https://doi.org/10.2352/issn.2694-118X.2020.LIM-13

Parameters optimization of the Structural Similarity Index

Illya Bakurov [a], *Marco Buzzelli* [b], *Mauro Castelli* [a], *Raimondo Schettini* [b], *and Leonardo Vanneschi* [a,c];
[a] *Nova Information Management School (NOVA IMS), Universidade Nova de Lisboa, 1070-312, Lisboa, Portugal*
[b] *Department of Computer Sciences Systems and Communications, University of Milano - Bicocca; Milan, Italy*
[c] *LASIGE, Departamento de Informática, Faculdade de Ciências, Universidade de Lisboa, 1749-016 Lisboa, Portugal*

Abstract

We exploit evolutionary computation to optimize the hand-crafted Structural Similarity method (SSIM) through a data-driven approach. We estimate the best combination of luminance, contrast and structure components, as well as the sliding window size used for processing, with the objective of optimizing the similarity correlation with human-expressed mean opinion score on a standard dataset. We experimentally observe that better results can be obtained by penalizing the overall similarity only for very low levels of luminance similarity. Finally, we report a comparison of SSIM with the optimized parameters against other metrics for full reference quality assessment, showing superior performance on a different dataset.

Introduction

Full-reference measures for Image Quality Assessment (IQA) provide a comparison value between a pristine reference image and a potentially corrupted version of the same image. These measures are used in a wide variety of applications, such as the assembly of pleasing photo-collages, perceptual image compression, and its subsequent transmission [21]. More recently, differentiable full-reference measures have also been exploited as a guiding loss for gradient-based learning of neural networks, in the fields of image generation and enhancement [10].

The Structural Similarity method (SSIM) [20] has been introduced as an effective full-reference measure for image quality assessment, showing a good correlation with the subjective evaluation provided by human observers (such as Mean Opinion Score - MOS) on standard datasets. Due to this correlation, it has been used as proxy evaluation for human assessment in different applications, such as image deblurring and super-resolution [7]. As a guiding optimization measure, however, other measures are often preferred, based on the comparison of visual features at different levels of abstraction [10]. A better approximation of human perception typically characterizes such measures, at the cost of higher computational time when compared to SSIM.

SSIM is a hand-crafted measure, computed from the comparison of luminance, structure, and contrast of the input pair. It is characterized by manually-defined parameters, such as exponentiation of each comparison element, and the window's size for its processing. In this work, we augment the hand-crafted SSIM through a data-driven approach, which allows reaching a higher correlation with the human response in terms of quality assessment. We exploit Evolutionary Computation (EC) techniques on standard image quality datasets to efficiently define the best combination of parameters for the application of the SSIM measure.

Background and Related Works

SSIM compares a reference image x and a corrupted image y, based on three independent components: luminance, contrast, and structure [21]. The luminance information is represented by each image's average (μ), thus the luminance comparison is:

$$l(x,y) = (2\mu_x\mu_y + C_1)/(\mu_x^2 + \mu_y^2 + C_1) \tag{1}$$

where C_1 is a small constant for numerical stability, as are C_2 and C_3 in the following equations for the other components. Contrast is represented through the use of standard deviation (σ), and consequently the contrast-based comparison is:

$$c(x,y) = (2\sigma_x\sigma_y + C_2)/(\sigma_x^2 + \sigma_y^2 + C_2) \tag{2}$$

Structure is computed by normalizing the images by the corresponding mean and variance. They are compared with the inner product, computed through their covariance σ_{xy}:

$$s(x,y) = (\sigma_{xy} + C_3)/(\sigma_x\sigma_y + C_3) \tag{3}$$

Finally, the three components are combined into:

$$SSIM(x,y) = [l(x,y)]^\alpha \cdot [c(x,y)]^\beta \cdot [s(x,y)]^\gamma \tag{4}$$

Statistics $\mu_{\{x,y\}}$, $\sigma_{\{x,y\}}$ and σ_{xy} are computed locally with a 11 × 11 gaussian weighting function, and eventually averaged.

