IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

*Digital Object Identifier 10.1109/ACCESS.2017.DOI*

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

**Francesco Banterle [1], Alessandro Artusi[2], Alejandro Moreo[1], Fabio Carrara[1], and Paolo Cignoni[1]**
[1]ISTI-CNR, Italy
[2]DeepCamera, CYENS CoE, Cyprus

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

This work has been supported by funding from the European Union Horizon 2020 research and innovation programme under grant agreement No 739578 for Dr. Artusi and No 820434 (ENCORE) for Dr. Banterle. The funding for the work of Dr. Artusi is also complemented by the Government of the Republic of Cyprus through the Directorate General for European Programmes, Coordination and Development.

**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 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 MetaIQA 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., NoRVDPNet 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 Figure2. 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.

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

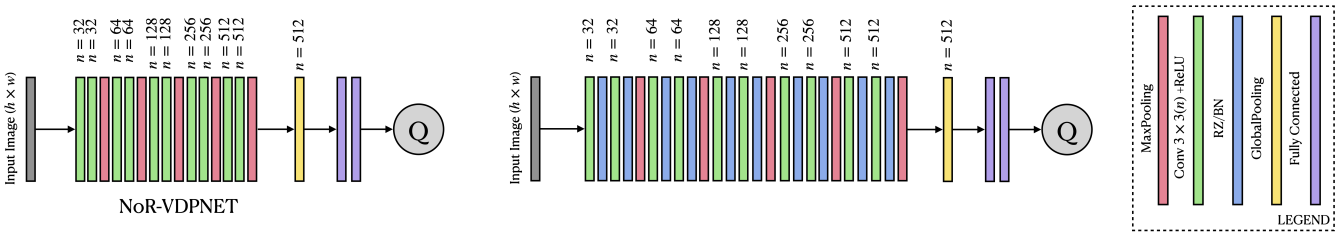



FIGURE 1: The network architecture of NoR-VDPNet (left) and NoRVDPNet++ (right): Batch Normalization or ReZero are added to each convolution layer.

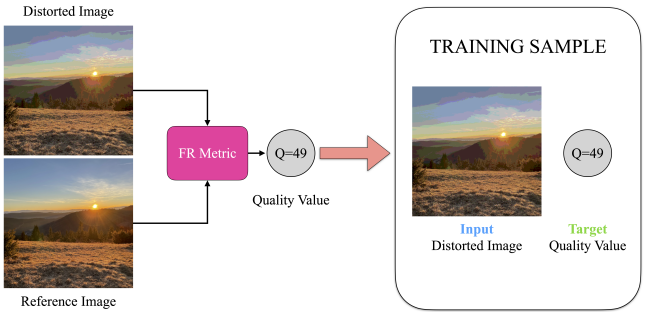



FIGURE 2: The process for creating a sample for our NR datasets. We use a FR metric for computing the quality value between the ground truth and the distorted images. Then, the sample is created by discarding the ground truth; the input for the network is the distorted image and the target output to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

examples. Equation 1 describes the computation for the $k$th dimension of a vector $\mathbf{x} = (x^{(1)}, \cdots, x^{(m)})$; $\mu_B^{(k)}$ and $\sigma_B^{(k)}$ stand for the sample mean and sample standard deviation, respectively, as computed on the batch of examples $B$.

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

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 ($\alpha_l$) 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:

