Department of Intelligent Systems, Delft University of Technology, Mekelweg 4, 2628CD Delft, The Netherlands

<sup>d</sup> Temasek Laboratories, National University of Singapore, 5A Engineering Drive 1, Singapore 117411, Singapore

### ARTICLE INFO

#### Key words:

Extreme learning machine  
Circular backpropagation  
Image quality assessment

### ABSTRACT

Providing a satisfactory visual experience is one of the main goals for present-day electronic multi-media devices. All the enabling technologies for storage, transmission, compression, rendering should preserve, and possibly enhance, the quality of the video signal; to do so, quality control mechanisms are required. These mechanisms rely on systems that can assess the visual quality of the incoming signal consistently with human perception. Computational Intelligence (CI) paradigms represent a suitable technology to tackle this challenging problem. The present research introduces an augmented version of the basic Extreme Learning Machine (ELM), the Circular-ELM (C-ELM), which proves effective in addressing the visual quality assessment problem. The C-ELM model derives from the original Circular BackPropagation (CBP) architecture, in which the input vector of a conventional MultiLayer Perceptron (MLP) is augmented by one additional dimension, the circular input; this paper shows that C-ELM can actually benefit from the enhancement provided by the circular input without losing any of the fruitful properties that characterize the basic ELM framework. In the proposed framework, C-ELM handles the actual mapping of visual signals into quality scores, successfully reproducing perceptual mechanisms. Its effectiveness is proved on recognized benchmarks and for four different types of distortions.

© 2012 Elsevier B.V. All rights reserved.

### 1. Introduction

Ensuring the excellence of the user's visual experience has become mandatory for modern imaging applications. To satisfy users, the final image or video has to be delivered at the highest visual quality. Limitations in both transmission and imaging technologies bring about distortions to the video signal, thus ending up in poor visualization. Modern displays often embed post-processing systems, which can restore the signal quality before the actual on-screen visualization. To address distortions in an appropriate way, post-processing technologies should act based on the assessment of the (perceived) visual quality of the incoming signal. Such assessment is required to be performed automatically and in real time, and can be used also in a feedback loop to maximize the effectiveness of the restoration process [1–3]. Systems in charge to automatically assess visual quality are generally known as objective quality metrics.

The design of objective quality metrics is a complex task because predictions must be consistent with human visual quality

preferences. Human preferences are inherently quite variable and, by definition, subjective; moreover, in the field of visual quality, they stem from perceptual mechanisms that are not fully understood yet. A common choice is to design metrics that replicate the functioning of the Human Visual System to a certain extent, or at least that take into account its perceptual response to visual distortions by means of numerical features (for an overview on the subject, see [4]). Although successful, these approaches come with a considerable computational cost, which makes them impractical for real-time applications. Objective systems include tools for image quality assessment at concept-level, i.e., systems that analyze image data and metadata for evaluating the perceived quality of the cognitive and affective information associated with personal pictures [5].

Computational Intelligence (CI) paradigms allow to tackle the quality assessment task from a different perspective, since they aim at mimicking quality perception instead of designing an explicit model of the human visual system. Quality assessment based on CI paradigms follows a 2-step approach. In the first step, a meaningful feature-based representation of the visual signal is defined. In the second step, a learning machine handles the actual mapping of the feature vector into quality scores. Since such an approach can rely on the ability of CI tools to deal with complex, non-linear problems,

\* Corresponding author. Tel.: +39 10 3532268.

E-mail address: [paolo.gastaldo@unige.it](mailto:paolo.gastaldo@unige.it) (P. Gastaldo).

relatively simple metrics can be designed. Most of the computational power is spent in the off-line training phase; indeed, a trained system can support real-time quality assessment with a minimal overhead to the feature extraction computational cost. Several studies proved the effectiveness of methodologies that exploit CI tools to address image quality [6–12].

The Extreme Learning Machine (ELM) [13–15] model is particularly suitable to support a real-time, objective quality prediction task, given its well-known capability of handling complex problems with low computational costs and straightforward training. Indeed, the ELM has been already successfully applied in domains such as bioinformatics [16–18], computer vision [19], data mining [20] and robotics [21].

This paper presents an augmented version of the standard ELM model [13], namely the Circular ELM (C-ELM), and shows that this model proves effective in mapping feature vectors into quality perception. The C-ELM structure originates from the Circular BackPropagation (CBP) architecture [22]. The CBP approach augments conventional MultiLayer Perceptrons (MLP) by one additional input dimension, which is worked out as the norm of the input vector itself. The C-ELM is obtained by applying the same structural enhancement to a standard ELM model.

The research described in this paper embeds a C-ELM network into a Reduced-Reference objective quality metric. Reduced-Reference (RR) methods [23–28] combine the information on the received signal with a (limited) number of features extracted from the original signal. These features are usually small enough to be tagged as metadata and transmitted along with the image/video. The RR paradigm overcomes the traditional, unrealistic Full-Reference approach [3], which requires the knowledge of the whole original image for quality prediction.

The challenging task in the design of a RR model is the selection of the smallest set of features that can support quality assessment effectively. Since acquisition, transmission and delivery processes can affect images with distortions that alter the original color distribution significantly, the color correlogram [29] provides an effective tool to extract a compact description of the signal and its degradation. Objective methods for quality assessment typically analyze the luminance component of the color information, but recent studies [12] showed that chrominance also plays a relevant role in quality perception. As a consequence, in this paper both luminance and chrominance information (specifically hue) are involved in the numerical representation of pictures, and color correlograms combine such information in low-dimensional numerical descriptors.

The C-ELM objective quality metric was tested on the second release of the LIVE database [30]. The performance evaluation considered prediction accuracy, and took into account four types of distortion: White noise, Gaussian Blur, JPEG compression and JPEG2000 compression. The results confirmed the validity of the C-ELM based framework, as well as of the advantages of color statistics in reduced-reference paradigms.

Section 2 gives an overview of the CI-based approach to image quality assessment. Section 3 introduces the Circular-ELM model.

Section 4 illustrates the experimental set up adopted in this research. Section 5 presents the experimental verification of the overall framework. Some concluding remarks are made in Section 6.

## 2. Reduced-reference quality assessment by exploiting computational intelligence

A CI-based approach to objective quality assessment allows to reduce the overall complexity of the perception prediction problem by decoupling its modeling into two, less complex tasks, namely:

1. The definition of a suitable descriptive basis for the input image, i.e. a feature-based description  $\mathbf{f} \in \mathcal{F}$ , where  $\mathcal{F}$  is a feature space:

$$\mathbf{f} = \phi(I, I') \quad (1)$$

where  $I$  is the input image and  $I'$  is the reference image (whenever required);

2. The design by empirical learning of the non-linear mapping function,  $\gamma$ , between the feature space  $\mathcal{F}$  and a scalar measure of the perceived quality,  $q$ :

$$\gamma: \mathcal{F} \rightarrow [0, 1] \quad (2)$$

where the quality score is normalized in the range [0,1] without loss of generality.

The first task aims at reducing the dimensionality of the original data space, which is virtually of infinite size since multimedia signals are involved. The ability of the feature space  $\mathcal{F}$  to characterize the underlying perceptual phenomenon is critical for the effectiveness of the whole framework.

The second task takes advantage of the ability of CI paradigms to deal with multidimensional data characterized by complex relationships, which are learned from examples by using a training algorithm. This in turn allows one to bypass the challenging issue of designing an explicit model of perceptual mechanisms that maps  $\mathcal{F}$  into quality scores.

Fig. 1 outlines the CI-based, reduced reference quality assessment system:  $I^{(n)}$  is the original, reference image, and  $I^{(n,r)}$  is the image resulting from applying to  $I^{(n)}$  some impairment with a distortion strength,  $r$ . The quantities  $q^{(n)}$  and  $q^{(n,r)}$  denote the perceived quality scores obtained from a subjective (i.e., human) evaluation of  $I^{(n)}$  and  $I^{(n,r)}$ , respectively. Likewise,  $\mathbf{f}^{(n)}$  and  $\mathbf{f}^{(n,r)}$  denote the feature-based description of  $I^{(n)}$  and  $I^{(n,r)}$ , respectively. The fact that in our approach only  $\mathbf{f}^{(n)}$  is needed for the quality estimation, and not the full information on the original image,  $I^{(n)}$ , clarifies its classification as reduced reference paradigm.

The overall objective-quality assessment model can be regarded as a regression problem, since feature vectors  $\mathbf{f}$  have to be mapped into continuous, scalar quality scores. Over the past

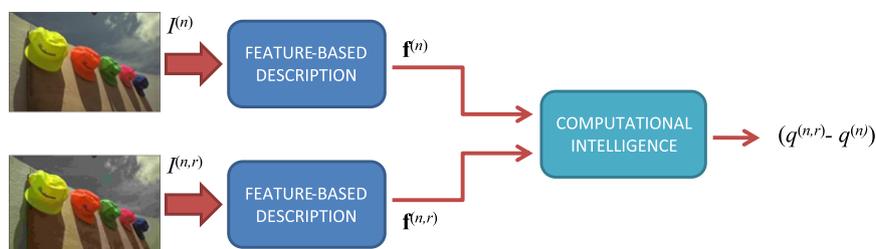


Fig. 1. A CI-based reduced-reference schema for image quality assessment.

decades, research in the CI area has lead to effective theoretical models that have been applied successfully to different application domains, including video/image quality assessment [6–12]. Feed-forward neural networks, in particular, proved to be a very powerful and flexible tool to address image quality assessment problems [12]. In the present work, a novel feed-forward model, the Circular-ELM, is adopted.

