

RESEARCH ARTICLE

NoR-VDPNet++: Real-Time No-Reference Image Quality Metrics

FRANCESCO BANTERLE¹, ALESSANDRO ARTUSI², ALEJANDRO MOREO¹,
FABIO CARRARA¹, AND PAOLO CIGNONI¹

¹ISTI-CNR, 56124 Pisa, Italy

²DeepCamera, CYENS CoE, 1016 Nicosia, Cyprus

Corresponding author: Francesco Banterle (francesco.banterle@isti.cnr.it)

The work of Francesco Banterle was supported by the European Union Horizon 2020 Programme under Grant “ENergy aware BIM Cloud Platform in a CoSt-effective Building RENovation Context” (ENCORE) No. 820434. The work of Alessandro Artusi was supported by the European Union Horizon 2020 Research and Innovation Program under Grant “Research Centre on Interactive Media, Smart System and Emerging Technologies” – RISE No. 739578. For this work, Alessandro Artusi has received funding from the Government of the Republic of Cyprus through the Deputy Ministry of Research, Innovation and Digital Policy.

ABSTRACT Efficiency and efficacy are desirable properties for any evaluation metric having to do with Standard Dynamic Range (SDR) imaging or with High Dynamic Range (HDR) imaging. However, it is a daunting task to satisfy both properties simultaneously. On the one side, existing evaluation metrics like HDR-VDP 2.2 can accurately mimic the Human Visual System (HVS), but this typically comes at a very high computational cost. On the other side, computationally cheaper alternatives (e.g., PSNR, MSE, etc.) fail to capture many crucial aspects of the HVS. In this work, we present NoR-VDPNet++, a deep learning architecture for converting full-reference accurate metrics into no-reference metrics thus reducing the computational burden. We show NoR-VDPNet++ can be successfully employed in different application scenarios.

INDEX TERMS Deep learning, HDR imaging, objective metrics, no-reference.

I. INTRODUCTION

The quality of natural/synthetic images is commonly assessed either through user studies or through objective metrics. This step is especially important to assess the quality of a compression/restoring/enhancing algorithm.

Although a user study is very reliable in terms of the quality of results, it is rather cumbersome to run since a considerable amount of time (spanning from weeks to months in some cases) is often required, given that a large number of participants and images should be involved. Furthermore, such studies require a careful design to avoid biases, and they can be very expensive since some money has to be invested in order to attract the participants' interest. As a result, objective metrics are typically preferred for assessing image quality; the monetary cost is greatly reduced while, at the same time, these metrics can be employed to evaluate real-time applications. Although such metrics do not require users, they represent fairly reliable solutions, especially when these

metrics take into account different aspects of the human visual system (HVS) or when they provide an accurate simulation of relevant aspects of it. An example of this last case is HDR-VDP 2.2 [32], which is, by now, a well-established metric for HDR and SDR imaging used in standardization committees. Unfortunately, HDR-VDP 2.2 presents two main limitations: i) its high computational cost prevents its use in real-time applications or large databases (e.g., standardization); ii) it requires a reference image, which may not be available in many cases (e.g., TV live broadcasting). Recently, some efforts have been paid into designing more computationally efficient metrics for specific problems. However, the most popular and reliable metrics, such as TMQI [27], [37] for assessing the quality of tone-mapped images, still require a reference image, which is a severe limiting factor.

All the above-mentioned problems make it evident the necessity of new objective metrics that (i) can be run in real-time, (ii) do not require a reference image or a ground truth, and (iii) mimic accurately the original reference-based metrics. In this paper, we propose NoR-VDPNet++, an efficient

The associate editor coordinating the review of this manuscript and approving it for publication was Yun Zhang¹.

deep learning architecture for converting full-reference accurate metrics into no-reference metrics.

The rest of this paper is organized as follows. In Section II, we review previous efforts in the field. In Section III, we explain our NoR-VDPNet++ architecture in detail. In Section IV, we turn to describe the dataset and the training strategy we use, while in Section V, we report our experiments. In Section VI, we demonstrate how our system fares in real applications. Finally, Section VII wraps up, also offering pointers to potential directions for future work.

II. RELATED WORK

Nowadays, image quality evaluation through the use of objective metrics has become of high importance. Objective metrics are not only used for quality assessment in benchmark studies but are also used to monitor/guide the performance of algorithms such as 3D renderers, encoders, enhancement, deep-learning training, etc. In this work, we consider Image Quality Metrics (IQMs) which predict a single global quality score for the entire image.

IQMs can be divided into Fully-Reference (FR) and No-Reference (NR) methods. While FR-metrics receive as input a pair of images (i.e., the ground truth and the distorted images), the NR-metrics has only the distorted image as input. In this section, we will focus on state-of-the-art NR-based IQMs approaches only, which is the scenario of our study.

Typically, NR IQMs are based on statistical information derived from the distorted image [19], [31], [35], [40]. For example, NIQE [31] first computes 36 highly regular natural scene statistics from an input image, to then compute the distance from these and a multi-variate data-driven Gaussian fit. Recently, NR metrics based on machine learning made their appearance. Mittal et al. [30] proposed to extract locally normalized luminance coefficients (LNLCS) to quantify possible losses of naturalness in the image due to the presence of distortions. Subsequently, a support vector regressor (SVR) is trained to predict, from LNLCS, a proxy of human perception called BRIQUE index. Similarly, Kundu et al. [26] introduced Higrade, an NR-metric for tone-mapped images based on the extraction of log-domain gradients and an SVR. Regarding convolutional neural networks (CNNs), Kang et al. [22] proposed one of the first approaches where they presented a simple NR CNN architecture for predicting quality scores that correlate with user experiments. Kottayil et al. [24] introduced an NR-IQA deep learning scheme for HDR images that correlates with mean opinion scores. Kim and Lee [23] deal with the absence of ground truth by employing local quality maps derived by FR-IQMs as intermediate regression targets. This approach requires pre-training the FR-IQM model on data where the ground truth is available. The approach by Bosse et al. [11] is purely data-driven and does not rely on hand-crafted features or other types of prior domain knowledge about the HVS or image statistics. Zhu et al. [39] proposed MetaQA to improve the prediction capabilities of

a CNN-based metric through pre-trained architectures. Here the meta-knowledge shared by people during the evaluation of the quality of images with various distortions is learned and adapted to unknown distortions.

Recently, CNN-based architectures have been employed to transfer the knowledge of an algorithm into the parameters of a convolutional network able to produce real-time predictions. This was achieved for both the FR scenario [2] (i.e., DIQM which mimics HDR-VDP 2.2 [32] and DRIIM [5] with uses a reference) and the NR scenario [9] (i.e., NoRVDP-Net which mimics HDR-VDP 2.2 without a reference).

In this work, we present NoR-VDPNet++, an improved variant of NoR-VDPNet [9] that achieves higher accuracy while maintaining its real-time nature. In particular, we present the following contributions with respect to previous art:

- NoR-VDPNet++ is a NR version of FR CNN-based metric [2], which distills HDR-VDP 2.2 and DRIIM [5] with high accuracy and efficiency. In this work, we extend NoR-VDPNet architecture testing normalization layers.
- We apply NoR-VDPNet and NoR-VDPNet++ to obtain a no-reference TMQI [37] to assess the quality of tone-mapped images and a no-reference HDR-VDP 2.2 to assess the quality of inverse tone-mapped images.
- We present two novel datasets: the former composed of tone-mapped HDR images using different tone mapping operators, and the latter composed of inverse tone-mapped images using different inverse tone mapping operators.