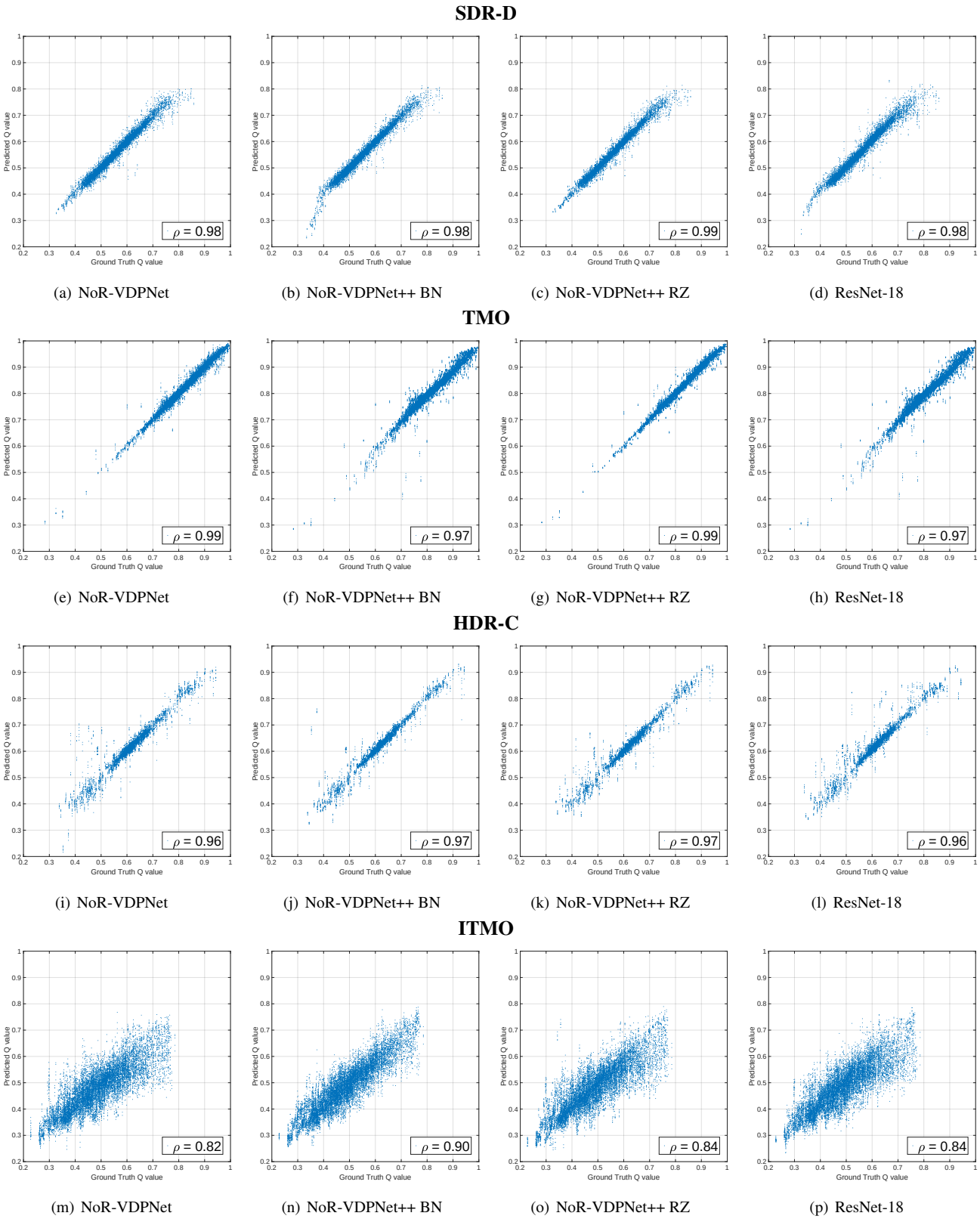
- **SDR-D**: Estimating HDR-VDP2.2 [32] quality value at different SDR distortions; e.g., blur, noise, quantization, etc. In this case, we converted 8-bit values into display referred values.
- **TMO**: Estimating TMQI [37] score under different tone mapping operators (TMOs).
- **HDR-C**: Estimating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