### 3. Circular-ELM for objective quality assessment

Previous research proved the effectiveness of using Circular BackPropagation (CBP) networks for image quality assessment [6,12,31,32]. The CBP network extends the conventional MLP with one additional input, being the sum of the squared values of all the network inputs. Such a structural enhancement eventually improves the ability of the resulting CI model to tackle complex, non-linear problems, such as modeling perceptual mechanisms.

In this work the same structural enhancement is applied to the basic ELM model. The resulting model, the Circular-ELM, is then used to support the non-linear quality prediction system presented in Section 2.

#### 3.1. Basic Extreme Learning Machine

The Extreme Learning Machine (ELM) approach [13] was introduced to overcome some issues in Back-Propagation network training [14], that is, possibly slow convergence rates, the critical tuning of optimization parameters [33], and the presence of local minima that call for multistart and re-training strategies. The ELM learning problem setting requires a training set,  $\Theta$ , of  $N_p$  labeled pairs  $(\mathbf{x}_i, y_i)$ , where  $\mathbf{x}_i \in \mathfrak{R}^M$  is the  $i$ th input vector and  $y_i \in \mathfrak{R}$  is the associate expected “target” value. Assuming a scalar output implies that the network has one output unit. This assumption comes without loss of generality, and nicely suits the visual quality prediction problem, where targets are scalar, subjective scores of overall image quality.

The input layer has  $M$  neurons and connects to the “hidden” layer (having  $N_h$  neurons) through a set of weights  $\{\hat{\mathbf{w}}_j \in \mathfrak{R}^M; j=1, \dots, N_h\}$ . The  $j$ th hidden neuron embeds a bias term,  $\hat{b}_j$ , and a nonlinear “activation” function,  $\zeta(\cdot)$ ; thus the neuron’s response to an input stimulus,  $\mathbf{x}$ , is

$$A_j(\mathbf{x}, \hat{\mathbf{w}}_j, \hat{b}_j) = \zeta(\hat{\mathbf{w}}_j \cdot \mathbf{x} + \hat{b}_j) \quad (3)$$

Note that (3) can be further generalized to a wider class of functions [34]. In the case of Radial Basis Functions (RBF),  $A_j(\mathbf{x}, \hat{\mathbf{w}}_j, \hat{b}_j)$  is given by

$$A_j(\mathbf{x}, \hat{\mathbf{w}}_j, \hat{b}_j) = \zeta(\hat{b}_j \cdot \|\mathbf{x} - \hat{\mathbf{w}}_j\|) \quad (4)$$

where  $\hat{\mathbf{w}}_j$  and  $\hat{b}_j$  are the center and the gain of the  $j$ th node, respectively.

A vector of weighted links,  $\mathbf{w} \in \mathfrak{R}^{N_h}$ , connects hidden neurons to the output neuron. The overall output function,  $y$ , of the network is

$$y_{ELM}(\mathbf{x}) = \sum_{j=1}^{N_h} w_j A_j(\mathbf{x}, \hat{\mathbf{w}}_j, \hat{b}_j). \quad (5)$$

To formalize the ELM training problem, it is convenient to define an “activation matrix”,  $\mathbf{H}$ , such that the entry  $\{h_{ij} \in \mathbf{H}; i=1, \dots, N_p; j=1, \dots, N_h\}$  is the activation value of the  $j$ th hidden neuron for the  $i$ th input pattern. The  $\mathbf{H}$  matrix is

$$\mathbf{H} \equiv \begin{pmatrix} A_1(\mathbf{x}_1, \hat{\mathbf{w}}_1, \hat{b}_1) & \dots & A_{N_h}(\mathbf{x}_1, \hat{\mathbf{w}}_{N_h}, \hat{b}_{N_h}) \\ \vdots & \ddots & \vdots \\ A_1(\mathbf{x}_{N_p}, \hat{\mathbf{w}}_1, \hat{b}_1) & \dots & A_{N_h}(\mathbf{x}_{N_p}, \hat{\mathbf{w}}_{N_h}, \hat{b}_{N_h}) \end{pmatrix} \quad (6)$$

In the ELM model, the quantities  $\{\hat{\mathbf{w}}_j, \hat{b}_j\}$  in (6) are set randomly and are not subject to any adjustment, and the quantities  $\mathbf{w}_j$  in (5) are the only degrees of freedom. Thus, the training problem reduces to the minimization of the convex cost:

$$\min_{\{\mathbf{w}\}} \|\mathbf{H}\mathbf{w} - \mathbf{y}\|^2 \quad (7)$$

A matrix pseudo-inversion yields the unique  $L_2$  solution, as proved in [13]

$$\mathbf{w} = \mathbf{H}^+ \mathbf{y} \quad (8)$$

The simple, efficient procedure to train an ELM therefore involves the following steps:

1. Randomly set the input weights  $\hat{\mathbf{w}}_j$  and bias  $\hat{b}_j$  for each hidden neuron;
2. Compute the activation matrix,  $\mathbf{H}$ , as per (5);
3. Compute the output weights by solving a pseudo-inverse problem as per (8).

In spite of the apparent simplicity of the ELM approach, the crucial result is that even random weights in the hidden layer endow a network with a satisfactory representation ability. The theory derived in [13] proves that the ELM with  $N_h = N_p$  hidden nodes can approximate every function.

#### 3.2. Circular-ELM

Circular BackPropagation (CBP) networks [22] improve over the basic formulation of MLP; the CBP model augments the input vector by one additional dimension, which is computed as the norm of the input vector itself. In a classic set-up involving a single layer network, the estimation process supported by the enhanced, CBP network is expressed as

$$y_{CBP}(\mathbf{x}) = \zeta \left( b + \sum_{j=1}^{N_h} \left[ w_j \zeta \left( \hat{b}_{j,0} + \sum_{k=1}^M \hat{w}_{j,k} x_k + \hat{w}_{j,M+1} \|\mathbf{x}\|^2 \right) \right] \right) \quad (9)$$

The model name results from the fact that a CBP classifier can map both linear and circular separation boundaries [22]. The additional term boosts the overall representation ability of the network, while not affecting the fruitful properties of an MLP structure. Moreover, setting  $\hat{w}_{j,M+1}$  to 0 reduces a CBP network to a classical MLP, but the actual value of the coefficient  $\hat{w}_{j,M+1}$  is determined during training, hence the choice between a conventional MLP and a CBP network need not be made in advance.

The structural enhancement provided by (9) can be easily extended to the basic ELM formulation. Actually, one only needs to rewrite the neuron’s response to an input stimulus (3) as

$$A_j(\mathbf{x}, \hat{\mathbf{w}}_j, \hat{b}_j) = \zeta(\hat{\mathbf{w}}_j \cdot \bar{\mathbf{x}} + \hat{b}_j) \quad (10)$$

where  $\hat{\mathbf{w}}_j \in \mathfrak{R}^{M+1}$  and  $\bar{\mathbf{x}}_i = [x_{i,1}, \dots, x_{i,M}, \|\mathbf{x}_i\|^2]$ . Indeed, it is possible to obtain another form of the neuron’s response (10):

$$A_j(\mathbf{x}, \hat{\mathbf{w}}_j, \hat{b}_j) = \zeta(z_j \cdot \|\mathbf{x} - \mathbf{c}_j\|^2 - \bar{b}_j) \quad (11)$$

where

$$z_j = \hat{w}_{j,M+1}$$

$$\mathbf{c}_j = \left[ -\frac{\hat{w}_{j,1}}{2\hat{w}_{j,M+1}}, \dots, -\frac{\hat{w}_{j,M}}{2\hat{w}_{j,M+1}} \right]$$

$$\bar{b}_j = \frac{1}{\hat{w}_{j,M+1}} \left( \sum_{k=1}^M \frac{\hat{w}_{j,k}^2}{4\hat{w}_{j,M+1}} - \hat{b}_j \right)$$

The latter formulation of the neuron’s response  $A_j(\mathbf{x}, \hat{\mathbf{w}}_j, \hat{b}_j)$  exhibits a radial symmetry, hence one can assert that the CBP

enhancement to the basic ELM eventually yields an RBF-like structure.

The training problem of the resulting Circular-ELM (C-ELM) model still implies the minimization of the convex cost (7), as one just needs to update the neuron's response (6) to switch from basic ELM to C-ELM. As a result, the training procedure of C-ELM does not differ from that of basic ELM.

The C-ELM model can actually benefit from the enhancement provided by the circular input without losing any of the valuable properties that characterize the standard ELM framework. Hence, C-ELM can be effectively implemented by an electronic embedded system for real-time prediction—thanks to the simple expression of the decision functions. In this regard, the computation of the circular input does not bring about any significant additional complexity.

Furthermore, regularization strategies can yet be applied to the cost formulation (7) by adopting the approaches proposed in [13]. The ultimate effect of the regularizing term is the possibility to use a high number of neurons without affecting the numerical stability of the convergence process; this in turn means that C-ELM performances can be improved both in terms of numerical stability and accuracy. Such aspect is crucial when considering that in dealing with image quality assessment, usually the size of the training set is rather small (i.e., hundreds of patterns).