III. DISTILLING IMAGE QUALITY METRICS

NoR-VDPNet [9] accomplishes the conversion of HDR-VDP 2.2 [32] into a NR model encoded as a CNN. This is attained by training a CNN architecture (see the left-most architecture in Figure 1) using a medium dataset (e.g., more than 50,000-100,000 examples with/without reference) of SDR/HDR images for different scenarios such as SDR distortions detection (blur, noise, quantization, etc.), JPEG-Xt [3] compression artifacts, tone/inverse tone mapping evaluation, etc. Each example pair consists of a distorted image and the ground truth quality value that HDR-VDP 2.2 or TMQI calculates using its reference; see Figure 2. Note that the key for distilling HDR-VDP 2.2 or TMQI into a no-reference metric comes down to omitting the reference during training.

In this work, we explore techniques aimed at improving the stability of the training phase of the previous version NoR-VDPNet and increasing accuracy at inference time. The resulting model, which we dub NoR-VDPNet++, comes in two flavors, one that uses Batch Normalization [21] as a way to counter the internal covariate shift, and another that instead uses the more recent ReZero [6] normalization layer to speed up training convergence. We experiment with both variants and discuss the merits of each.

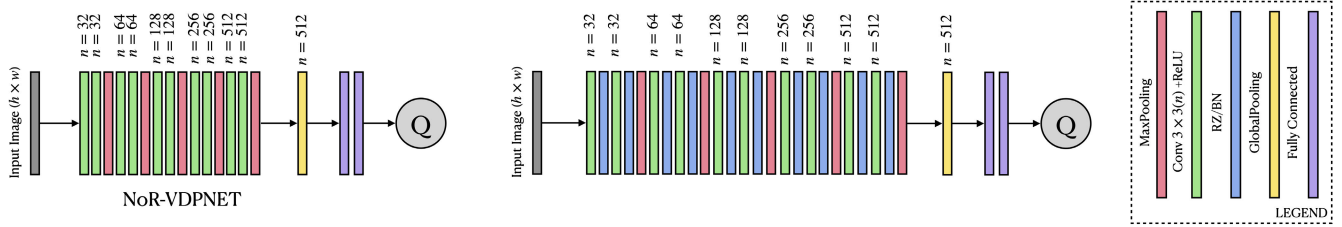


FIGURE 1. The network architecture of NoR-VDPNet (left) and NoR-VDPNet++ (right): Batch Normalization or ReZero are added to each convolution layer.

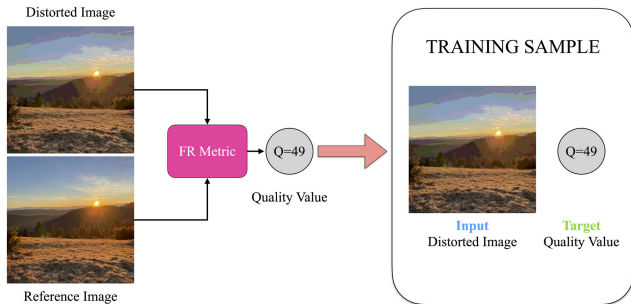


FIGURE 2. The process for creating a sample for our NR datasets. We use a FR metric for computing the quality value between the ground truth and the distorted images. Then, the sample is created by discarding the ground truth; the input for the network is the distorted image and the target output to minimize is the computed quality value Q .

Batch Normalization [21] has been shown to effectively help reduce the covariate shift between layers and to allow for faster and more robust training. Batch Normalization comes down to independently re-centering and re-scaling the dimensions of data tensors by using an approximation of the mean and standard deviation computed on the batch of examples. Equation 1 describes the computation for the k th dimension of a vector $\mathbf{x} = (x^{(1)}, \dots, x^{(m)})$; $\mu_B^{(k)}$ and $\sigma_B^{(k)}$ stand for the sample mean and sample standard deviation, respectively, as computed on the batch of examples B .

$$\hat{x}^{(k)} = \frac{x^{(k)} - \mu_B^{(k)}}{\sqrt{(\sigma_B^{(k)})^2 + \epsilon}} \quad (1)$$

ReZero [6], on the other hand, was recently proposed as a novel way for reducing the problems of vanishing and exploding gradients typical of deep learning training with residual layers. As a residual block, it allows deep architectures to become deeper while at the same time being much more efficient than other normalization techniques. The computation of ReZero between two subsequent layers (l and $l + 1$) is described by Equation 2 and comes down to a residual connection with a trainable parameter (α_l) used to modulate the transformation F of the data tensor.

$$\mathbf{x}_{l+1} = \mathbf{x}_l + \alpha_l F(\mathbf{x}_l). \quad (2)$$

Both variants of NoR-VDPNet++ achieve a lower prediction error than the original NoR-VDPNet and still preserve

real-time performance. When equipped with the ReZero connections, NoR-VDPNet++ produces lower errors in some scenarios; Figure 1 shows NoR-VDPNet before (left) and after (right) these changes.

IV. DATASETS AND TRAINING

We trained NoR-VDPNet++ for different scenarios:

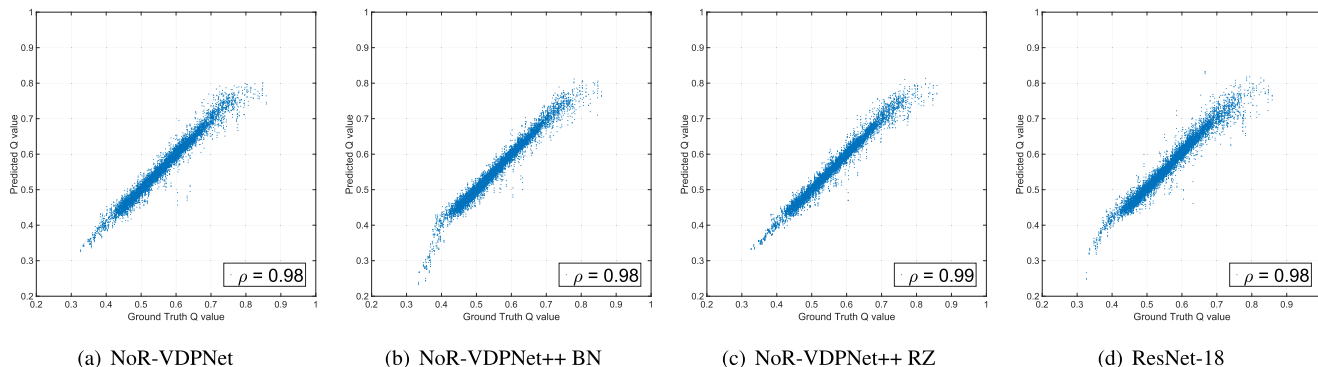
- **SDR-D:** Estimating HDR-VDP2.2 [32] quality value at different SDR distortions; e.g., blur, noise, quantization, etc. In this case, we converted 8-bit values into display referred values.
- **TMO:** Estimating TMQI [37] score under different tone mapping operators (TMOs).
- **HDR-C:** Estimating HDR-VDP2.2 [32] quality value at different JPEG-XT [4] quality settings.
- **ITMO:** Estimating HDR-VDP2.2 [32] score under different inverse tone mapping operators (TMOs).

For HDR-C, TMO, and ITMO datasets, we employed 1,478 HDR images or I_{HDR} : HDR Survey [14], Stanford HDR Dataset [36], the 100-sample from the Laval HDR Panorama dataset [18], Funt et al.’s HDR Dataset [16], Akyüz’s Dataset [1], the UBC HDR video dataset [7], and Stuttgart HDR Video dataset [15].