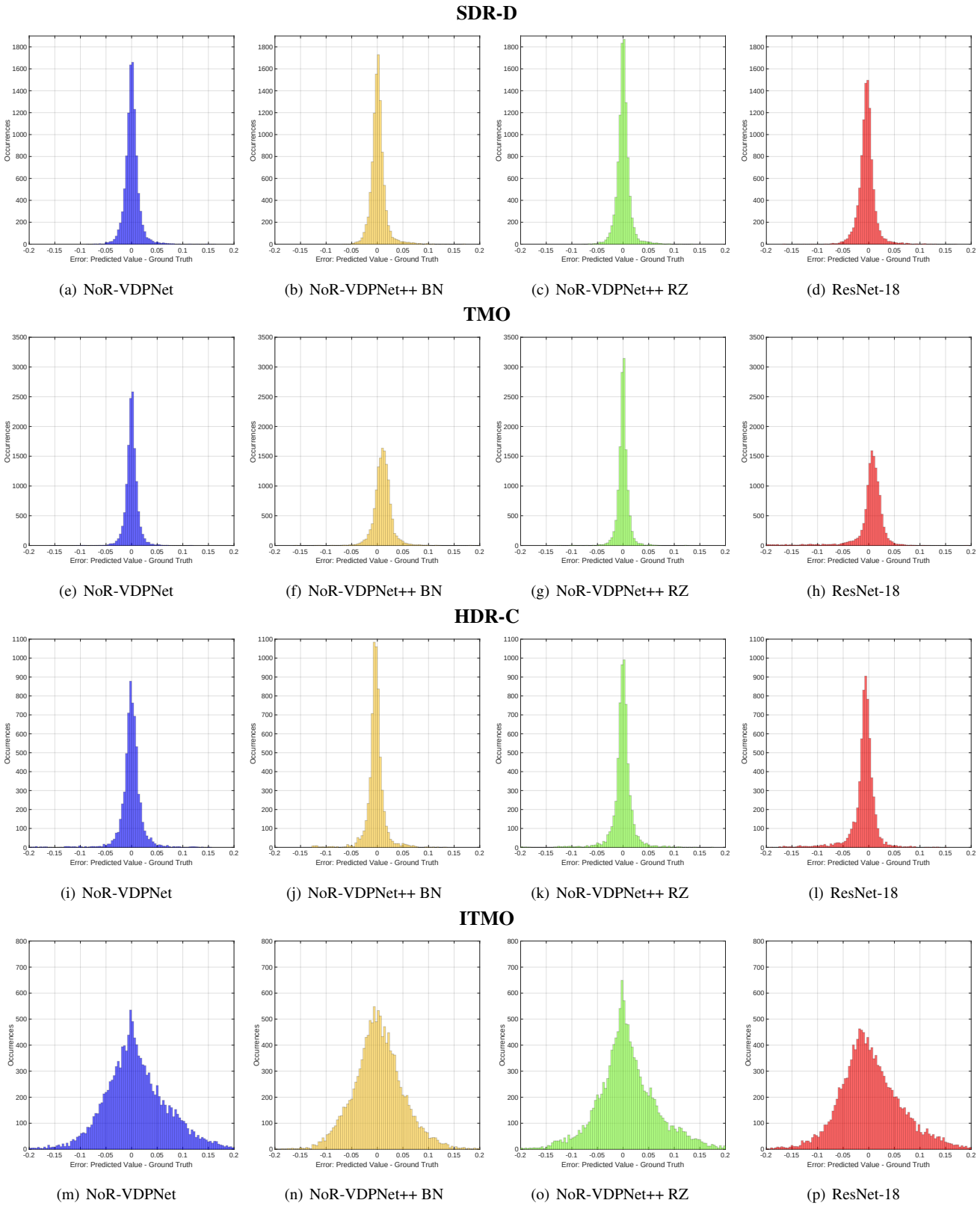

(a) NoR-VDPNet (b) NoR-VDPNet++ BN (c) NoR-VDPNet++ RZ (d) ResNet-18

(e) NoR-VDPNet (f) NoR-VDPNet++ BN (g) NoR-VDPNet++ RZ (h) ResNet-18

(i) NoR-VDPNet (j) NoR-VDPNet++ BN (k) NoR-VDPNet++ RZ (l) ResNet-18

(m) NoR-VDPNet (n) NoR-VDPNet++ BN (o) NoR-VDPNet++ RZ (p) ResNet-18

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




(a) NoR-VDPNet (b) NoR-VDPNet++ BN (c) NoR-VDPNet++ RZ (d) ResNet-18

(e) NoR-VDPNet (f) NoR-VDPNet++ BN (g) NoR-VDPNet++ RZ (h) ResNet-18

(i) NoR-VDPNet (j) NoR-VDPNet++ BN (k) NoR-VDPNet++ RZ (l) ResNet-18

(m) NoR-VDPNet (n) NoR-VDPNet++ BN (o) NoR-VDPNet++ RZ (p) ResNet-18

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 TMQI.

(a) 0-stop (b) 1.5-stop (c) 3.0-stop

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.

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 $I_{\text{HDR}}$ for HDR-C.

To further increase the size of the dataset, we performed further data augmentation by applying 90°/180°/270° 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 standard[1] with a peak luminance of $1,400$ 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.