#### 4. Experimental set up

The present research aims to evaluate the performance of the C-ELM model in the quality-assessment problem. At the same time, one wants to compare the performances of C-ELM, ELM and CBP when they are embedded in a visual quality-assessment framework. To this purpose, this work adopts the experimental protocol that has already been established in [12] (Fig. 2). A two-layer architecture can tackle images affected by a set of possible different distortions, without any previous knowledge on the actual distortion effect that altered the image. The first layer performs distortion identification, so that images can be forwarded to distortion-specific sub-modules in the second layer. Each sub-module has been trained to predict the loss in visual quality that is brought about by one specific distortion. As a result, the two-layer approach ensures the prediction accuracy

typical of distortion-specific systems, while retaining the ability to cover a spectrum of different distortion effects.

This paper focuses on the implementation of the second layer of system [12], thus evaluating the ability of the CI models to attain valuable performances when addressing the regression problem that maps feature vectors into quality scores. Therefore, in the following it will be assumed that Layer I embeds the SVM-based distortion classifier adopted in [12].

##### 4.1. Color correlogram for reduced-reference quality assessment

The RR quality assessment framework presented in [12] uses the second order statistics of color information to work out numerical descriptors from images  $I^{(n)}$  and  $I^{(n,r)}$ . Second-order histograms have been proved effective, in the B/W domain, to convey significant information about perceived image quality [6]. The research presented in [12] adopts the *color correlogram* [29] for extracting numerical features. This second-order histogram expresses how the spatial correlation of pairs of colors changes with distance.

The color correlogram for an image region  $\Gamma$  (including  $W_\Gamma \times H_\Gamma$  pixels) describes the spatial correlation of pairs of colors with respect to the distance  $k$ . Formally, given a set of  $L$  colors  $C = \{C_1, \dots, C_L\}$ , an element of the correlogram matrix  $Z_\Gamma^k$  is defined as

$$z_\Gamma^k(i,j) = \left\{ \begin{array}{l} (m,n), m < W_\Gamma, n < H_\Gamma, \text{ s.t.} \\ \Gamma[m,n] = C_i; \Gamma[p,q] = C_j; \text{dist}(\Gamma[m,n], \Gamma[p,q]) = k \end{array} \right\} \\ i,j = 1, \dots, L \quad (12)$$

where  $\text{dist}()$  is a measure of distance between a pair of pixels. Each element of the matrix gives the probability of finding a pixel having color  $C_j$  at a distance  $k$  from a pixel having color  $C_i$ . In this paper,  $\text{dist}()$  embeds the  $L_1$ -norm: only the pairs of pixels lying at a distance  $k$  in the horizontal/vertical direction are considered in the histogram computation.

To preserve local information, every image is split into non-overlapping square regions (blocks), each holding  $H_\Gamma \times H_\Gamma$  pixels. An optimal setting of the block size,  $H_\Gamma$ , is crucial. This parameter should be set to  $H_\Gamma > 8$  to ensure a significant statistical sampling

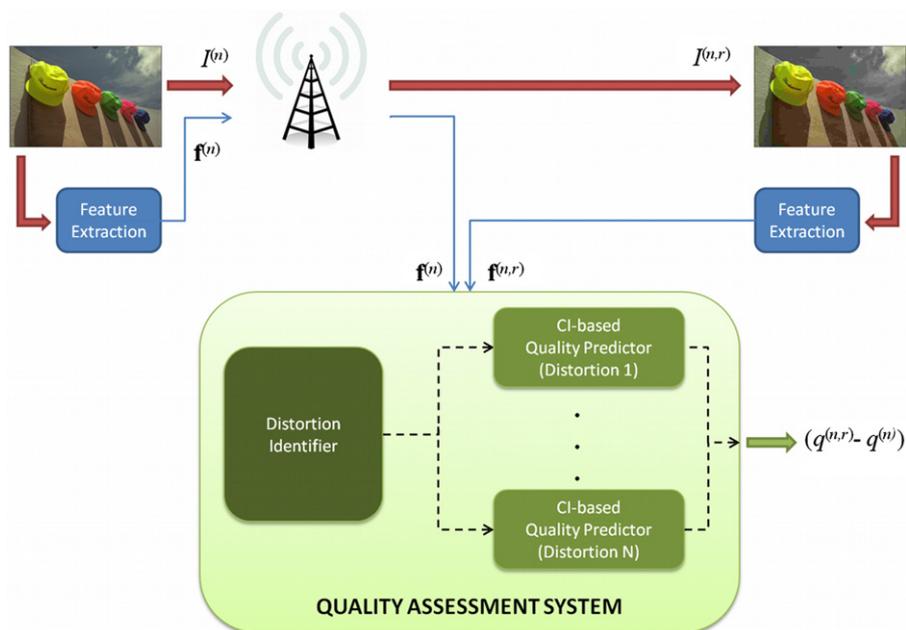


Fig. 2. Scheme of the proposed reduced-reference framework for objective quality assessment [12].

of the region; on the other hand, one should limit the value of  $H_I$  to actually work out local information.

For each block, the color correlogram is computed and a set of features,  $\Phi$ , is worked out that extracts relevant statistical information from the image (see Appendix for a more detailed overview of such features). The local-level feature extraction gives, for each feature  $f_u \in \Phi$ , as many values as the number of blocks into which the input image was divided, each describing local color distribution divergences. Humans, however, express a single judgment for the overall impression of the quality of the image; therefore a single numerical image descriptor (i.e. one vector of feature values) should be associated with the quality score. Toward this end, the present approach aggregates block-based information by means of statistical approach to render the distribution of a feature,  $f_u \in \Phi$ , over the image. The pseudo-code of Fig. 3 outlines the procedure to assemble the global descriptors for both the reference and the distorted image.

The eventual result of the procedure is, for each feature  $f_u \in \Phi$ , the description of an image by means of a global-level vector,  $\mathbf{x}_u$ , whose dimensionality depends on how many percentiles are worked out to characterize  $X_u$  and that outlines the distribution of the values of feature  $f_u$  across the image. The global descriptor  $\mathbf{f}^{(n,r)}$  for image  $I^{(n,r)}$  will result from gathering the objective vectors  $\mathbf{x}_u^{(n,r)}$ , being  $n_f$  the number of features  $f_u \in \Phi$ :

$$\mathbf{f}^{(n,r)} = \{\mathbf{x}_u^{(n,r)}, u = 1, \dots, n_f\} \quad (13)$$

#### 4.2. Numerical description

As a first step, features are extracted from the image to feed the CI-based predictor. Both luminance and chrominance information of the original and the received images are used. The procedure outlined above applies separately to (1) the  $Y$  component, yielding the luminance descriptor  $\mathbf{y}_{RR}^{(n,r)} = [\mathbf{y}^{(n)} \mathbf{y}^{(n,r)}]$  in the  $YCbCr$  color space, and (2) the hue layer of the image represented in the HSV color space, resulting in the hue descriptor,  $\mathbf{h}_{RR}^{(n,r)} = [\mathbf{h}^{(n)} \mathbf{h}^{(n,r)}]$ .

For calculating the features, the image is split in blocks of  $32 \times 32$  pixels. For each block the features described in Table 1 are calculated from the color correlogram. For a given feature  $f_u \in \Phi$ , the objective

vectors  $\mathbf{y}_{u,RR}^{(n,r)}$  and  $\mathbf{h}_{u,RR}^{(n,r)}$  are constructed by assembling six percentiles (i.e., {0, 20, 40, 60, 80, 100}) of the distributions  $Y_u, H_u$  over all blocks. As a result, global descriptive vectors span a 6-dimensional space. Eventually, for a feature  $f_u$ , the overall objective representation of image  $I^{(n,r)}$  is based on the two vectors  $\mathbf{y}_{RR}^{(n,r)}$  and  $\mathbf{h}_{RR}^{(n,r)}$ , both sized 12 times the number of considered features.

In practice, the feature-based descriptors feed both the first and the second layer of the framework sketched in Fig. 2, and different set of features may be preferred for the different tasks [12]. As anticipated above, the present paper focuses only on the second layer. Section 5 will deal with the process of feature selection concerning the CI-based regression modules.

#### 4.3. C-ELM ensembles for predicting the loss/gain in quality

The quality-level prediction layer includes as many elements as the number,  $n$ , of distortions. Each element  $\Omega(d_i)$  is implemented by an ensemble of several regularized C-ELM networks, which are trained to quantify the effect on quality produced by the specific distortion  $d_i$ .

An “ensemble” [35] gathers parallel networks [35,36] that have been trained on the same problem. The use of ensembles is justified by the statistical fluctuations in the empirical training set, which give rise to the problem of getting robust estimators; indeed, by averaging the predictions of  $N$  independent estimators [35,36] one reduces the variance  $\sigma$ , brought about by statistical

**Table 1**  
Features used for the C-ELM predictors.

	JP2K	WN	GB	JPEG
Luminance	Entropy Homogeneity	Entropy Contrast	Entropy Homogeneity	Entropy Homogeneity
Hue	Homogeneity Contrast	Contrast Energy Ratio	Entropy Homogeneity	Diagonal Energy Entropy

### Feature Extraction

#### 0. Inputs

a picture  $I$ , a descriptive feature  $f_u \in \Phi$ ,  $u = 1, \dots, N_u$  and a value for distance  $k$

#### 1. Block-level feature extraction

a. Split  $I$  into  $N_b$  non-overlapping square blocks, and obtain the set:

$$\mathbf{B} = \{b_m; m = 1, \dots, N_b\}$$

b. For each block  $b_m \in \mathbf{B}$ : compute the associate color correlogram  $\mathbf{Z}_{b_m}^k$

c. For each matrix  $\mathbf{Z}_{b_m}^k$ : compute the value  $x_{u,m}$  of feature  $f_u$ , and obtain the set:

$$\mathbf{X}_u = \{x_{u,m}; m = 1, \dots, N_b\}$$

#### 2. Global level numerical representation

a. Compute a percentile-based description of  $X_u$ ; let  $p_\alpha$  be the  $\alpha$ -th percentile:  $\varphi_{\alpha,u} = p_\alpha(X_u)$

b. Assemble the objective descriptor vector,  $\mathbf{x}_u$ , for the feature  $f_u$  on the image  $I$

$$\mathbf{x}_u = \{\varphi_{\alpha,u}; \alpha \in [0,100]\} \quad u = 1, \dots, N_u,$$

#### 3. Output

The global descriptor  $\mathbf{x}_u$  for image  $I$

**Fig. 3.** The pseudo-code that outlines the procedure to assemble the global descriptors.

noise in training data. Ideally, the decrease in variance reduces to

$$\bar{\sigma}^2 = \sigma^2/N \quad (14)$$

Building independent estimators is crucial: when the number of patterns is small as compared with the data dimensionality (as is the case in quality estimation), an approach based on the theory of receptive fields [29,37] can apply. This method partitions the high-dimensional input space into several, lower-dimensional subspaces, and provides a specialized neural network for each subspace.