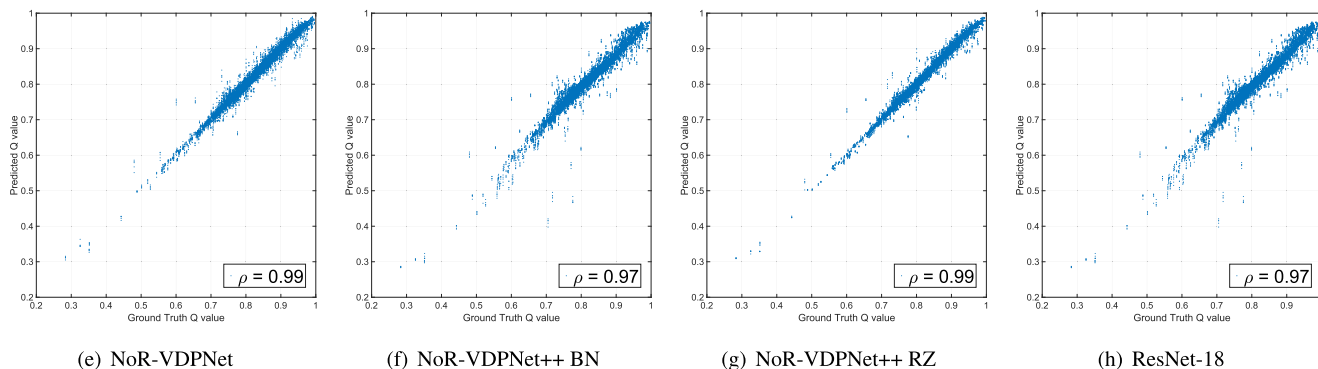
For the TMO dataset, we applied 18 TMOs (see Figure 6) to all images in I_{HDR} using the HDR Toolbox [8]. Then, we ran TMQI using the original HDR images and their tone-mapped versions, storing the TMQI score as the target output. The no-reference dataset comprises the tone-mapped images stored at 8-bit per color channel in the sRGB color space and its TMQI score.

Regarding ITMO, we applied six inverse tone mapping operators (ITMOs) to the SDR versions (i.e., with a f-stop that maximizes the total well-exposed pixels) of the HDR images in I_{HDR} . These operators are: Akyuz et al. [1], Huo et al. [20], Kovaleski and Oliveira [25], and Masia et al. [29], Eilertsen et al. [13], and Santos et al. [33]. We ran HDR-VDP 2.2 between the original HDR images and their inverse tone-mapped one storing the HDR-VDP 2.2 Q value. The no-reference dataset comprises inverse tone-mapped images stored at 32-bit per color channel in the sRGB color space and its HDR-VDP 2.2 score. To further stress-test different input conditions, we applied an exposure augmentation; i.e., we applied a +1.5-stop and a 3.0-stop increase from a

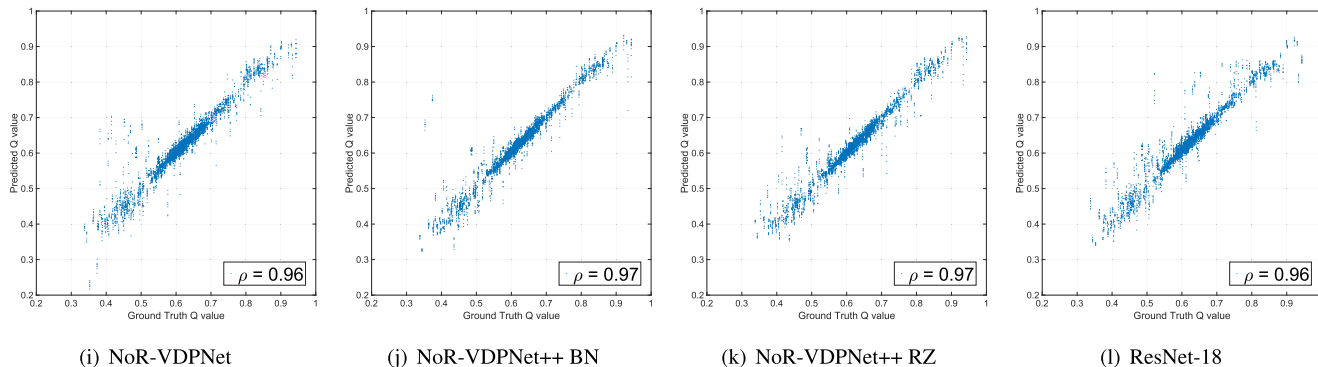
SDR-D



TMO



HDR-C



ITMO

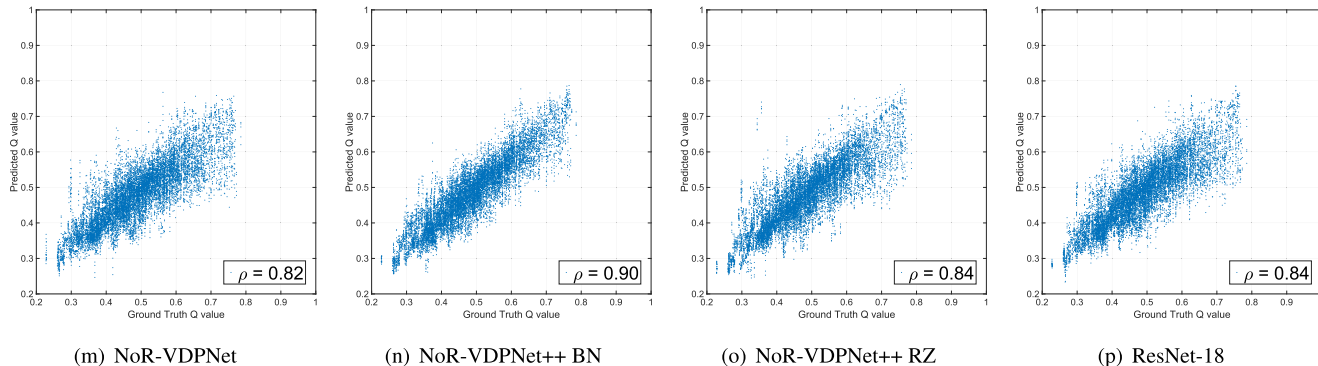
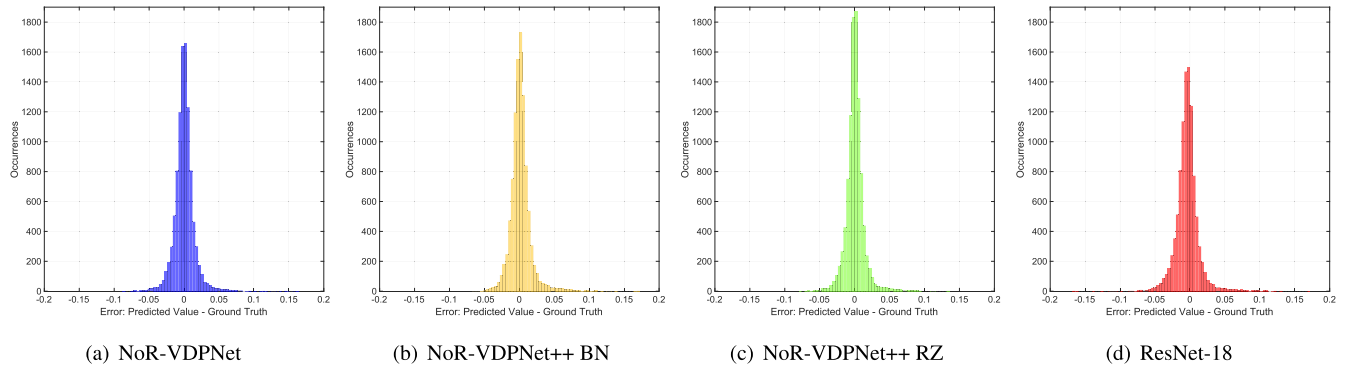
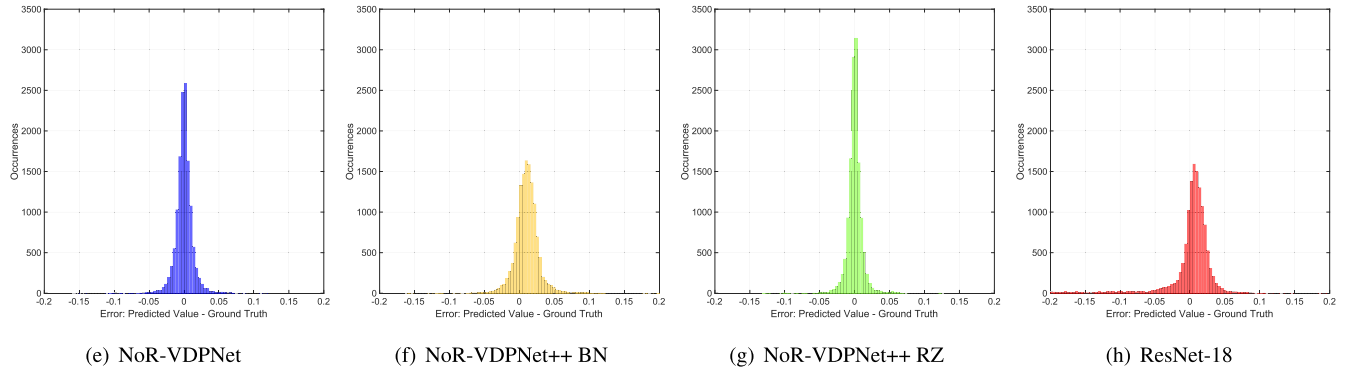