| 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 |

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

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 code of NoR-VDPNet[3] 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.

[1]https://displayhdr.org/
[2]https://github.com/banterle/NoR-VDPNetpp
[3]https://github.com/banterle/NoR-VDPNet

| 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 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-tailored t-test in the differences in performance: symbol ‡ indicates $0.01 < p-\text{value}$; i.e., the methods behave similarly with very high confidence.

## V. RESULTS

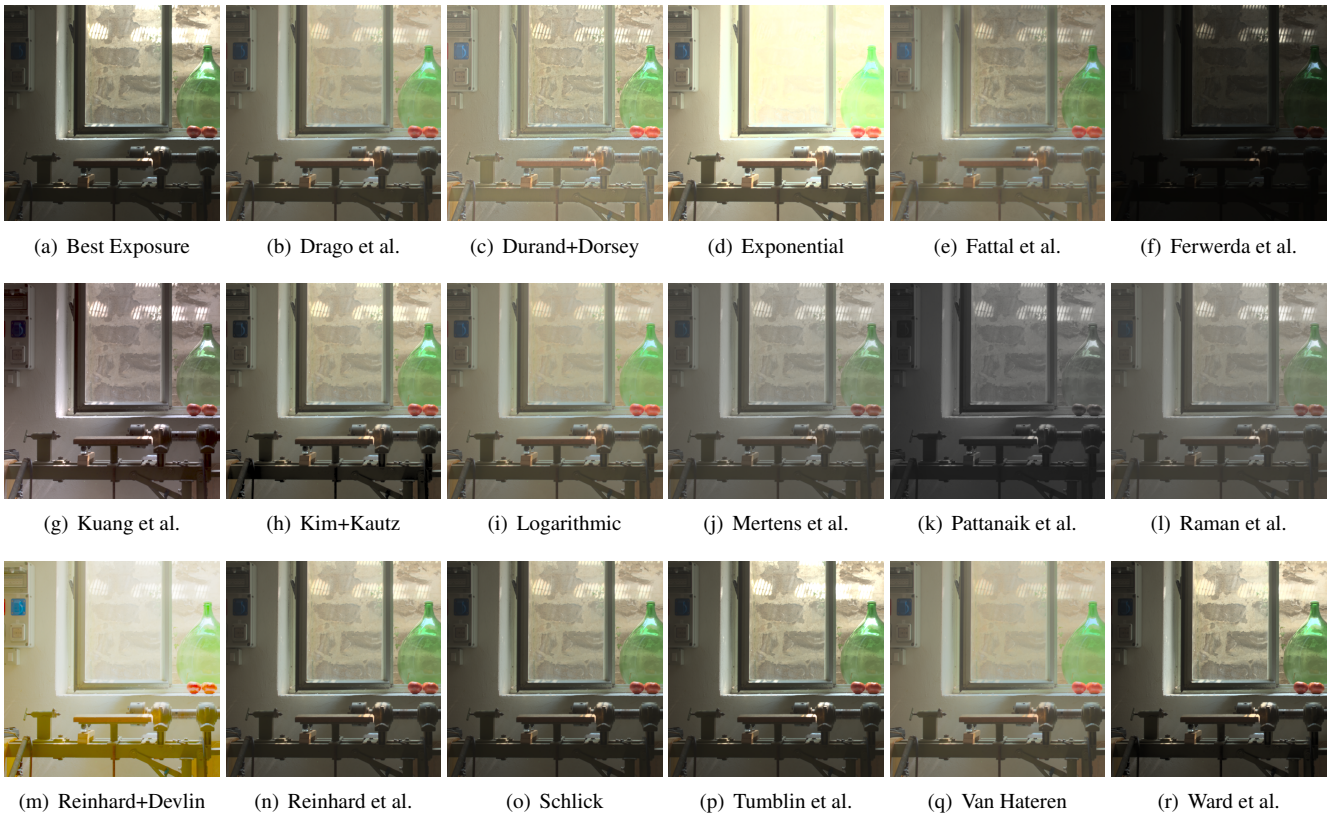
In order to assess the quality of the predictions that our new NR model yields, we compared the Mean Squared Error (MSE) of the predictions against the FR target quality values (as produced by HDR-VDP 2.2) for the test datasets of SDR-D, HDR-C, TMO, and ITMO.