In the specific problem of quality prediction, the coordinate-partitioning approach leads to predictors that involve different features and thus validate the hypothesis of disjoint subspaces. Moreover, such a strategy shrinks the dimensionality of the input space for each neural network, thus enhancing the network's generalization ability.

In the current implementation, the  $i$ th ensemble, predicting the quality degradation for distortion  $d_i$ , involves a pair of features, hence includes a pair of C-ELM networks. The features should be as much informative as possible about the effect of the specific artifact on visual quality, and the feature selection process is performed independently for each artifact, since the ensembles of C-ELM networks are independent of one another.

The neural networks employed share the same architecture and always use a sigmoid as a neuron-activation function. The input layer has the dimensionality of vector  $\mathbf{y}_{u,RR}^{(n,r)}$ . One output neuron provides the objective estimate of gain/loss in quality of the distorted sample with respect to the original one.

## 5. Experimental results

The second release of the LIVE database [30] is used as a testbed for the performance evaluation of the presented model. In particular, the four datasets including samples distorted with JPEG compression, JPEG 2000 Compression, White noise and Gaussian Blur are considered.

A  $k$ -fold test strategy rules the overall performance evaluation of the CI-based quality estimator, mainly because this method proves effective to obtain reliable results when few data are available [38]. The LIVE datasets contain several distorted versions of the same 29 original images, varying in content. In the following, each dataset is divided into 5 groups ('folds'), each containing few original images, in such a way that none of the 29 originals belongs to more than one group. Then, for each content, the corresponding distorted versions are added to the fold. To test all the learning machines in the system, 5 runs are performed, for each of which 4 of the 5 folds are used as training data, and the remaining one is used as test set. In this way, the network is tested on distorted versions of contents not previously seen in the training. Thus, each sub-module is proven to be able to generalize independently from the specific set of contents presented in the training phase. To allow a fair comparison with the performance obtained in [12] by exploiting CBP neural networks, the whole experimental session adopts the same composition of the five folds used in [12].

For all the experiments reported in the following, the color correlogram is computed for distance  $k=1$  using the L1 norm. To evaluate the single and the joint contribution of luminance and hue information, the system is first trained entirely with the luminance component  $\mathbf{y}_u^{(n,r)}$  of the global descriptor, and then with the Hue component  $\mathbf{h}_u^{(n,r)}$ . Finally, the outputs of the two systems trained separately are averaged, to evaluate the benefits of exploiting the combined information.

The feature selection procedure presented in [12] is exploited to define the set of features needed by each predictor to process

the four kinds of distortion involved in the present research. Since hue and luminance information are consistently different, features were selected separately for the two color components. The obtained sets of features to be used as inputs for the four predictors are reported in Table 1.

### 5.1. Accuracy in predicting gain/loss in quality

As anticipated in Section 4, each single artifact is addressed by a specific CI-based quality estimator. Hence, each module in the second layer was trained on a different dataset taken from the LIVE database: the JP2K module was trained for both the JP2K(1) and JP2K(2) sets, the Noise and Blur predictors were trained on the Noise and Blur datasets respectively, and the JPEG module was trained for the sets JPEG(1) and JPEG(2). The two JP2K datasets included 82 and 87 patterns, respectively (undistorted images were not considered in the test); both the White noise and Gaussian blur datasets each contained 145 patterns; finally the JPEG(1) set included 87 images and the JPEG(2) set included 88 images.

For each image, a Difference Mean Opinion Score (DMOS) is provided in the LIVE database, indicating the perceived difference in quality between the original and processed image, and ranging between [1, 100]. For computational reasons, these scores were further remapped for this study into the range  $[-1, +1]$ .

Thanks to the flexibility of the system, it was possible to design the hidden layer of the C-ELM models specifically for each task. Table 2 reports the configurations adopted for the different predictors involved in the proposed framework, resulting from a model selection procedure. The quality prediction system was first trained with the luminance component of the patterns, and then with the hue component. The input layer of the networks was therefore sized as the dimensionality of  $\mathbf{y}_u^{(n,r)}$  and  $\mathbf{h}_u^{(n,r)}$  (12 input neurons). The evaluation of the accuracy in prediction involves two different descriptors, which measure the discrepancy between the predicted change in quality,  $\hat{d}_s(q^{(n)}, q^{(n,r)})$  and the actual change in quality provided by the LIVE database,  $d_s(q^{(n)}, q^{(n,r)})$ . The two evaluation quantities are:

- The Pearson's correlation coefficient,  $\rho$  between  $d_s$  and  $\hat{d}_s$ ;
- the root mean square error (RMSE), computed on the prediction error between  $d_s$  and  $\hat{d}_s$ .

Figs. 4–7 report the results obtained for the different distortions included in the LIVE database, detailed per each-run of the  $k$ -fold validation. In each figure, for a given dataset two plots are presented: the first refers to Pearson correlation and compares for each run the results obtained by using as input luminance information, hue information, and the combination of luminance and hue, respectively. The second plot, analogously, compares the RMSE values. The approach based on luminance information seems to

**Table 2**  
Number of neurons in the hidden layer for the different C-ELM networks involved in the experimental session.

	Luminance		Hue	
	Predictor #1	Predictor #2	Predictor #1	Predictor #2
JPEG—Set1	150	10	20	20
JPEG—Set2	90	190	30	120
JP2K—Set1	90	70	30	150
JP2K—Set2	70	50	10	10
Gaussian Blur	150	80	160	200
White Noise	120	170	110	140

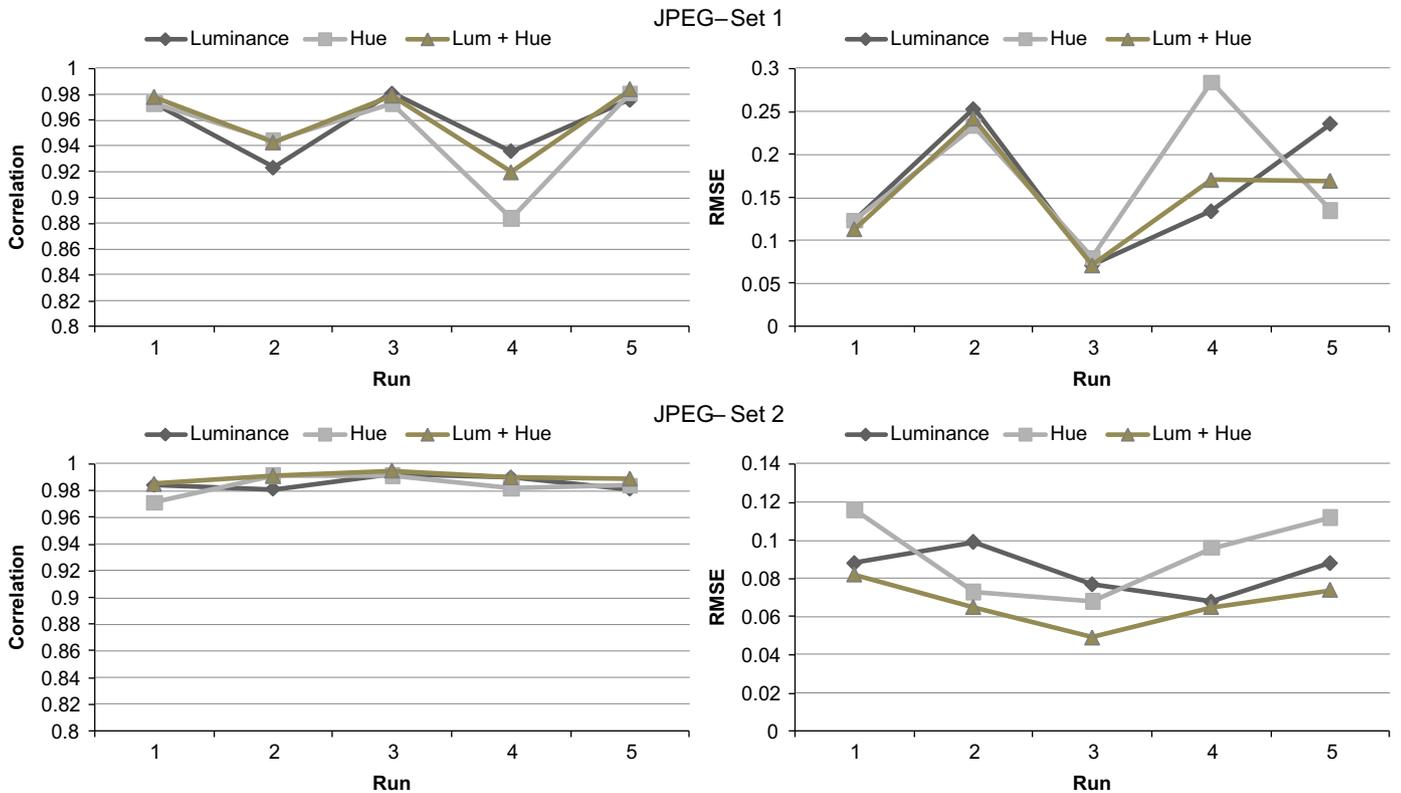


Fig. 4. Accuracy in assessing the quality of image compressed with JPEG: comparison between luminance, hue and combination of luminance and hue.