Multiscale SSIM (MS-SSIM) [22] is an extension of SSIM based on its efficient application at different resolutions. The Visual Information Fidelity (VIF) [16] is computed by comparing the mutual information contained in each image. Several measures have also been recently developed in a data-driven fashion. Perceptual loss (PL) [10] exploits a neural network pretrained for image classification to extract features of each image at different levels of abstraction, which are then compared through mean square error. Amirshahi et al. [3] perform the comparison using traditional image quality metrics such as SSIM. Bosse et al. [4] explicitly train a neural network for regression of local full-reference image quality.

Proposed Method

Search Space

The work presented in this paper is conceptually divided into two steps, each exploring a different search space (S). First, we explore the optimization of relative importance of the three components of SSIM: luminance (α), contrast (β) and structure (γ); in this case, the sliding window's size is fixed at 11 as suggested in [21]; we denote this search space as $S_{(\alpha,\beta,\gamma)}$. Second, we extend the previous idea by including the sliding window's size as a parameter (w) in our optimization procedure; we denote this search space as $S_{(\alpha,\beta,\gamma,w)}$. Further experiments might include the optimization of stability constants C_1, C_2, C_3, although we reserve this for future works. By adapting nomenclature from EC, a given candidate-solution can be seen as a fixed-length chromosome of real-values which size varies from 3 to 4, depending if the window's size is included or not; formally, assuming $S_{(\alpha,\beta,\gamma,w)}$ a chromosome c at iteration i assumes the form $\vec{X}_{c,i} = [X_{\alpha,c,i},\ X_{\beta,c,i},\ X_{\gamma,c,i},\ X_{w,c,i}]$. Since the chromosome's values represent the exponents associated to each SSIM component, we bound them in the $(0,\ 3]$ interval, thus giving each component the chance to range from rooted form, to linear, quadratic, and up to cubic order. As a consequence of this decision, initial candidate-solutions were generated under continuous uniform distribution $\sim U(0,\ 3]$, regardless of the search algorithm and the search space.

Taking into consideration that the sliding window's size is an integer number and the fact we are approaching the problem from the perspective of continuous optimisation, we decided to create a special mapping from the set of admissible real-values in the chromosome, to the set of admissible values for the window's size. More specifically, the $(0,\ 3]$ interval was divided in 5 even sub-intervals, each representing an admissible window size $w \in [7,\ 9,\ 11,\ 13,\ 15]$.

Fitness Function

Since the goal of our Optimization Problem (OP) is to find a set of parameters for SSIM which maximises its similarity with MOS, we formalize the similarity measure f as Spearman's rank correlation coefficient (SRCC) between both measures, such that $f\ :\ S \rightarrow [-1,1]$, with higher values representing higher similarity, i.e., better fitness. The domain $\forall s \in S$ can be formally defined as $\mathbb{R}^3$ and $\mathbb{R}^4$ in $(0,\ 3]$, for $S_{(\alpha,\beta,\gamma)}$ and $S_{(\alpha,\beta,\gamma,w)}$ respectively.

Optimization Algorithms

The motivation behind the application of EC-based techniques is related to their adaptive capacity to learn and control their environments [14]. It is important to highlight that the objective of this paper is not to perform an exhaustive hyper-parameter exploration for the considered algorithms. Instead, our goal is to prove the suitability of the proposed method to optimize SSIM's parameters. For that, we experiment with a set of semantically diverse algorithms: Genetic Algorithm (GA), Differential Evolution (DE) and Particle Swarm Optimization (PSO).