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-tailored 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 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 reporting the Pearson correlation coefficient $\rho$. The scatter



(a) Best Exposure (b) Drago et al. (c) Durand+Dorsey (d) Exponential (e) Fattal et al. (f) Ferwerda et al.

(g) Kuang et al. (h) Kim+Kautz (i) Logarithmic (j) Mertens et al. (k) Pattanaik et al. (l) Raman et al.

(m) Reinhard+Devlin (n) Reinhard et al. (o) Schlick (p) Tumblin et al. (q) Van Hateren (r) Ward et al.

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

| 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 |

TABLE 3: Training time per epoch in seconds.

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 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.

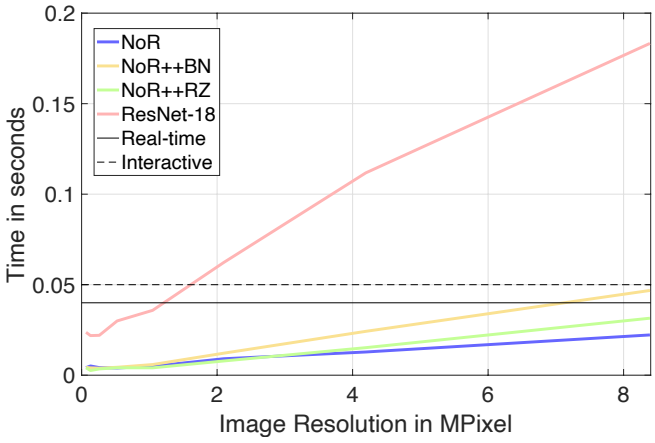



FIGURE 7: Timings at inference time for all different tested architectures.

## 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 col-


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.

lection; 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:

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

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; $\alpha$ and $\mu$ are tone-curve parameters, and $\gamma$ 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 CONCLUSIONS

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] Ahmet Oğuz Akyüz, Roland Fleming, Bernhard E. Riecke, Erik Reinhard, and Heinrich H. Bülthoff. Do HDR Displays Support LDR Content? A Psychophysical Evaluation. ACM Trans. Graph., 26(3):38–es, Jul 2007.

[2] Alessandro Artusi, Francesco Banterle, Alejandro Moreo, and Fabio Carrara. Efficient Evaluation of Image Quality via Deep-Learning Approximation of Perceptual Metrics. IEEE Trans. on Image Processing, 29:1843–1855, Oct 2019.