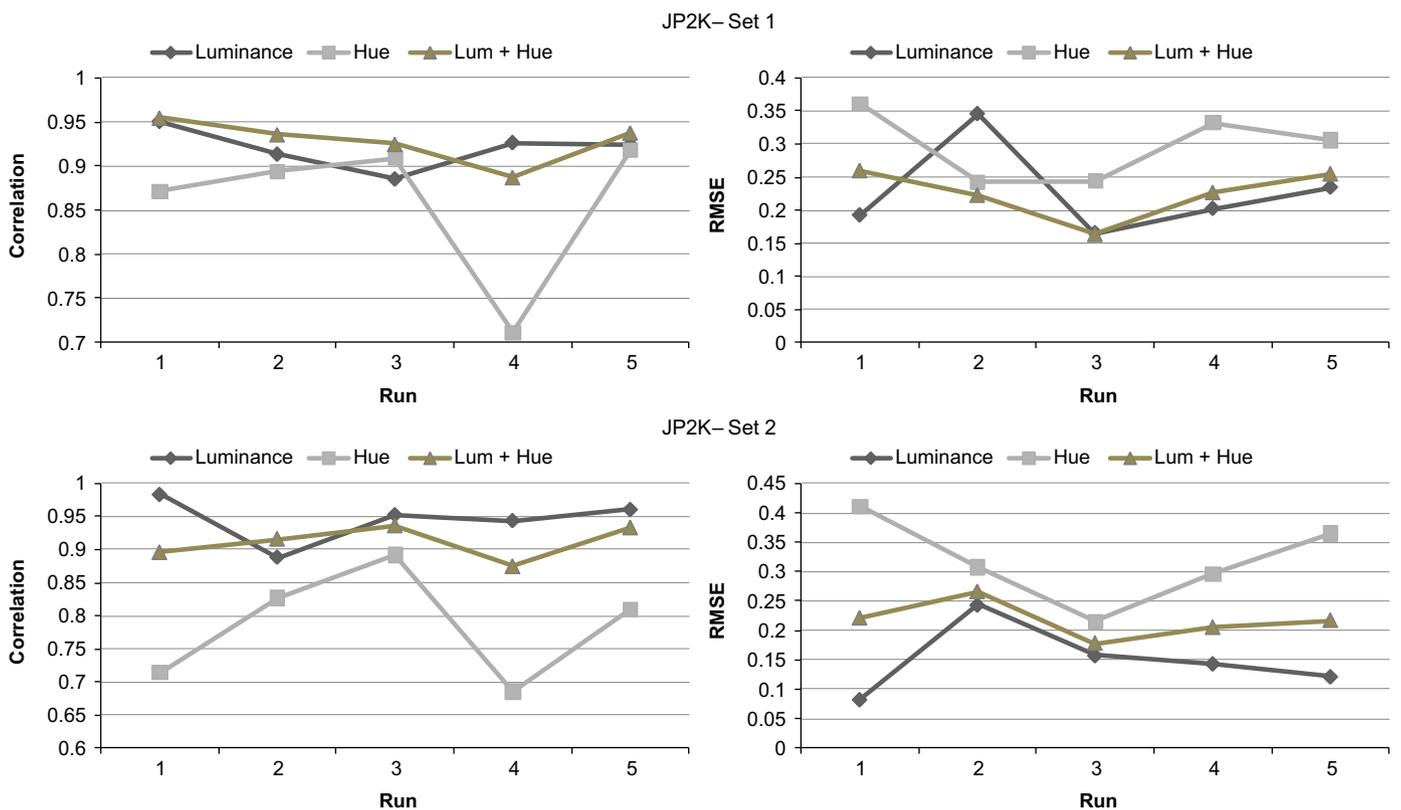


Fig. 5. Accuracy in assessing the quality of image compressed with JPEG 2000: comparison between luminance, hue and combination of luminance and hue.

perform better than the one using hue information. Predicting the quality of JPEG compressed images though, the performance is comparable. The system is weakest in dealing with JPEG 2000

compressed images. The corresponding RMSE, however, is relatively low, especially for the luminance-based system, and can be considered acceptable for many applications.

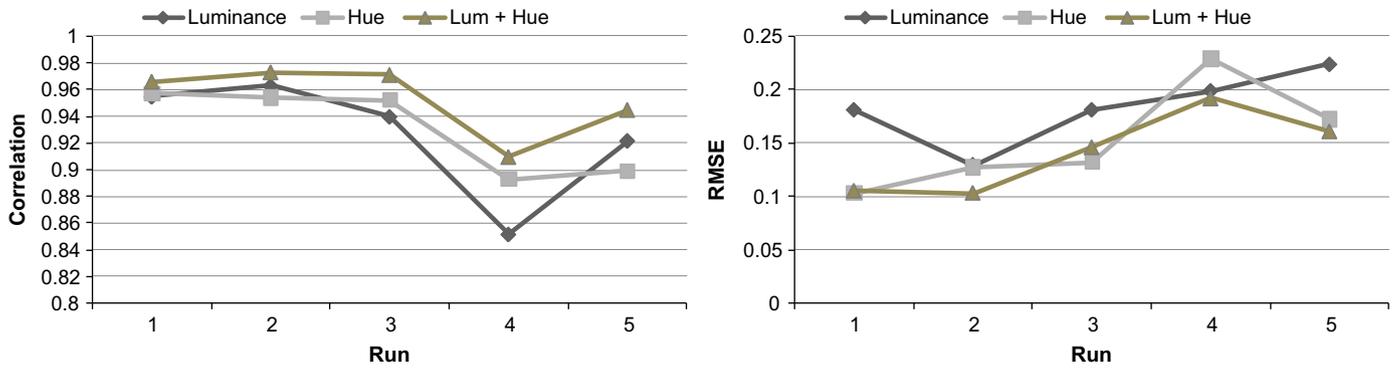


Fig. 6. Accuracy in assessing the quality of image distorted with Gaussian Blur: comparison between luminance, hue and combination of luminance and hue.

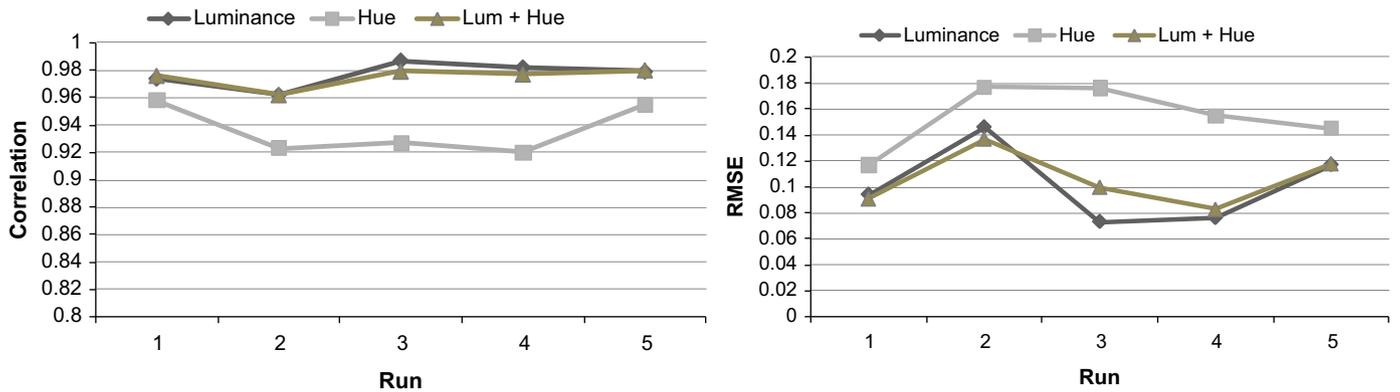


Fig. 7. Accuracy in assessing the quality of image distorted with White Noise: comparison between luminance, hue and combination of luminance and hue.

**Table 3**  
Comparison between the performances achieved with different neural networks.

	JPEG1	JPEG2	JP2K1	JP2K2	Noise	Blur
C-ELM	0.961	0.990	0.928	0.911	0.975	0.953
ELM-sigmoid	0.937	0.979	0.882	0.868	0.949	0.942
ELM-RBF	0.928	0.971	0.884	0.877	0.909	0.920
CBP [12]	0.938	0.979	0.880	0.874	0.980	0.924

To evaluate the joint contribution of luminance and hue information, the scores provided for each sample by the two systems are averaged. When combining hue and luminance information, the performance of the system in general improves. For five out of the six datasets, the Pearson correlation increases and the mean percentage prediction error decreases with respect to systems based on hue or luminance information only. Only in the case of JPEG 2000 dataset 2 is slightly worse with respect to the luminance-only system. Therefore, the increase in complexity corresponding to the ensemble of the two systems is worth the cost, considering the general improvement.