Grid Search Given that the number of fitting parameters can be said small, one could be tempted to simply apply an exhaustive search of the parameters' space. However, this does not seem to be such an easy task for the following reasons. Although we bound the search-space in $(0,\ 3]$ hype-cube, the space is continuous which means that the set of candidate solutions is, in theory, infinite. However, even if one admits discretization of the continuous search-space, following results presented in this paper we consider 3 decimal points, and admitting only 3 fitting parameters (α, β, γ), there will be 3000^3 candidate-solutions to evaluate. Considering that one candidate-solution takes, in average terms, 5 seconds to be evaluated on the set of 1700 images, divided in batches of size 100, using a MSI GS65 Stealth Thin 8RF computer and GPU capabilities, then 3000^3 candidate-solutions will take 135000000000 seconds or 1562500 days. Whereas a single execution of an optimization heuristic like Genetic Algorithm parametrized as in our experiments, takes 250 seconds to generate one solution.

Genetic Algorithm (GA) Genetic Algorithm is a meta-heuristic introduced by Holland [9]. The algorithm starts with a random-like population of the candidate-solutions (called chromosomes). Then, by mimicking natural selection and genetically-inspired variation operators such as crossover and mutation, the algorithm breeds a population of next-generation candidate-solutions (called offspring population), which replaces the previous population (a.k.a. parents population). The procedure is iterated until reaching some stopping criteria, like a pre-defined number of iterations (also called generations).

In our experiments, GA was used with a tournament selection of size 2. The survival was elitist, always copying the best individual into the next generation. Given that chromosomes are vectors of real-valued numbers, we opted to use a geometric crossover and ball-mutation with probabilities 0.7 and 0.3 respectively. In this case, the resulting offspring always stands on the segment joining the points representing the parents in the search space. Box-mutation, which consists of a random perturbation of chromosome's values in a given range, was applied at every chromosome's position with a probability of 0.2 and the bound of perturbation was set to 1. No inversion was used.

Differential Evolution (DE) Differential Evolution is a stochastic and population-based meta-heuristic originally designed for solving continuous OPs by Storn and Price in 1995 [19, 18]. In DE, parents' selection is performed at random, meaning that all chromosomes have an equal probability of being selected for mating, regardless of their fitness value. The variation consists of two steps: mutation and crossover. The mutation creates an offspring based on a scaled difference between two randomly selected parents, added to a third population member.

The scaling factor F usually lies in $[0.4,\ 1]$ as reported in [6]. In a binomial crossover, the type of crossover we have used in our experiments, the elements of the resulting *mutant* (called the *donor*) are exchanged, with probability C_r, with the elements of one of the previously selected parents (called the *target*); that is, the crossover is performed on each of the D indexes of the donor with a probability C_r, by exchanging its values with the target. Finally, the best solution passes to the next iteration. In our experiments, DE was used with $F = 0.5$ and $C_r = 0.2$.

Particle Swarm Optimization (PSO) Particle Swarm Optimization is another form of stochastic and population-based meta-heuristic, developed by Eberhart and Kennedy in 1995 [11]. Following PSO's nomenclature, a population is called a swarm, and a candidate-solution a particle. In PSO, the position of particle p at iteration i, $\vec{x}_{p,i}$, is updated at each iteration based on a procedure that takes two components into account: the particle's and swarm's best-so-far positions. Formally, the procedure for particle's position update is defined as [17]: $\vec{x}_{(p,i)} = \vec{x}_{(p,\ i-1)} + \vec{v}_{(p,i)}$, such that $\vec{v}_{(p,i)} = w * \vec{v}_{(p,\ i-1)} + K_1 \vec{\phi}_1 (\vec{lbest}_p - \vec{x}_{(p,\ i-1)}) + K_2 \vec{\phi}_2 (\vec{gbest} - \vec{x}_{(p,\ i-1)})$, where $\vec{x}_{(p,i)}$ represents the position of particle p at iteration i, $\vec{lbest}_p$ and $\vec{gbest}$ represent the local and global best, respectively, with K_1 and K_2 being two positive constants used to scale their contribution. $\vec{\phi}_1$ and $\vec{\phi}_2$ are random vectors which follow $\sim U(0,\ 1)$ at each dimension. Following the above-mentioned definition of PSO's update-rule, the swarm's positions are updated taking into consideration the same version of the current global best. Alternatively, Carlisle and Dozier [5] have proposed an Asynchronous update (A-PSO), where global best is identified immediately after updating the position of each particle.