[3] Alessandro Artusi, Rafal K. Mantiuk, Thomas Richter, Pavel Korshunov, Philippe Hanhart, Touradj Ebrahimi, and Massimiliano Agostinelli. JPEG XT: A Compression Standard for HDR and WCG Images [Standards in a Nutshell]. IEEE Signal Processing Magazine, 33(2):118–124, 2016.

[4] Alessandro Artusi, Rafał K. Mantiuk, Thomas Richter, Philippe Hanhart, Pavel Kurshunov, Massimiliano Agostinelli, arkady Ten, and Touradj Ebrahimi. Overview and evaluation of the JPEG XT HDR image compression standard. In Journal of Real-Time Image Processing, pages 413–428, 2019.

[5] Tunç Ozan Aydın, Rafał Mantiuk, Karol Myszkowski, and Hans-Peter Seidel. Dynamic range independent image quality assessment. ACM Trans. Graph. (TOG), 27(3), Aug 2008.

[6] Thomas Bachlechner, Bodhisattwa Prasad Majumder, Huanru Henry Mao, Garrison W Cottrell, and Julian McAuley. Rezero is all you need: Fast convergence at large depth. arXiv preprint arXiv:2003.04887, 2020.

[7] Amin Banitalebi-Dehkordi, Mehran Azimi, Mahsa T. Pourazad, and Panos Nasiopoulos. Compression of high dynamic range video using the HEVC and H.264/AVC standards. In 10th International Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness, QShine 2014, Rhodes, Greece, August 18-20, 2014, pages 8–12. IEEE, 2014.

[8] Francesco Banterle, Alessandro Artusi, Kurt Debattista, and Alan Chalmers. Advanced High Dynamic Range Imaging: Theory and Practice (2nd Edition). AK Peters (CRC Press), Natick, MA, USA, Jul 2017.

[9] Francesco Banterle, Alessandro Artusi, Alejandro Moreo, and Fabio Carrara. NoR-VDPNet: A No-Reference High Dynamic Range Quality Metric Trained on HDR-VDP 2. In IEEE International Conference on Image Processing (ICIP). IEEE, Oct 2020.

[10] Francesco Banterle, Rui Gong, Massimiliano Corsini, Fabio Ganovelli, Luc Van Gool, and Paolo Cignoni. A Deep Learning Method for Frame Selection in Videos for Structure from Motion Pipelines. In 2021 IEEE International Conference on Image Processing, ICIP 2021, Anchorage, AK, USA, September 19-22, 2021, pages 3667–3671. IEEE, 2021.

[11] Sebastian Bosse, Dominique Maniry, Klaus-Robert Müller, Thomas Wiegand, and Wojciech Samek. Deep Neural Networks for No-Reference and Full-Reference Image Quality Assessment. IEEE Trans. Image Processing, 27(1):206–219, 2018.

[12] Massimiliano Corsini, Francesco Banterle, Federico Ponchio, and Paolo Cignoni. Image Sets Compression Via Patch Redundancy. In 8th European

(a) $\hat{Q} = 0.903$ /$Q = 0.885$ (b) $\hat{Q} = 0.906$/$Q = 0.930$ (c) $\hat{Q} = 0.933$ /$Q = 0.914$ (d) $\hat{Q} = 0.918$/$Q = 0.903$

(e) $\hat{Q} = 0.902$/$Q = 0.889$ (f) $\hat{Q} = 0.841$/$Q = 0.771$ (g) $\hat{Q} = 0.951$/$Q = 0.831$ (h) $\hat{Q} = 0.875$/$Q = 0.909$

(i) $\hat{Q} = 0.951$/$Q = 0.967$ (j) $\hat{Q} = 0.958$/$Q = 0.974$ (k) $\hat{Q} = 0.967$/$Q = 0.976$ (l) $\hat{Q} = 0.997$/$Q = 0.979$

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, $Q$.


Workshop on Visual Information Processing, EUVIP 2019, Rome, Italy, October 28-31, 2019, pages 10–15. IEEE, 2019.