### 5.2. The added value of the Circular input in ELM

To evaluate the benefits brought about the addition of the Circular input to the standard ELM paradigm, the performance attained by C-ELM on the image quality assessment problem is compared to that of the basic ELM (Table 3). Two configurations are considered: one based on additive hidden nodes with sigmoidal activation function (Eq. (3), ELM-sigmoid in Table 3) and one with RBF hidden nodes with Gaussian activation function (Eq. (4), ELM-RBF in Table 3). For reference purposes, the table also reports

the performance attained by the CBP neural network [12]. The comparison is based on the Pearson's Correlation Coefficient between predictions and actual subjective quality judgments; the table reports the average prediction accuracy of the ideal system, i.e. when one assumes that Layer I does not bring about any error in identifying the distortion affecting the images. Numerical results clearly show that C-ELM improves over basic ELM and CBP, as its predictions are the closest to human quality judgments (highest Pearson's correlation coefficient) for each of the six datasets involved in the experimental session.

### 5.3. A comparison between C-ELM-based quality assessment and other systems

To fairly evaluate the robustness of the proposed system against other methods presented in literature, we analyze its performance when the two layers are combined, i.e. when layer I is no more considered as an ideal system. Hence, we consider the case where layer I supports the distortion identification task and actually redirects the input image to the predictor that is entitled to evaluate the quality score. For this final validation, the SVM-based distortion identifier introduced in [12] is adopted. For the sake of brevity, only one of the runs of the 5-fold procedure is going to be discussed, and considered representative for the performance of the complete system. Based on the partial results for layer I [12] and layer II, run #3 is selected. Run #3 attained a satisfactory, but not the best performance. As such, it seems fair to estimate the overall performance of the proposed system based on this specific run.

In practical terms, evaluating the joint performance of layers I and II means that the quality of each image is assessed through the predictor specialized on the distortion that layer I identified

for that image. The confusion matrix in Table 4 reports on the performance of the SVM-based distortion identifier for Run #3. Two JPEG images are recognized as being distorted by blur, and one JPEG 2000 as being distorted by noise. As a consequence, the two JPEG images enter the blur-specialized predictor, and the JPEG 2000 image enters the noise-specialized predictor. The robustness of the system in dealing with these misclassification cases is easily verifiable by comparing the results reported in the first two rows of Table 5.

Table 5 also allows the comparison of the complete system with other image quality assessment methodologies. The data included in the table are organized as follows. On the first and second rows precision is reported for the prediction accuracy of the complete system (row 1) and for the system with ideal classification (row 2), respectively. On the third row the prediction accuracy attained also on Run #3 by the approach presented in [12], which embeds the SVM-based distortion identifier in layer I and exploits CBP networks as quality predictors in Layer II. In the following rows, three general purpose reduced-reference systems are considered: those proposed by Wang et al. [26], by Li and Wang [28] and by Narwaria and Lin [11]; the latter one recently presented a quality assessment framework based on machine learning. Additionally, for reference purposes, also the well-known full-reference metrics MSSIM [39] and PSNR are given; though, one should take into account that the comparison of FR with RR metrics is not completely fair. The comparison is based on the Pearson's Correlation Coefficient between predictions and actual subjective quality judgments. It should be noted that the work by Narwaria and Lin [11] does not give separate values of prediction accuracy for the two different datasets for the JPEG2000 and the JPEG distortions. As such, the table reports only one value of Pearson's Correlation Coefficient for each distortion.

Table 5 provides two important outcomes. The first is that the quality assessment system exploiting C-ELM improves over the one using CBP. Indeed, C-ELM significantly outperforms CBP for the two datasets involving JPEG2000 compressed images and the Blur dataset. This in turn confirms that C-ELM is able to properly address a complex, strongly non-linear problem such as modeling perceptual mechanisms.

The second important outcome is that the proposed framework when exploiting C-ELM compares satisfactorily with the

other quality assessment methods proposed in the literature. In this regard, one should notice that the estimated performances of the proposed methodology appear quite robust, as they were estimated on five independent runs involving different compositions of the training and the test set. By contrast, some of the other methods considered [18,20,31] measure prediction accuracy after applying a non-linear regression fit against the subjective scores. Although prescribed by VQEG [40], this practice lacks a test of the effectiveness of the trained regression function on images not included in the training process.

## 6. Conclusions

This paper introduced a new feed-forward neural network, the Circular-ELM, that augments the standard ELM paradigm by introducing a structural enhancement derived from the CBP network. The theoretical discussion and the experimental session showed that standard ELM networks can actually benefit from the enhancement provided by the circular input without losing any of the fruitful properties that characterize the basic ELM framework. Indeed, experimental results demonstrated that C-ELM can outperform other powerful models such as the standard ELM and the CBP network. Furthermore, the satisfactory performances attained by C-ELM on the complex domain of image quality assessment endorse future research activities on this neural network model. In this sense, follow-up research might focus on a thorough comparison of CBP, standard ELM and C-ELM on different real-world problems to assess the effectiveness of the latter model.

From another perspective, the C-ELM based system showed high accuracy and robustness in assessing visual quality when compared to state-of-the-art methods. This is a further confirmation that Computational Intelligence methods are particularly suitable to tackle visual perception modeling problems in general, and the one of objective quality assessment in particular. Specifically, the advantages brought by CI to objective quality assessment are:

1. The overall complexity of the system is decreased, as trained CI components can reliably predict quality based on relatively simple metrics.
2. The framework is flexible and can be adapted to any kind of media.
3. CI-based systems can be efficiently implemented into embedded electronic systems.
4. The accuracy is comparable or even higher than that of conventional objective metrics.

As a consequence, future research should focus on further exploring the possibilities offered by ELM in the visual quality assessment domain. In particular, a C-ELM based implementation of layer I should be tried, in order to minimize the misclassification of images had with SVM in [12]. Furthermore, C-ELM might prove adequate also in supporting the more complex (but also

**Table 4**  
Confusion matrix for the SVM-based distortion identifier (Layer I) using Run #3.

Original	Classified as			
	JPEG	JP2K	Noise	Blur
JPEG	30	0	0	2
JP2K	0	35	1	0
Noise	0	0	30	0
Blur	0	0	0	30

**Table 5**  
Comparison between different approaches to image quality assessment.

	JPEG1	JPEG2	JP2K1	JP2K2	Noise	Blur
Proposed (run#3+classifier [12])	0.968	0.995	0.918	0.935	0.980	0.950
Proposed (run#3+ideal classifier)	0.979	0.995	0.925	0.936	0.979	0.971
Redi et al. (CBP) [12]	0.934	0.991	0.777	0.887	0.986	0.859
Li and Wang [28]	0.820	0.957	0.948	0.965	0.965	0.956
Wang et al. [26]	0.845	0.969	0.935	0.949	0.889	0.887
Narwaria and Lin [11]	0.947		0.956		0.989	0.951
MSSIM [39]	0.960	0.903	0.955	0.964	0.974	0.950
PSNR	0.886	0.916	0.933	0.874	0.985	0.783

more crucial) problem of video quality assessment. If coupled with satisfactory accuracy, the reduced computational complexity proper of C-ELM could indeed allow the successful solution of the emerging problem of on-line Quality of Experience optimization in internet-based multimedia delivery.

**Appendix. Correlogram features**

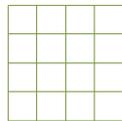
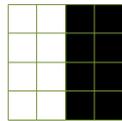
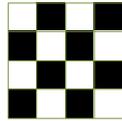
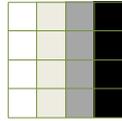
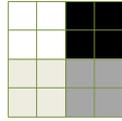
Table 6 presents the objective descriptors  $f_u \in \Phi$  used in the feature-extraction process. Every single feature among those reported in Table 6 assesses a specific property of the correlogram matrix. For the sake of clarity and completeness, in the following a

short analysis of the six features will be proposed. To this end, Fig. 8 gives five basic examples of image regions  $\Gamma$ , along with the correlogram matrices that eventually are extracted from those regions. In these examples, pixels are expressed in gray levels, and at most four different gray levels are represented. The correlograms refer to the set up  $k=1$ . Furthermore, the correlogram elements have been normalized; hence  $0 \leq z_T^k(i,j) \leq 1$  and  $\sum_{i,j} z_T^k(i,j) = 1$ .