Throughout our experiments, in both variants of PSO, we have used equal weights for acceleration coefficients $K_1 = K_2 = 1.0$, inertia weight w was set equal to 0.6, and the components of $\vec{v}_{p,i}$ were constrained in $[-2,\ 2]$.

Experimental Setup and Results

Datasets

Experiments are conducted on two well-known datasets for assessment of image quality: TID2008 [15] and CSIQ [12]. These share a similar set of distortions, therefore allowing an investigation of the generalizability across datasets sampled from

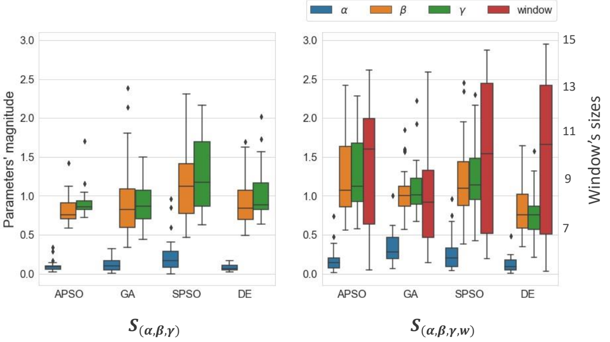

Figure 1: Distribution of SSIM parameters according to each algorithm. The left chart is related to the optimization of luminance, contrast and structure exponents. The right chart considers the joint optimization of exponents and window size.

similar distributions.

The first, **TID2008** [15], was used for the estimation of SSIM's parameters with the proposed method. TID2008 is composed of 25 reference images, each corrupted with 17 types of distortions at 4 different levels for each type of distortion, resulting in 1700 reference-distortion pairs. The visual quality of distorted images was subjectively evaluated through MOS of more than 800 volunteers of different cultural level from three countries have participated to the experiments.

The second dataset, **CSIQ** [12], was used to assess the method's generalization ability, i.e., the generalization of proposed parameters on a dataset created upon completely different reference images. CSIQ was built from 30 reference images, each corrupted with 6 types of distortions at 5 different levels, resulting in 900 reference-distortion pairs. The visual quality of distorted images was subjectively evaluated by 35 different volunteers. Unlike for TID2008, authors reported their results in the form of Differential Mean Opinion Scores (DMOS), where larger values stand for greater visual distortion when compared to the reference. For this reason, a negative correlation is expected between SSIM and DMOS.

Wang et al. [1] argument that SS-SSIM is most effective if used at the appropriate scale, which depends on both the image resolution and the viewing distance. For this reason, all images are rescaled according to the empirical formula provided by the authors. The viewing distance is fixed for both datasets, although it would be interesting to further experiment on datasets which incorporate evaluation at varying distance levels, such as VDID [8] and CID:IQ [13].

Parameters

All algorithms were executed for 30 generations with a population size equal to 20. To account for the algorithms' stochastic nature and provide a statistically sustained analysis of the experimental results, we repeated the experiments for 30 times (runs), each with a different seed for the pseudo-random number generator. During training, we have left away 30% of the reference-distortion pairs for estimation of the algorithms' generalization ability to then have the possibility to compare these estimates with the (real) fitness observed on a previously unseen dataset. To accelerate our algorithmic procedures, we decided to use only 50% of the reference-distortion pairs from the training partition, selected at random and without replacement at the beginning of each iteration. When estimating SSIM's parameters on TID2008 we defined the similarity measure f as SRCC between the proposed SSIM and MOS. Nevertheless, after estimating the parameters based on SRCC, we also assessed both Pearson's (PCC) and Kendall's (KRCC) correlation coefficients.

Table 1: Statistics on SRCC between optimized SSIM and MOS, obtained with different optimization algorithms. We report the sample mean ($\bar{x}$), median ($\widetilde{x}$) and standard deviation (s).