FIGURE 3. Scatter plots for the test datasets of all scenarios where we compute the Pearson coefficient of correlation, ρ . From the top to the bottom: HDR-C, ITMO, SDR-D, and TMQI.

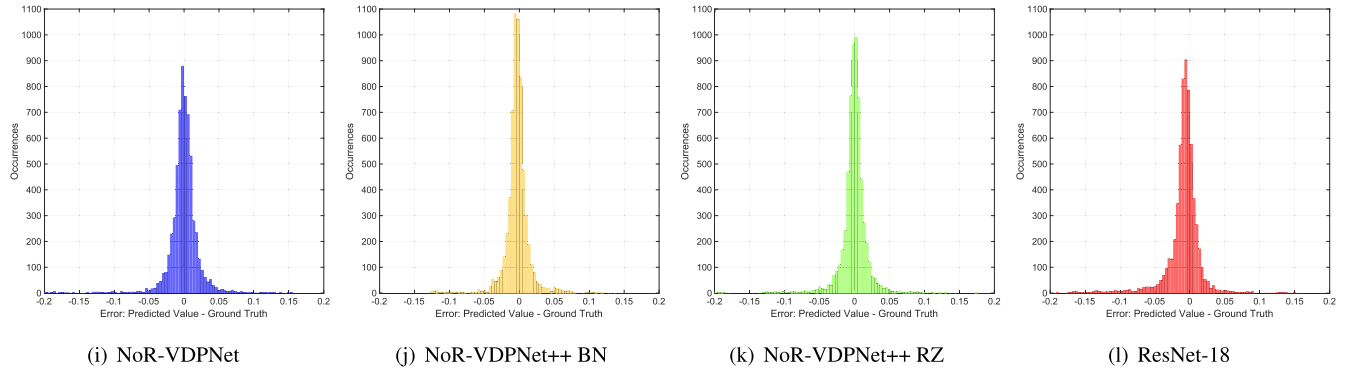
SDR-D



TMO



HDR-C



ITMO

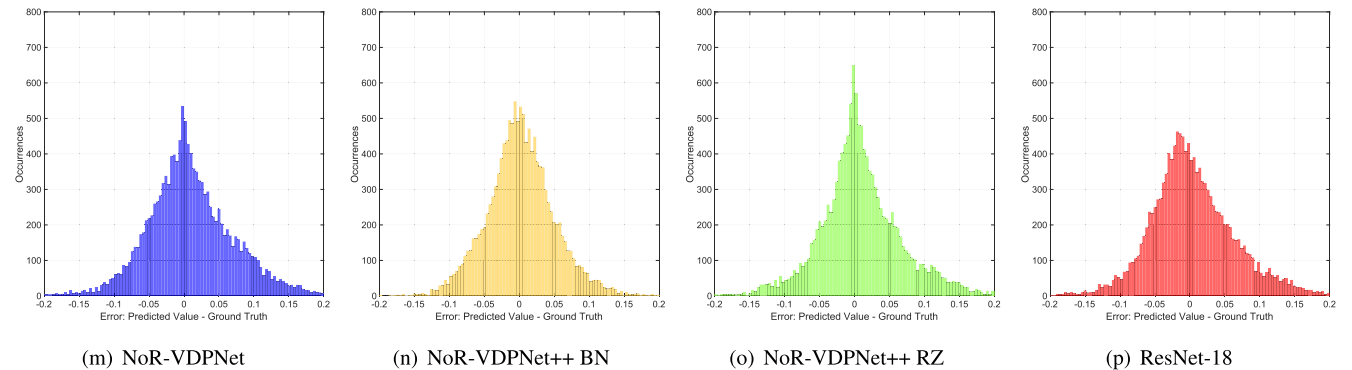


FIGURE 4. Histograms plots of the error between the predicted Q value and its ground truth for images belonging to the test dataset of each scenario. From the top to the bottom: HDR-C, ITMO, SDR-D, and TMOI .

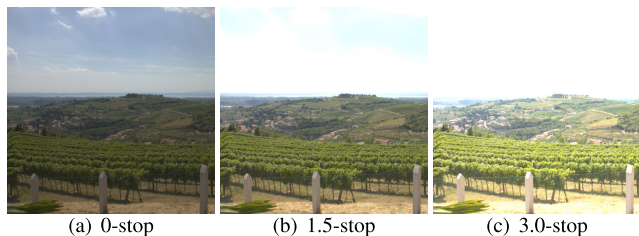


FIGURE 5. An example of the exposure augmentation of the input images for generating expanded HDR images for the ITMO dataset: (a) The well-exposed reference HDR image using Gallo et al.'s histogram method [17]. (b) The same image with the exposure set to +1.5-stop. (c) The same image with the exposure set to +3.0-stop.

TABLE 1. The size (number of images) of each dataset employed in this paper.

Dataset	Training	Validation	Test	Total
SDR-D	80,244	10,025	10,044	100,313
TMO	106,290	13,320	13,320	132,930
HDR-C	49,602	6,216	6,216	62,034
ITMO	106,290	13,320	13,320	132,930

well-exposed input image (with only clipped highlights); see Figure 5.