Feature  $f_0$  (energy) can take values that range between 0 and 1. The maximum value is attained when all the elements of the correlogram but one are null; in that case, the only non-null element takes value 1. Regions  $\Gamma_a$  and  $\Gamma_c$  of Fig. 8 give two examples of array of pixels that lead to a correlogram with  $f_0=1$ . On the other hand, the definition of  $f_0$  clarifies that the energy

**Table 6**  
Objective features derived from the color correlogram

Feature name	Definition	Feature name	Definition
Energy	$f_0 = \sum_{i,j} [z_T^k(i,j)]^2$	Diagonal energy	$f_1 = \sum_i [z_T^k(i,i)]^2$
Entropy	$f_2 = -\sum_{i,j} z_T^k(i,j) \log_2 z_T^k(i,j)$	Contrast	$f_3 = \sum_q q^2 \left( \sum_{ i-j =q} z_T^k(i,j) \right)$
Homogeneity	$f_4 = \sum_{i,j} z_T^k(i,j) / [1 + (i-j)^2]$	Energy ratio	$f_5 = f_0 / f_1$

Region	Correlogram matrix	Feature values																									
	<table border="1"> <tr><td></td><td>0</td><td>1</td><td>2</td><td>3</td></tr> <tr><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>1</td><td></td><td>0</td><td>0</td><td>0</td></tr> <tr><td>2</td><td></td><td></td><td>0</td><td>0</td></tr> <tr><td>3</td><td></td><td></td><td></td><td>1</td></tr> </table>		0	1	2	3	0	0	0	0	0	1		0	0	0	2			0	0	3				1	$f_0 = 1$ $f_1 = 1$ $f_2 = 0$ $f_3 = 0$ $f_4 = 1$ $f_5 = 1$
	0	1	2	3																							
0	0	0	0	0																							
1		0	0	0																							
2			0	0																							
3				1																							
	<table border="1"> <tr><td></td><td>0</td><td>1</td><td>2</td><td>3</td></tr> <tr><td>0</td><td>0.41</td><td>0</td><td>0</td><td>0.18</td></tr> <tr><td>1</td><td></td><td>0</td><td>0</td><td>0</td></tr> <tr><td>2</td><td></td><td></td><td>0</td><td>0</td></tr> <tr><td>3</td><td></td><td></td><td></td><td>0.41</td></tr> </table>		0	1	2	3	0	0.41	0	0	0.18	1		0	0	0	2			0	0	3				0.41	$f_0 = 0.36$ $f_1 = 0.33$ $f_2 = 1.5$ $f_3 = 1.62$ $f_4 = 0.83$ $f_5 = 0.91$
	0	1	2	3																							
0	0.41	0	0	0.18																							
1		0	0	0																							
2			0	0																							
3				0.41																							
	<table border="1"> <tr><td></td><td>0</td><td>1</td><td>2</td><td>3</td></tr> <tr><td>0</td><td>0</td><td>0</td><td>0</td><td>1</td></tr> <tr><td>1</td><td></td><td>0</td><td>0</td><td>0</td></tr> <tr><td>2</td><td></td><td></td><td>0</td><td>0</td></tr> <tr><td>3</td><td></td><td></td><td></td><td>0</td></tr> </table>		0	1	2	3	0	0	0	0	1	1		0	0	0	2			0	0	3				0	$f_0 = 1$ $f_1 = 0$ $f_2 = 0$ $f_3 = 9$ $f_4 = 0.1$ $f_5 = 0$
	0	1	2	3																							
0	0	0	0	1																							
1		0	0	0																							
2			0	0																							
3				0																							
	<table border="1"> <tr><td></td><td>0</td><td>1</td><td>2</td><td>3</td></tr> <tr><td>0</td><td>0.125</td><td>0.16</td><td>0</td><td>0</td></tr> <tr><td>1</td><td></td><td>0.125</td><td>0.16</td><td>0</td></tr> <tr><td>2</td><td></td><td></td><td>0.125</td><td>0.16</td></tr> <tr><td>3</td><td></td><td></td><td></td><td>0.125</td></tr> </table>		0	1	2	3	0	0.125	0.16	0	0	1		0.125	0.16	0	2			0.125	0.16	3				0.125	$f_0 = 0.13$ $f_1 = 0.06$ $f_2 = 2.7$ $f_3 = 0.5$ $f_4 = 0.75$ $f_5 = 0.46$
	0	1	2	3																							
0	0.125	0.16	0	0																							
1		0.125	0.16	0																							
2			0.125	0.16																							
3				0.125																							
	<table border="1"> <tr><td></td><td>0</td><td>1</td><td>2</td><td>3</td></tr> <tr><td>0</td><td>0.083</td><td>0.083</td><td>0</td><td>0.083</td></tr> <tr><td>1</td><td></td><td>0.083</td><td>0.083</td><td>0</td></tr> <tr><td>2</td><td></td><td></td><td>0.083</td><td>0.083</td></tr> <tr><td>3</td><td></td><td></td><td></td><td>0.083</td></tr> </table>		0	1	2	3	0	0.083	0.083	0	0.083	1		0.083	0.083	0	2			0.083	0.083	3				0.083	$f_0 = 0.05$ $f_1 = 0.02$ $f_2 = 2.3$ $f_3 = 0.99$ $f_4 = 0.46$ $f_5 = 0.5$
	0	1	2	3																							
0	0.083	0.083	0	0.083																							
1		0.083	0.083	0																							
2			0.083	0.083																							
3				0.083																							

**Fig. 8.** Examples of correlograms and features values.

value lowers as long as the number of non-null coefficients in the correlogram increase. Regions  $\Gamma_b$ ,  $\Gamma_d$  and  $\Gamma_e$  in Fig. 8 provide a few examples of such behavior. Feature  $f_1$  (*diagonal energy*) limits the computation of the energy value to the elements lying on the diagonal of the correlogram matrix. This in turn means that only regions characterized by a uniform color can attain  $f_1=1$ . The examples presented in Fig. 8 confirm that the value of  $f_1$  is somewhat correlated to the presence of uniform subregions in  $\Gamma$ . In this regard, regions  $\Gamma_a$  and  $\Gamma_c$ , which shared the energy value  $f_0=1$ , lie in the opposite extremes of the range in term of  $f_1$ .

Feature  $f_2$  (*entropy*) measures the average number of bits per symbol needed to encode  $Z_i^k$ . Therefore,  $f_2$  estimates the information content of a correlogram. In Fig. 8, regions  $\Gamma_a$  and  $\Gamma_c$  are characterized by  $f_2=0$ , since both of them can be entirely described by using only a single symbol. Conversely, entropy increases in those cases where several symbols (i.e.  $Z_i^k(i,j)$ ) should be used to encode the correlogram: regions  $\Gamma_d$  and  $\Gamma_e$  represent two examples of the latter behavior.

Feature  $f_3$  (*contrast*) basically associates different weights to the elements  $Z_i^k(i,j)$ ; in particular, weights increase as long as the elements lie farther from the principal diagonal. Hence,  $f_3$  estimates the contrast level in a region  $\Gamma$ . Fig. 8 shows that  $f_3$  takes on the highest values in regions  $\Gamma_b$  and  $\Gamma_c$ , which are characterized by numerous edges between the maximum and the minimum luminance value. Feature  $f_4$  (*homogeneity*) also is computed by associating different weights to the elements  $Z_i^k(i,j)$ . In this case, the goal is to evaluate the level of uniformity that characterizes the region  $\Gamma$ . Hence, as shown in Fig. 8, the presence of edges decreases the homogeneity and thus lowers the value of  $f_4$ .

Finally, feature  $f_5$  (*energy ratio*) computes the ratio between  $f_1$  and  $f_0$ . A high value of  $f_5$  is thus assigned to those regions that include large uniform areas.