Algorithms	TID2008$_{training}$			TID2008$_{test}$		
	$\bar{x}_{SRCC}$	$\widetilde{x}_{SRCC}$	s_{SRCC}	$\bar{x}_{SRCC}$	$\widetilde{x}_{SRCC}$	s_{SRCC}
Baseline	0.767	0.767	0.007	0.770	0.771	0.016
$APSO_{(\alpha,\beta,\gamma)}$	0.830	0.831	0.008	0.798	0.797	0.017
$DE_{(\alpha,\beta,\gamma)}$	0.832	0.832	0.007	0.804	0.807	0.014
$GA_{(\alpha,\beta,\gamma)}$	0.806	0.807	0.013	0.809	0.810	0.012
$SPSO_{(\alpha,\beta,\gamma)}$	0.826	0.826	0.007	0.806	0.807	0.014
$APSO_{(\alpha,\beta,\gamma,w)}$	0.840	0.838	0.009	0.795	0.797	0.028
$DE_{(\alpha,\beta,\gamma,w)}$	0.841	0.840	0.008	0.816	0.822	0.024
$GA_{(\alpha,\beta,\gamma,w)}$	0.834	0.834	0.008	0.794	0.796	0.025
$SPSO_{(\alpha,\beta,\gamma,w)}$	0.839	0.838	0.012	0.799	0.800	0.027

Experimental Results

After 30 iterations of each run, we selected the best solution of each algorithm based on its performance on unseen data.

Figure 1 reports the distribution of parameters obtained by each algorithm, after repeating the experiments 30 times. The sub-figure on the left regards the study of $S_{(\alpha,\beta,\gamma)}$, whereas the sub-figure on the right regards the study of $S_{(\alpha,\beta,\gamma,w)}$. All the algorithms, regardless of the search space, suggest that the three components should have a different impact on the overall SSIM computation. The luminance, contrast, and structure components can be considered as independent pseudo-probabilities, since each is constrained between 0 and 1, and they are multiplied to compute the joint probability associated to the overall image similarity. By raising all components to a lower-than-one exponent, as found by our optimization process, we are in fact increasing each probability, with the effect of being less penalizing on the final similarity score. In particular, the luminance component is subject to a stronger distortion due to the extremely-low exponent found, meaning that it will impact the SSIM only for very low values of luminance similarity. We consider this data-driven finding as an important achievement since scientific community essentially uses $\alpha = \beta = \gamma = 1$, as suggested in [21].

The right sub-figure shows that, in median terms, almost all the algorithms agree upon window's size, 11, which is consistent with the literature [21]. GA is the only algorithm in which the median is one level below (9); nevertheless, this does not decrease GA's performance. This can be seen from Table 1, which reports the sample mean ($\bar{x}$), median ($\widetilde{x}$) and standard deviation (s) of SRCC achieved by each algorithm, on the training and test partitions of the TID2008 database, after 30 runs. It also includes SSIM's statistics when its components are subject to no exponentiation ($\alpha = \beta = \gamma = 1$) and the window's size equal to 11, as suggested in [21]. The latter is reported as the *baseline*.

From the analysis of Table 1, one can observe not only the large difference with the baseline but also the fact that the algorithms' performance significantly differs from one search space, denoted as $Algorithm_{(\alpha,\beta,\gamma)}$, to another, denoted as $Algorithm_{(\alpha,\beta,\gamma,w)}$. One can clearly see that the latter achieves higher SRCCs on training partition whereas the former higher SRCCs on unseen partitions. This factor suggests that algorithms overfit more on $S_{(\alpha,\beta,\gamma,w)}$; nevertheless, the best-expected performance is achieved on $S_{(\alpha,\beta,\gamma,w)}$ by Differential Evolution (DE). Another insight can be derived from the analysis of standard deviation s: although on $S_{(\alpha,\beta,\gamma)}$ algorithms tend to present higher stability from one run to another, which can be related to previous observation about their generalization ability, in general

Table 2: Enumeration of the suggested parameters, and comparison of the corresponding performance with other image quality assessment measures on the TID2008 and CSIQ datasets.