Given that the same HDR image is tone/inverse tone mapped with different TMOs/ITMOs, this is actually equivalent to performing data augmentation. Therefore, for each image, all different tone/inverse tone mapped images are placed either in the training set, in the evaluation set, or in the test set.

Note that for HDR-C and SDR-D, we extended Scenario 1 and Scenario 2 from Artusi et al.'s work [2] by increasing the number of samples by 3.8 times and 7 times, respectively. We achieved that using images from [12] for SDR-D, and the new images from I_{HDR} for HDR-C.

To further increase the size of the dataset, we performed further data augmentation by applying $90^\circ/180^\circ/270^\circ$ rotations and horizontal/vertical image flips. Note that HDR-VDP 2.2 requires physical values in order to obtain meaningful results, so images were converted from relative values to display-referred values. For the SDR-D dataset, the reference display had the characteristics of nowadays standard 8-bit display; i.e., the display peak brightness and black level were, respectively, set to 250 cd/m^2 and 0.5 cd/m^2 . Regarding ITMO and HDR-C datasets, the reference HDR display was the DisplayHDR1400 standard¹ with a peak luminance of $1,400 \text{ cd/m}^2$ and a black level of 0.02 cd/m^2 . The TMO dataset had no reference display because TMQI [37] works on normalized values for both the HDR and tone-mapped images.

Our training machine was an NVIDIA DGX Server 5.2.0 machine equipped with four AMD Epyc 7742 (64-core) CPUs with 1 TB of memory, and we used a NVIDIA A100 GPU with 40 GB of memory (CUDA 11.3). For implementing NoR-VDPNet++,² we modified the publicly available

¹<https://displayhdr.org/>

²<https://github.com/banterle/NoR-VDPNetpp>

TABLE 2. Performance evaluation in terms of MSE (lower is better). **Boldface** indicates the best method overall for each scenario. **Superscripts ‡** denotes the method (if any) whose MSE score is not statistically significantly different from the best one in terms of a two-tailed t-test in the differences in performance: symbol ‡ indicates $0.01 < p$ -value; i.e., the methods behave similarly with very high confidence.

Method\Dataset	SDR-D	TMO	HDR-C	ITMO
NoR	2.177E-04	1.598E-04	7.887E-04	4.097E-03
NoR++BN	2.289E-04	5.321E-04	6.364E-04‡	2.375E-03
NoR++RZ	1.822E-04	1.270E-04	4.802E-04	3.748E-03
ResNet-18	3.075E-04	5.312E-03	8.150E-04	3.739E-03

TABLE 3. Training time per epoch in seconds.

Method\Dataset	SDR-D	TMO	HDR-C	ITMO
NoR	343	426	215	450
NoR++BN	542	657	314	675
NoR++RZ	405	552	244	566
ResNet-18	2297	2693	1378	2787

code of NoR-VDPNet³ that uses PyTorch 1.3.1 deep-learning framework. For ResNet-18, we employed the PyTorch implementation using its original weights and fine-tuning weights using SDR-D, TMO, HDR-C, and ITMO training sets. During training, we employed Adam as the optimizer with default parameters and learning rate initialized to 10^{-5} ; we halved the learning rate whenever a plateau was reached. We trained all our networks for 100 epochs and certified that the optimization search converged in all cases. Typically, convergence is reached after 60 or 70 epochs.

V. RESULTS

In order to assess the quality of the predictions that our new NR model yields, we compared the Mean Squared Error (MSE) of the predictions against the FR target quality values (as produced by HDR-VDP 2.2) for the test datasets of SDR-D, HDR-C, TMO, and ITMO.

Table 2 reports performance comparisons in terms of MSE between the original NoR-VDPNet [9], ResNet-18, and the new variants NoR-VDPNet++ when equipped with Batch Normalization (BN) or with ReZero (RZ), for SDR-D, HDR-C, TMO, and ITMO. Statistical significance of the averaged scores is tested according to a two-tailed t-test at different confidence levels ($\alpha = 0.01$ and $\alpha = 0.001$).

These results reveal some interesting facts. First of all, there is a clear advantage (i.e., a statistically significant improvement), in terms of error score, when equipping the network with sophisticated normalization layers, when compared to the classical NoR variant; see Table 2. Another interesting aspect that jumps to the eye, is that NoR++BN and NoR++RZ both perform substantially better, in a statistically significant sense, than ResNet-18 in terms of error score. Interestingly enough, this improvement does not come

³<https://github.com/banterle/NoR-VDPNet>

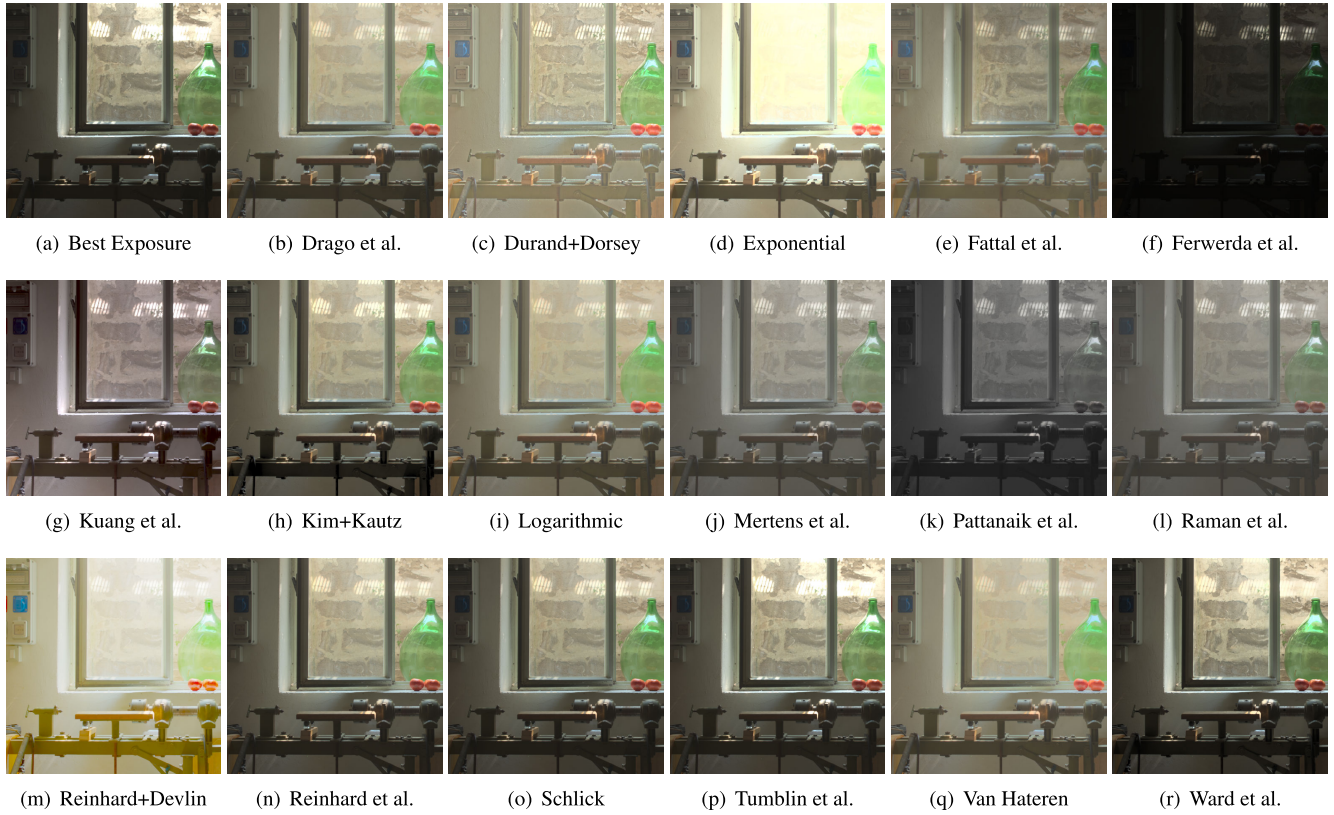


FIGURE 6. An example of the 18 TMOs from the HDR Toolbox [8] applied to the So-de-soto HDR image.