## References

- [1] A.M. Eskicioglu, P.S. Fisher, Image quality measures and their performance, *IEEE Trans. Commun.* 43 (12) (1995) 2959–2965.
- [2] T.N. Pappas, R.J. Safranek, Perceptual criteria for image quality evaluation, in: A. Bovik (Ed.), *Handbook of Image and Video Processing*, Academic, New York, 2000.
- [3] Z. Wang, H.R. Sheikh, A.C. Bovik, Objective video quality assessment, in: B. Furth, O. Marques (Eds.), *The Handbook of Video Databases: Design and Applications*, CRC Press, Boca Raton, FL, 2003.
- [4] B.W. Keelan, *Handbook of Image Quality*, Marcel Dekker, New York, 2002.
- [5] E. Cambria, A. Hussain, Sentic Album: content, concept and context based online personal photo management system, *Cognitive Computation*, <http://dx.doi.org/10.1007/s12559-012-9145-4>, in press.
- [6] P. Gastaldo, R. Zunino, Neural networks for the no-reference assessment of perceived quality, *J. Electron. Imag.* 14 (3) (2005) 033004-1–033004-11.
- [7] R.V. Babu, S. Suresh, A. Perkis, No-reference JPEG image quality assessment using GAP-RBF, *Signal Process.* 87 (6) (2007) 1493–1503.
- [8] P. Le Callet, C. Viard-Gaudin, D. Barba, A convolutional neural network approach for objective video quality assessment, *IEEE Trans. Neural Networks* 17 (5) (2006) 1316–1327.
- [9] F.-H. Lin, R.M. Mersereau, Rate-quality tradeoff MPEG video encoder, *Signal Proc.: Image Commun.* 14 (4) (1999) 297–309.
- [10] W. Lu, K. Zeng, D. Tao, Y. Yuan, X. Gao, No-reference image quality assessment in contourlet domain, *Neurocomputing* 73 (4–6) (2010) 784–794.
- [11] M. Narwaria, W. Lin, Objective image quality assessment based on support vector regression, *IEEE Trans. Neural Networks* 21 (3) (2010) 515–519.
- [12] J. Redi, P. Gastaldo, I. Heynderickx, R. Zunino, Color distribution information for the reduced-reference assessment of perceived image quality, *IEEE Trans. Circuits Syst. Video Technol.* 20 (12) (2010) 1757–1769.
- [13] G.-B. Huang, D. Wang, Y. Lan, Extreme learning machines: a survey, *Int. J. Mach. Learn. Cybern.* 2 (2) (2011) 107–122.
- [14] G.-B. Huang, L. Chen, Enhanced random search based incremental extreme learning machine, *Neurocomputing* 71 (2008) 3460–3468.
- [15] G.-B. Huang, H. Zhou, X. Ding, R. Zhang, Extreme learning machine for regression and multiclass classification, *IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics* 42 (2) (2012) 513–529.
- [16] G. Wang, Y. Zhao, D. Wang, A protein secondary structure prediction framework based on the Extreme Learning Machine, *Neurocomputing* 72 (1–3) (2008) 262–268.
- [17] Y. Lan, Y.C. Soh, G.-B. Huang, Extreme Learning Machine based bacterial protein subcellular localization prediction, in: *Proceedings of the IEEE International Joint Conference on Neural Networks, IJCNN 2008, Hong Kong, 2008*, pp. 1859–1863.
- [18] R. Zhang, G.-B. Huang, N. Sundararajan, P. Saratchandran, Multicategory classification using an Extreme Learning Machine for microarray gene expression cancer diagnosis, *IEEE/ACM Trans. Comput. Biol. Bioinform.* 4 (3) (2007) 485–495.
- [19] A.A. Mohammed, R. Minhas, Q.M. Jonathan Wu, M.A. Sid-Ahmed, Human face recognition based on multidimensional PCA and extreme learning machine, *Pattern Recognition* 44 (10–11) (2011) 2588–2597.
- [20] A.H. Nizar, Z.Y. Dong, Y. Wang, Power utility nontechnical loss analysis with Extreme Learning Machine Method, *IEEE Trans. Power Syst.* 23 (3) (2008) 946–955.
- [21] S. Decherchi, P. Gastaldo, R.S. Dahiya, M. Valle, R. Zunino, Tactile data classification of contact materials using computational intelligence, *IEEE Trans. Robotics* 27 (3) (2011) 635–639.
- [22] S. Ridella, S. Rovetta, R. Zunino, Circular back-propagation networks for classification, *IEEE Trans. Neural Networks* 8 (1) (1997) 84–97.
- [23] S. Saha, R. Vemuri, An analysis on the effect of image activity on lossy coding performance, in: *Proceedings of the IEEE International Symposium on Circuits and Systems*, vol. 3, Geneva, Switzerland, 2000, pp. 295–298.
- [24] T.M. Kusuma, H.-J. Zepernick, On perceptual objective quality metrics for in-service picture quality monitoring, in: *Proceedings of the Third ATCrc Telecommunications and Networking Conference and Workshop*, Melbourne, Australia, 2003.
- [25] Z. Wang, E.P. Simoncelli, Reduced-reference image quality assessment using a wavelet-domain natural image statistic model, in: *Proceedings of the SPIE Human Vision and Electronic Imaging X*, vol. 5666, San Jose, CA, 2005, pp. 149–159.
- [26] Z. Wang, G. Wu, H.R. Sheikh, E.P. Simoncelli, E.H. Yang, A.C. Bovik, Quality-aware images, *IEEE Trans. Image Process.* 15 (6) (2006) 1680–1689.
- [27] M. Carney, P. Le Callet, D. Barba, Objective quality assessment of color images based on a generic perceptual reduced reference, *Signal Process.: Image Commun.* 23 (4) (2008) 239–256.
- [28] Q. Li, Z. Wang, General-purpose reduced-reference image quality assessment based on perceptually and statistically motivated image representation, in: *Proceedings of the IEEE International Conference on Image Processing, ICIP 2008, San Diego, CA, 2008*.
- [29] J. Huang, S. Ravi Kumar, M. Mitra, W.J. Zhu, R. Zabih, Image indexing using color correlograms, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR'97*, 1997, pp. 762–768.
- [30] H.R. Sheikh, Z. Wang, L. Cormack, A.C. Bovik: LIVE Image Quality Assessment Database at <http://live.ece.utexas.edu/research/quality>.
- [31] J. Redi, P. Gastaldo, R. Zunino, I. Heynderickx, Co-occurrence matrixes for the quality assessment of coded images, in: *Proceedings of the International Conference on Artificial Neural Networks, ICANN'08*, 2008.
- [32] J. Redi, P. Gastaldo, R. Zunino, I. Heynderickx, Reduced Reference Assessment of perceived quality by exploiting color information, in: *Proceedings of the Fourth International Workshop on Video Processing and Quality Metrics for Consumer Electronics, VPQM'09*, 2009.
- [33] T.P. Vogl, J.K. Mangis, A.K. Rigler, W.T. Zink, D.L. Alkon, Accelerating the convergence of the back-propagation method, *Biol. Cybern.* 59 (4–5) (1988) 257–263.
- [34] T. Evgeniou, M. Pontil, T. Poggio, Regularization networks and support vector machines, *J. Adv. Comput. Math.* 13 (1) (2000) 1–50.
- [35] M. Perrone, Improving Regression Estimates: Averaging Methods for Variance Reduction With Extension to General Convex Measure Optimization, Ph.D. Dissertation, Brown University, Providence, RI, 1993.
- [36] S. Geman, E. Bienenstock, R. Doursat, Neural networks and the bias/variance dilemma, *Neural Comput.* 4 (1) (1992) 1–48.
- [37] D.E. Rumelhart, J.L. McClelland, *Parallel Distributed Processing*, MIT Press, Cambridge, MA, 1986.
- [38] P. Bartlett, S. Boucheron, G. Lugosi, Model selection and error estimation, *Mach. Learn.* 48 (1–3) (2011) 85–113.
- [39] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, *IEEE Trans. Image Process.* 13 (4) (2004) 600–612.
- [40] VQEG: Final report from the video quality experts group on the validation of objective models of video quality assessment, 2003, Available: <http://www.vqeg.org>.



**Sergio Decherchi** obtained the "Laurea" degree summa cum laude in Electronics Engineering in 2007 from Genoa University, Italy. Since 2005 he started collaborating with the Department of Biophysics and Electronics Engineering of Genoa University. In April 2011 he obtained a Ph.D. on Machine Learning and Data Mining. From February 2011 he is Post-Doc researcher at the Department of Drug Discovery and Design of IIT, Genoa, Italy, where he develops computational and learning methods for biophysics. His main research areas include: theoretical and practical aspects of Machine Learning/Data Mining, computational methods for biophysics and drug design, molecular surfaces definition and processing. Sergio

Decherchi is author/co-author of more than 20 papers in peer reviewed journals and conferences; he is reviewer for several IEEE and Elsevier journals on computational intelligence.



**Paolo Gastaldo** obtained the “Laurea” degree in Electronic Engineering and a Ph.D. in Space Sciences and Engineering (2004), both from Genoa University, Italy. Since 2004 he is with the Department of Biophysical and Electronics Engineering (DIBE) of Genoa University; he is currently with DIBE as Assistant Professor. His main research area include innovative systems for visual signal understanding, neural network-based methods for nonlinear information processing, and DSP-based architectures for advanced signal interpretation, such as intelligent object tracking for video surveillance and cryptography.



**Rodolfo Zunino** (born Genoa, Italy, 1961) obtained the “Laurea” degree cum laude in Electronic Engineering from Genoa University in 1985. From 1986 to 1995 he was a research consultant with the Department of Biophysical and Electronic Engineering (DIBE) of Genoa University. He is currently with DIBE as an Associate Professor, teaching Electronics for Embedded Systems and Electronics for Security. His main scientific interests include intelligent systems for Computer Security, network security and Critical Infrastructure Protection, embedded electronic systems for neural networks, efficient models for data representation and learning, massive-scale text-mining and text-

clustering methods, and advanced techniques for multimedia data processing. Rodolfo Zunino coauthored more than 170 scientific papers in International Journals and Conferences; he has been the Co-Chairman of the two Editions of the International Workshop on Computational Intelligence for Security in Information Systems (CISIS’08 and CISIS’09). Since 2001 he is contributing as Associate Editor of the IEEE Transactions on Neural Networks, and has participated in the Scientific Committees of several International Events (ICANN’02, ICANN’09, IWPAAMS2004, IWPAAMS2005, Applied Computing 2006). Rodolfo Zunino is a Senior Member of IEEE (CIS—Computational Intelligence Society).



**Erik Cambria** received his B.Sc. and M.Sc. cum laude in Electronic Engineering from the University of Genoa, in 2005 and 2008 respectively. From 2009, he is the principal investigator of a project funded by the UK Engineering and Physical Sciences Research Council (EPSRC Grant Reference no. EP/G501750/1), which was born from the collaboration between the University of Stirling, the MIT Media Laboratory and Sitekit Labs for the development of next-generation intelligent web applications. Today Erik is also an associate scientist at the National University of Singapore, where he is part of a new Cognitive Science Program at Temasek Laboratories in collaboration with Singapore Ministry

of Defense. His interests include AI, Semantic Web, NLP, opinion mining and sentiment analysis, affective and cognitive modeling, HCI and e-health. Erik is also a chair of several international workshops, e.g., Health Web Science at WebSci11, SADMA at ISNN11 and SENTIRE at ICDM11, a reviewer of top AI journals, e.g., IEEE Transactions on Neural Networks, Elsevier Knowledge-Based Systems and Springer Language Resources and Evaluation, a Web Science lecturer at University of the Highlands and Islands, and a research affiliate at HP Labs India, Chinese Academy of Sciences and Microsoft Research Asia.



**Judith Redi** obtained the M.Sc. in Electronic Engineering in 2006 from Genoa University, Italy, and her Ph.D. in 2010 from the same University. Her Ph.D. thesis “Visual Overall Quality Assessment: a perception-driven framework based on Computational Intelligence techniques” received the “best engineering thesis award” from the University of Genoa in the same year. After being appointed as a post-doc at Eurecom, where she focused on digital image forensics and 3D face recognition, she moved to the Interactive Intelligence group of Delft University of Technology, where she is currently working as an Assistant Professor. Her main research interests concern the

understanding of Visual perception and Quality of the Visual Experience, and in the use of Computational Intelligence for modeling them.