Algorithms	Parameters				TID2008_{full}			CSIQ_{full}		
	α	β	γ	w	SRCC	PCC	KRCC	SRCC	PCC	KRCC
SSIM-$SPSO_{(\alpha,\beta,\gamma)}$	0.054	0.789	0.843	11	0.811	0.769	0.613	-0.923	-0.843	-0.751
SSIM-$GA_{(\alpha,\beta,\gamma)}$	0.062	0.731	0.883	11	0.811	0.770	0.612	-0.925	-0.848	-0.752
SSIM-$DE_{(\alpha,\beta,\gamma,w)}$	0.063	0.529	0.554	13	0.821	0.756	0.623	-0.916	-0.826	-0.743
SSIM-$DE'_{(\alpha,\beta,\gamma,w)}$	0.009	0.826	0.779	7	0.821	0.775	0.620	-0.923	-0.833	-0.751
SSIM (default) [21]	1	1	1	11	0.773	0.739	0.575	-0.861	-0.780	-0.673
MS-SSIM [22]	-	-	-	11	0.838	0.784	0.641	-0.893	-0.709	-0.714
VIFP [16]	-	-	-	-	0.653	0.637	0.494	-0.880	-0.883	-0.696
VGG-based PL [10]	-	-	-	-	-0.572	-0.545	-0.409	0.788	0.738	0.571

Table 3: Average SRCC on each distortion type for TID2008.

Distortion type	$\overline{SRCC}_{SSIM(default)}$	$\overline{SRCC}_{SSIM-DE}$
Noise	0.825	0.833
Noise2	0.783	0.794
Safe	0.851	0.854
Hard	0.824	0.828
Simple	0.896	0.900
Exotic	0.679	0.668
Exotic2	0.742	0.735

terms all algorithms present a fair stability: from 0.007 to 0.013 on training and from 0.012 to 0.027 on the test partition.

The algorithms' median was compared to the baseline's employing Wilcoxon's paired signed-rank test with a Bonferroni correction and significance level $a = 0.05$, under the null hypothesis that the median difference between pairs of observations is zero. Since for all the pairs algorithm-baseline the null hypothesis was rejected, we do not report the tests' statistics table.

Recommended parameters

The objective of this sub-section is to provide a concrete solution for the initially defined objective: the maximization of SSIM's similarity with MOS. Table 2 reports the performance of eight image quality assessment measures on the two described datasets: TID2008 [15] and CSIQ [12]. It is worth noticing that the table's results were obtained from executing the measures on the full set of reference-distortion pairs in each dataset.

The first rows regard four sets of suggested parameters for SSIM, obtained during our experiments. More specifically, the $SPSO_{(\alpha,\beta,\gamma)}$ and $GA_{(\alpha,\beta,\gamma)}$ were obtained from the study of SSIM's components relative importance. The last two, $DE_{(\alpha,\beta,\gamma,w)}$ and $DE'_{(\alpha,\beta,\gamma,w)}$, were obtained from the extended study which also included the window's size.

For comparison, we report the performance of four existing measures. The first measure is the baseline SSIM, parametrized as in [21]. The second measure is a multi-scale extension of SSIM (MS-SSIM), which incorporates different variations of viewing conditions [22]. It is a more complete measure, although computationally more demanding since it involves computation of several SSIM's components at 5 different scales of the reference-distortion pairs. The third measure is the pixel-based version of Visual Information Fidelity (VIFP) [16], a measure that combines the information present in the reference image with the quantification of how much of this reference information can be extracted from the distorted image. The fourth measure is called Perceptual Loss (PL) and it is based on a pre-trained Convolutional Neural Network (CNN), called VGG-16, on ImageNet database [10]. More specifically, PL calculates the aggregated perceptual differences in content and style between features of the reference-distortion pairs, obtained at different levels of the network.