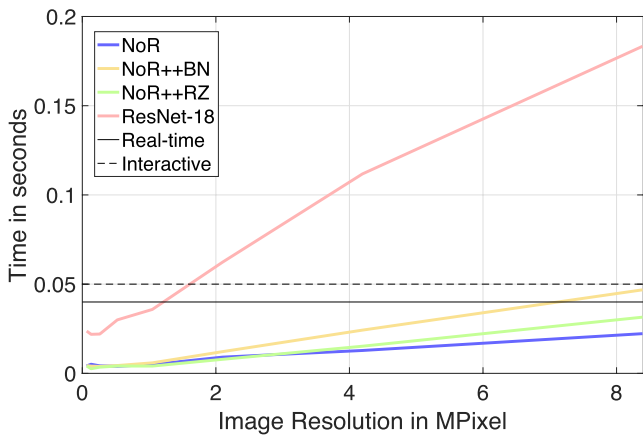


FIGURE 7. Timings at inference time for all different tested architectures.

at an extra cost. Indeed, ResNet-18 requires 58 hours for training on the SDR-D dataset, while NoR++RZ requires only 11 hours on the same dataset.

Figure 4 shows the error distributions for the testing datasets. Note that, amongst all methods, NoR-VDPNet++RZ displays the narrowest histogram centered around 0 for the majority of scenarios.

For a clearer picture, Figure 3 shows the scatter plots between the predicted value \hat{Q} and its ground truth Q by also

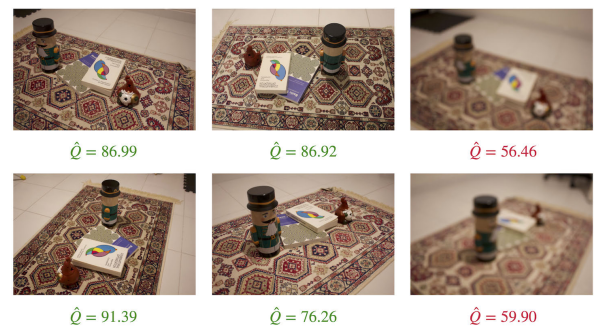


FIGURE 8. An example in which NoR-VDPNet++ is used to choose high-quality images from an image collection. NoR-VDPNet++ predicts a high Q -score (i.e., $Q > 70$) for sharp images, and a low one (i.e., $Q < 60$) for blurred images.

reporting the Pearson correlation coefficient ρ . The scatter plots exhibit a linear relationship between the inputs and the predicted values that tend to lie close to the main diagonal. From these plots, we can notice that ITMO is the most difficult case overall. This is due to the fact that an inverse tone mapping operator (both classic methods and especially deep-learning-based ones) applies many different processing operations at the same time on the same image.

Training times are reported in Table 3. It is worth noting that NoR++RZ displays comparable training times to NoR, while yielding better performance in terms of quality; see

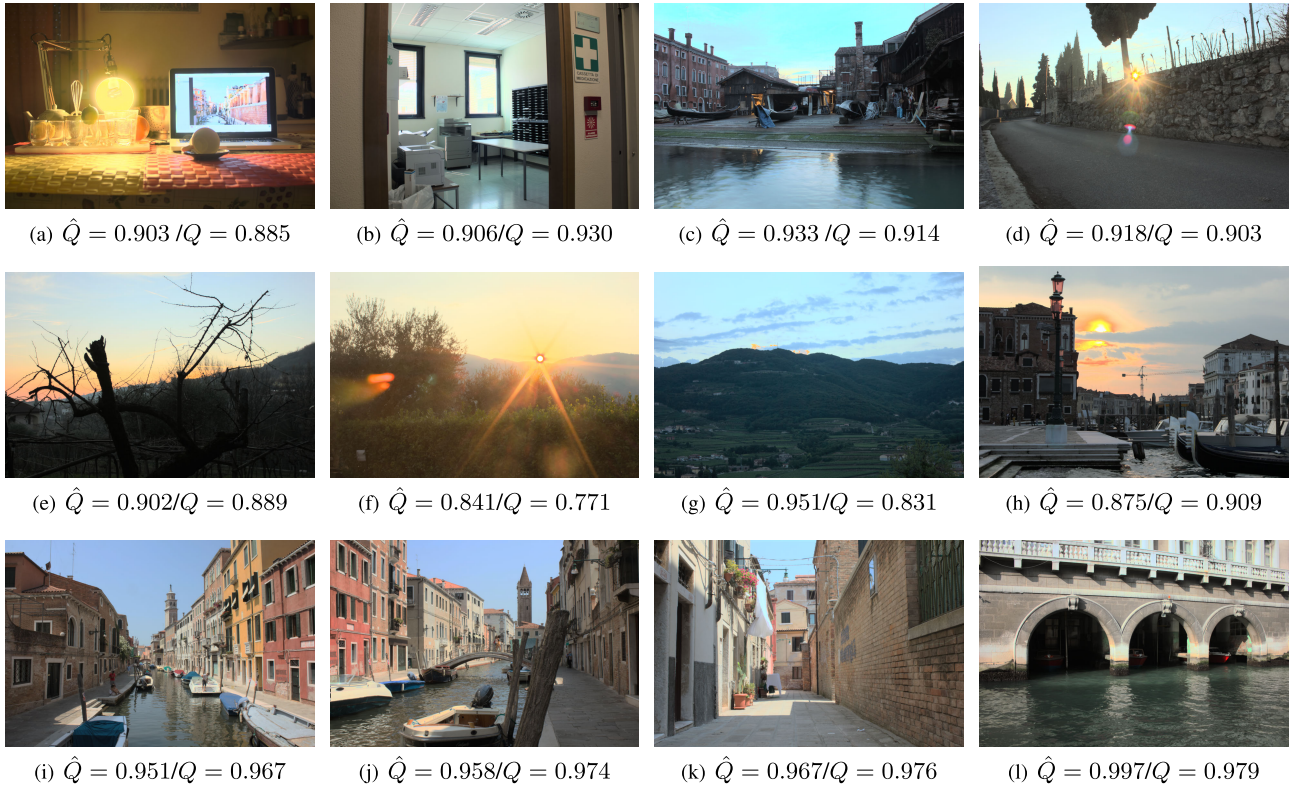


FIGURE 9. The results of our tone mapping application that maximises the TMQI. We report the maximized predicted TMQI, \hat{Q} , and the real value, Q .

Table 2. In terms of computational efficiency at inference time, the new architectures maintain real-time performance; i.e., both variants BN and RZ can issue predictions for 4-MPixel images in less than 24ms; see Figure 7. In our implementation, RZ is 44% faster than BN at high resolutions (i.e., >2-MPixel) because the implementation of Equation 1 is computationally more expensive than that of Equation 2.

VI. APPLICATIONS

NoR-VDPNet++ is a real-time metric, meaning that it can be employed in several optimization-based applications in which the parameters need to be optimized for a specific quality metric. A straightforward application of our work is the selection of high-quality images from an image collection; see Figure 8. This might be particularly useful for sorting vacation photographs or eliminating low-quality images in computer vision applications such as Structure-from-Motion [10] (e.g., removing blurred frames in a 3D reconstruction). Another interesting application is to use our metric trained on TMQI to optimize tone mapping operators parameters. To prove this possibility, we made an application that try to optimize the parameter of this sigmoid TMO:

$$L_d = \frac{L_w \alpha}{L_w \alpha + \mu} \quad C_d = \left(\frac{C_w}{L_w} \right)^\gamma L_d, \quad (3)$$

where C_w and C_d are, respectively, a HDR and a SDR color channel; L_w and L_d are, respectively, the HDR and SDR

luminance; α and μ are tone-curve parameters, and γ is a color saturation parameter. Figure 9 shows tone mapped images using this optimization process, displaying the TMQI predicted by the network and its corresponding real value.

The proposed tone mapping optimization can be also employed for selecting TMO parameters for JPEG-Xt [28] compression using HDR-C results. In a similar way, our metric trained on ITMO can be employed to optimize inverse tone mapping operators (be them relying on neural or non-neural implementations).

VII. DISCUSSION AND CONCLUSION

We have shown that CNN architectures can successfully distill the knowledge of existing reference metrics like HDR-VDP 2.2 [32] and TMQI [37]. In this work, we have presented NoR-VDPNet++, an improved variant of NoR-VDPNet [9]. This variant achieves more reliable results in general, and also in a newly introduced scenario, i.e., the evaluation of inverse tone mapped images. We also showed NoR-VDPNet++ outperforms other comparatively more complex networks like ResNet-18, while at the same time requiring less time to train, and being faster at inference time.

NoR-VDPNet++ maintains real-time performance, allowing it to be employed in any real-time constrained applications such as optimization processes for parameter selections like tone mapping, image selection from collections

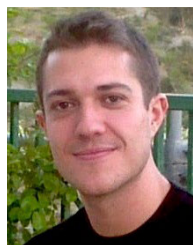
of photographs, or Structure-from-Motions tasks, to name a few.

Recent efforts have been paid in order to better understand how intermediate feature maps of pre-trained CNNs can be used to predict image distortion similarly to how humans do. For example, Zhang et al. [38] show a systematic study on how to evaluate feature maps across different CNN architectures, obtaining important improvements with respect to classical objective metrics. Tariq et al. [34], have shown the existing correlation between the capabilities of pre-trained CNN features in optimizing the perceptual quality, with their accuracy in capturing basic human visual perception characteristics. This altogether suggests that more efforts have to be devoted to better understanding the potential benefits that using feature maps from pre-trained CNNs as an objective metric can bring to bear in image/video evaluation. In future work, we plan to carry out a systematic study in this direction, analyzing ways for employing these feature maps in NoR-VDPNet++ in an effective manner.

REFERENCES

- [1] A. O. Akyüz, R. Fleming, B. E. Riecke, E. Reinhard, and H. H. Bühlhoff, "Do HDR displays support LDR content? A psychophysical evaluation," *ACM Trans. Graph.*, vol. 26, no. 3, p. 38, Jul. 2007.
- [2] A. Artusi, F. Banterle, F. Carra, and A. Moreo, "Efficient evaluation of image quality via deep-learning approximation of perceptual metrics," *IEEE Trans. Image Process.*, vol. 29, pp. 1843–1855, 2020.
- [3] A. Artusi, R. K. Mantiuk, T. Richter, P. Korshunov, P. Hanhart, T. Ebrahimi, and M. Agostinelli, "JPEG XT: A compression standard for HDR and WCG images [standards in a nutshell]," *IEEE Signal Process. Mag.*, vol. 33, no. 2, pp. 118–124, Mar. 2016.
- [4] A. Artusi, R. K. Mantiuk, T. Richter, P. Hanhart, P. Korshunov, M. Agostinelli, A. Ten, and T. Ebrahimi, "Overview and evaluation of the JPEG XT HDR image compression standard," *J. Real-Time Image Process.*, vol. 16, no. 2, pp. 413–428, Apr. 2019.
- [5] T. O. Aydın, R. Mantiuk, K. Myszkowski, and H.-P. Seidel, "Dynamic range independent image quality assessment," *ACM Trans. Graph.*, vol. 27, no. 3, pp. 1–10, Aug. 2008.
- [6] T. Bachlechner, B. P. Majumder, H. Henry Mao, G. W. Cottrell, and J. McAuley, "ReZero is all you need: Fast convergence at large depth," 2020, *arXiv:2003.04887*.
- [7] A. Banitalebi-Dehkordi, M. Azimi, M. T. Pourazad, and P. Nasiopoulos, "Compression of high dynamic range video using the HEVC and H.264/AVC standards," in *Proc. 10th Int. Conf. Heterogeneous Netw. Quality, Rel., Secur. Robustness*, Rhodes, Greece, Aug. 2014, pp. 8–12.
- [8] F. Banterle, A. Artusi, K. Debattista, and A. Chalmers, *Advanced High Dynamic Range Imaging: Theory and Practice*, 2nd ed. Natick, MA, USA: CRC Press, Jul. 2017.
- [9] F. Banterle, A. Artusi, A. Moreo, and F. Carrara, "NoR-VDPNet: A no-reference high dynamic range quality metric trained on HDR-VDP 2," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2020, pp. 126–130.
- [10] F. Banterle, R. Gong, M. Corsini, F. Ganovelli, L. V. Gool, and P. Cignoni, "A deep learning method for frame selection in videos for structure from motion pipelines," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Anchorage, AK, USA, Sep. 2021, pp. 3667–3671.
- [11] 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 Trans. Image Process.*, vol. 27, no. 1, pp. 206–219, Jan. 2017.
- [12] M. Corsini, F. Banterle, F. Ponchio, and P. Cignoni, "Image sets compression via patch redundancy," in *Proc. 8th Eur. Workshop Vis. Inf. Process. (EUVIP)*, Rome, Italy, Oct. 2019, pp. 10–15.
- [13] G. Eilertsen, J. Kronander, G. Denes, R. K. Mantiuk, and J. Unger, "HDR image reconstruction from a single exposure using deep CNNs," *ACM Trans. Graph.*, vol. 36, no. 6, pp. 1–15, Nov. 2017.
- [14] D. Mark Fairchild, "The HDR photographic survey," in *Proc. 15th Color Imag. Conf. (CIC)*, Albuquerque, NM, USA: The Society for Imaging Science and Technology, Nov. 2007, pp. 233–238.
- [15] J. Froehlich, S. Grandinetti, B. Eberhardt, S. Walter, A. Schilling, and H. Brendel, "Creating cinematic wide gamut HDR-video for the evaluation of tone mapping operators and HDR-displays," *Proc. SPIE*, vol. 9023, pp. 1–10, Mar. 2014.
- [16] F. Brian and L. Shi, "The effect of exposure on MaxRGB color constancy," *Proc. SPIE*, vol. 7527, Jan. 2010, Art. no. 75270Y.
- [17] O. Gallo, M. Tico, R. Manduchi, N. Gelfand, and K. Pulli, "Metering for exposure stacks," *Comput. Graph. Forum*, vol. 31, pp. 479–488, May 2012.
- [18] M.-A. Gardner, K. Sunkavalli, E. Yumer, X. Shen, E. Gambaretto, C. Gagné, and J.-F. Lalonde, "Learning to predict indoor illumination from a single image," *ACM Trans. Graph.*, vol. 36, no. 6, pp. 1–14, Dec. 2017.
- [19] Z. Gu, L. Zhang, and H. Li, "Learning a blind image quality index based on visual saliency guided sampling and Gabor filtering," in *Proc. ICIP*, Sep. 2013, pp. 186–190.
- [20] Y. Huo, F. Yang, L. Dong, and V. Brost, "Physiological inverse tone mapping based on retina response," *Vis. Comput.*, vol. 30, no. 5, pp. 507–517, May 2014.
- [21] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.
- [22] L. Kang, P. Ye, Y. Li, and D. Doermann, "Convolutional neural networks for no-reference image quality assessment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Columbus, OH, USA: IEEE Computer Society, Jun. 2014, pp. 1733–1740.
- [23] J. Kim and S. Lee, "Fully deep blind image quality predictor," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 1, pp. 206–220, Feb. 2017.
- [24] N. K. Kottayil, G. Valenzise, F. Dufaux, and I. Cheng, "Blind quality estimation by disentangling perceptual and noisy features in high dynamic range images," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1512–1525, Mar. 2018.
- [25] R. P. Kovaleski and M. M. Oliveira, "High-quality reverse tone mapping for a wide range of exposures," in *Proc. 27th SIBGRAPI Conf. Graph., Patterns Images*, Rio de Janeiro, Brazil: IEEE Computer Society, Aug. 2014, pp. 49–56.
- [26] D. Kundu, D. Ghadiyaram, A. C. Bovik, and B. L. Evans, "Large-scale crowdsourced study for tone-mapped HDR pictures," *IEEE Trans. Image Process.*, vol. 26, no. 10, pp. 4725–4740, Oct. 2017.
- [27] K. Ma, H. Yeganeh, K. Zeng, and Z. Wang, "High dynamic range image compression by optimizing tone mapped image quality index," *IEEE Trans. Image Process.*, vol. 24, no. 10, pp. 3086–3097, Oct. 2015.
- [28] R. K. Mantiuk, T. Richter, and A. Artusi, "Fine-tuning JPEG-XT compression performance using large-scale objective quality testing," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 2152–2156.
- [29] B. Masia, A. Serrano, and D. Gutierrez, "Dynamic range expansion based on image statistics," *Multimedia Tools Appl.*, vol. 76, no. 1, pp. 631–648, Jan. 2017.
- [30] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.
- [31] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2012.
- [32] M. Narwaria, R. K. Mantiuk, M. P. Da Silva, and P. Le Callet, "HDR-VDP-2.2: A calibrated method for objective quality prediction of high-dynamic range and standard images," *Proc. SPIE*, vol. 24, no. 1, Jan. 2015, Art. no. 010501.
- [33] M. S. Santos, T. I. Ren, and N. K. Kalantari, "Single image HDR reconstruction using a CNN with masked features and perceptual loss," *ACM Trans. Graph.*, vol. 39, no. 4, pp. 80-1–80-10, Aug. 2020.
- [34] T. Tariq, O. Tursun, M. Kim, and P. Diddy, "Why are deep representations good perceptual quality features?" in *Proc. ECCV*, 2020, pp. 445–461.
- [35] Z. Wang, A. C. Bovik, and B. L. Evan, "Blind measurement of blocking artifacts in images," in *Proc. IEEE Int. Conf. Image Process.*, vol. 3, Sep. 2000, pp. 981–984.
- [36] F. Xiao, J. M. DiCarlo, P. B. Catrysse, and B. A. Wandell, "High dynamic range imaging of natural scenes," in *Proc. 10th Color Imag. Conf., Color Sci. Eng. Syst., Technol., Appl. (CIC)*, Scottsdale, AZ, USA: The Society for Imaging Science and Technology, Nov. 2002, pp. 337–342.
- [37] H. Yeganeh and Z. Wang, "Objective quality assessment of tone-mapped images," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 657–667, Feb. 2013.