From Table 2 it is possible to observe the superior performance of SSIM with the proposed parameters. On the TID2008 database, one can see that SSIM's similarity was raised almost to the level of semantically more complete and computationally more complex MS-SSIM. The difference between them, in terms of SRCC, was reduced from 0.065 to 0.017. On the CSIQ database, one can observe that SSIM under the parameters we propose exhibits the highest similarity when compared to other measures, except for MS-SSIM when measured in terms of PCC.

Table 3 shows a more detailed comparison between SSIM with default parameters, and SSIM optimized through Differential Evolution, grouping the distortion types into clusters as indicated in [2]. The column $\overline{SRCC}_{SSIM(default)}$ represents the average SRCC calculated from SSIM with default parameter set, whereas $\overline{SRCC}_{SSIM-DE}$ represents calculations from one of the proposed sets of parameters (see Table 2).

Conclusions

In this work, we exploited Evolutionary Computation (EC) meta-heuristics to explore SSIM parameters to increase its similarity with human Mean Opinion Score (MOS) on the TID2008 dataset. Our experiments proved the suitability of the proposed approach, as the obtained results pointed to significantly superior similarity with MOS when compared to the baseline. Moreover, SSIM with suggested parameters is proved to be better or comparable to other more computationally expensive measures on a completely different dataset (CSIQ). The optimized parameters provided by our approach present an interesting insight in the application of SSIM, suggesting that the luminance-based component should negatively impact the overall similarity score only for very low levels of luminance similarity. In the future we intend to further investigate the sensitivity of our optimization approach to different amounts of training data, and to evaluate its generalization ability in optimizing other measures for image quality assessment.

Acknowledgments

This work was supported by FCT (Fundação para a Ciência e a Tecnologia), Portugal, through funding of LASIGE Research Unit (UIDB/00408/2020) and projects BINDER (PTDC/CCI-INF/29168/2017), GADgET (DSAIPA/DS/0022/2018), AICE (DSAIPA/DS/0113/2019) and PREDICT (PTDC/CCI-CIF/29877/2017). Mauro Castelli also acknowledges financial support from the Slovenian Research Agency (research core funding No. P5-0410) and support from NVIDIA Corporation.

References

[1] The ssim index for image quality assessment. `https://www.cns.nyu.edu/~lcv/ssim/`, 2003. Accessed: 01.10.2019.

[2] Tampere image database 2008 - tid2008, version 1.0. `http://www.ponomarenko.info/tid2008.htm`, 2008. Accessed: 01.10.2019.

[3] S. A. Amirshahi, M. Pedersen, and A. Beghdadi. Reviving traditional image quality metrics using cnns. In *Color and Imaging Conference*, volume 2018, pages 241–246. Society for Imaging Science and Technology, 2018.

[4] S. Bosse, D. Maniry, K.-R. Müller, T. Wiegand, and W. Samek. Deep neural networks for no-reference and full-reference image quality assessment. *IEEE Transactions on Image Processing*, 27(1):206–219, 2017.

[5] A. Carlisle and G. Dozier. An off-the-shelf pso. In *Proceedings of Workshop on Particle Swarm Optimization*, pages 1–6, 2001.

[6] S. Das and P. Suganthan. Differential evolution: A survey of the state-of-the-art. *IEEE Trans. Evolutionary Computation*, 15:4–31, 01 2011.

[7] W. Dong, L. Zhang, G. Shi, and X. Wu. Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization. *IEEE Transactions on Image Processing*, 20(7):1838–1857, 2011.

[8] K. Gu, M. Liu, G. Zhai, X. Yang, and W. Zhang. Quality assessment considering viewing distance and image resolution. *IEEE Transactions on Broadcasting*, 61(3):520–531, 2015.

[9] J. H. Holland. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence*. MIT Press, Cambridge, MA, USA, 1992.