- [38] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 586–595.
- [39] H. Zhu, L. Li, J. Wu, W. Dong, and G. Shi, "MetalQA: Deep meta-learning for no-reference image quality assessment," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2020, pp. 14143–14152.
- [40] X. Zhu and P. Milanfar, "A no-reference sharpness metric sensitive to blur and noise," in *Proc. Int. Workshop Quality Multimedia Exp.*, Jul. 2009, pp. 64–69.



ALEJANDRO MOREO received the Ph.D. degree in computer sciences and information technologies from the University of Granada, in 2013. He is a Researcher with the Artificial Intelligence for Multimedia and Humanities Laboratory, ISTI-CNR, Italy, and a member of the Text Learning Group. His research interests include deep learning and representation learning, with a particular focus on quantification and transfer learning for text classification.



FRANCESCO BANTERLE received the Ph.D. degree in engineering from Warwick University, in 2009. He developed Inverse Tone Mapping that bridges the gap between SDR and HDR imaging. He is a full-time Researcher with the Visual Computing Laboratory, ISTI-CNR, Italy. He is the first author of the book *Advanced High Dynamic Range Imaging* and a reference book for HDR imaging research and the coauthor of the book *Image Content Retargeting*. His main research interests

include HDR imaging, computer graphics (image-based lighting), computer vision, and deep learning.



FABIO CARRARA received the master's and Ph.D. degrees in computer engineering from the University of Pisa, Italy, in 2015 and 2019, respectively. He is a Researcher with the Artificial Intelligence for Multimedia and Humanities Laboratory, ISTI-CNR, Italy. His research interests include deep learning for multimedia data with a focus on visual perception, image classification, and content-based and cross-media image retrieval and analysis.



ALESSANDRO ARTUSI received the Ph.D. degree in computer science from the Vienna University of Technology, in 2004. He is currently a Research Associate Professor and the Managing Director of the DeepCamera Laboratory, CYENS, Cyprus. Recently, he joined as a Founding Member of the Moving Picture, Audio and Data Coding by Artificial Intelligence (MPAI), a not-for-profit standards organization established in Geneva. His research interests include visual

perception, image/video processing, HDR technology, objective/subjective imaging/video evaluation, deep learning, computer vision, and color science. He was a recipient of the prestigious BSI Award for his work on the JPEG-Xt standard. He was appointed as the Head of Delegation for the Cypriot National Body for the ISO/IEC/JCT 1 SC29 Committee. In the past, he has been a BSI Member of the IST/37 Committee and the U.K. Representative for the JPEG and MPEG Standardization Committees.



PAOLO CIGNONI is a Research Director with CNR-ISTI, where he leads the Visual Computing Laboratory. He has published more than 180 papers in international refereed journals/conferences and has served on the program committee for all the most important conferences of computer graphics. His research interests include computer graphics fields including geometry processing and the use of machine-learning technologies for 3-D, applied to 3-D scanning data processing, digital fabrication, scientific visualization, and digital heritage. His laboratory has provided the community with many successful and widely distributed advanced open-source software tools, such as MeshLab and TagLab, that have helped the research and professional activities of millions of people around the world. He has been a member of ACM SIGGRAPH Academy, since 2022. He was awarded Outstanding Technical Contributions by the Eurographics Association, in 2021.

...

Open Access funding provided by 'Consiglio Nazionale delle Ricerche-CARI-CARE' within the CRUI CARE Agreement