[10] J. Johnson, A. Alahi, and F. F. Li. Perceptual losses for real-time style transfer and super-resolution. volume 9906, pages 694–711, 10 2016.

[11] J. Kennedy and R. Eberhart. Particle swarm optimization. In *Proceedings of ICNN'95 - International Conference on Neural Networks*, volume 4, pages 1942–1948 vol.4, Nov 1995.

[12] E. Larson and D. Chandler. Most apparent distortion: Full-reference image quality assessment and the role of strategy. *J. Electronic Imaging*, 19:011006, 01 2010.

[13] X. Liu, M. Pedersen, and J. Y. Hardeberg. Cid: Iq–a new image quality database. In *International Conference on Image and Signal Processing*, pages 193–202. Springer, 2014.

[14] M. Mitchell. *An Introduction to Genetic Algorithms*. MIT Press, Cambridge, MA, USA, 1998.

[15] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, and F. Battisti. Tid2008-a database for evaluation of full-reference visual quality assessment metrics. *Advances of Modern Radioelectronics*, 10(4):30–45, 2009.

[16] H. R. Sheikh and A. C. Bovik. Image information and visual quality. *Trans. Img. Proc.*, 15(2):430–444, Feb. 2006.

[17] Y. Shi and R. Eberhart. A modified particle swarm optimizer. *1998 IEEE International Conference on Evolutionary Computation Proceedings. IEEE World Congress on Computational Intelligence (Cat. No.98TH8360)*, pages 69–73, 1998.

[18] R. Storn. On the usage of differential evolution for function optimization. pages 519 – 523, 07 1996.

[19] R. Storn and K. Price. Differential evolution: A simple and efficient adaptive scheme for global optimization over continuous spaces. *Technical Report TR-95-012, ICSI*, 23, 03 1995.

[20] Z. Wang and A. C. Bovik. A universal image quality index. *IEEE Signal Processing Letters*, 9:81–84, 2002.

[21] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE TRANSACTIONS ON IMAGE PROCESSING*, 13(4):600–612, 2004.

[22] Z. Wang, E. P. Simoncelli, and A. C. Bovik. Multiscale structural similarity for image quality assessment. In *The Thirty-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, volume 2, pages 1398–1402. Ieee, 2003.

Author Biography

Illya Bakurov is a PhD student at NOVA Information Management School (NOVA IMS), Universidade Nova de Lisboa, Portugal. His research activity focuses on Evolutionary Computation and its applications in field of medicine, image processing and computer vision.

Marco Buzzelli obtained his Ph.D. in Computer Science at University of Milano - Bicocca (Italy) in 2019, focusing on automatic description and annotation of complex scenes in digital images. He is currently a post-doctoral researcher, working on various Image Processing and Computer Vision tasks. His main topics of research include characterization of digital imaging devices, and object recognition in complex scenes.

Mauro Castelli has completed a PhD degree in Computer Science by University of Milano - Bicocca. He was assistant lecturer at the University of Milano - Bicocca and at the University of Bergamo. Currently he is associate professor at NOVA Information Management School (NOVA IMS), Universidade Nova de Lisboa, Portugal. His scientific activity spans the following areas of Evolutionary Computation: Genetic Programming, Genetic Algorithms, Swarm Intelligence, Artificial Intelligence, Machine Learning, Soft Computing, Heuristics for Combinatorial Optimization.

Raimondo Schettini is a professor at the University of Milano - Bicocca (Italy). Currently he is head of the Imaging and Vision Lab. He has been a team leader in several research projects and published more than 300 refereed papers and six patents about color reproduction, and image processing, analysis, and classification. He is a fellow of the International Association of Pattern Recognition for his contributions to pattern recognition research and color image analysis.

Leonardo Vanneschi is a full professor at NOVA Information Management School, Universidade Nova de Lisboa, Portugal. His main research interests involve Machine Learning, in particular Evolutionary Computation. His work can be partitioned into theoretical studies of Evolutionary Computation, and applicative work. He has published more than 200 contributions and he has led several research projects in the area.