[13] Gabriel Eilertsen, Joel Kronander, Gyorgy Denes, Rafał K. Mantiuk, and Jonas Unger. HDR Image Reconstruction from a Single Exposure Using Deep CNNs. ACM Trans. Graph., 36(6), Nov 2017.

[14] Mark D. Fairchild. The HDR Photographic Survey. In 15th Color and Imaging Conference, CIC 2007, Albuquerque, New Mexico, USA, November 5-9, 2007, pages 233–238. IS&T - The Society for Imaging Science and Technology, 2007.

[15] Jan Froehlich, Stefan Grandinetti, Bernd Eberhardt, Simon Walter, Andreas Schilling, and Harald Brendel. Creating Cinematic Wide Gamut HDR-Video for the Evaluation of Tone Mapping Operators and HDR-Displays. In Proc. SPIE, Digital Photography X, volume 9023, pages 1–10, 2014.

[16] Brian V. Funt and Lilong Shi. The effect of exposure on MaxRGB color constancy. In Bernice E. Rogowitz and Thrasyvoulos N. Pappas, editors, Human Vision and Electronic Imaging XV, part of the IS&T-SPIE Electronic Imaging Symposium, San Jose, CA, USA, January 18-21, 2010, Proceedings, volume 7527 of SPIE Proceedings, page 75270. SPIE, 2010.

[17] Orazio Gallo, Marius Tico, Roberto Manduchi, Natasha Gelfand, and Kari Pulli. Metering for Exposure Stacks. Comput. Graph. Forum, 31(2pt2):479–488, 2012.

[18] Marc-André Gardner, Kalyan Sunkavalli, Ersin Yumer, Xiaohui Shen, Emiliano Gambaretto, Christian Gagné, and Jean-François Lalonde. Learning to predict indoor illumination from a single image. ACM Trans. Graph., 36(6):176:1–176:14, 2017.

[19] Zhongyi Gu, Lin Zhang, and Hongyu Li. Learning a blind image quality index based on visual saliency guided sampling and Gabor filtering. In ICIP, pages 186–190. IEEE, 2013.

[20] Yongqing Huo, Fan Yang, Le Dong, and Vincent Brost. Physiological inverse tone mapping based on retina response. Vis. Comput., 30(5):507–517, 2014.

[21] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In International conference on machine learning, pages 448–456. PMLR, 2015.

[22] Le Kang, Peng Ye, Yi Li, and David S. Doermann. Convolutional Neural Networks for No-Reference Image Quality Assessment. In 2014 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014, Columbus, OH, USA, June 23-28, 2014, pages 1733–1740. IEEE Computer Society, 2014.

[23] Jongyoo Kim and Sanghoon Lee. Fully Deep Blind Image Quality Predictor. J. Sel. Topics Signal Processing, 11(1):206–220, 2017.

[24] Navaneeth Kamballur Kottayil, Giuseppe Valenzise, Frederic Dufaux, and Irene Cheng. Blind Quality Estimation by Disentangling Perceptual and Noisy Features in High Dynamic Range Images. IEEE Trans. on Image Processing, 27(3):1512–1525, Mar 2018.

[25] Rafael Pacheco Kovaleski and Manuel Menezes de Oliveira Neto. High-Quality Reverse Tone Mapping for a Wide Range of Exposures. In 27th SIBGRAPI Conference on Graphics, Patterns and Images, SIBGRAPI 2014, Rio de Janeiro, Brazil, August 27-30, 2014, pages 49–56. IEEE Computer Society, 2014.

[26] Debarati Kundu, Deepti Ghadiyaram, Alan C. Bovik, and Brian L. Evans. Large-Scale Crowdsourced Study for Tone-Mapped HDR Pictures. IEEE Trans. Image Processing, 26(10):4725–4740, 2017.

[27] Kede Ma, Hojatollah Yeganeh, Kai Zeng, and Zhou Wang. High Dynamic Range Image Compression by Optimizing Tone Mapped Image Quality Index. IEEE Trans. Image Process., 24(10):3086–3097, 2015.

[28] Rafał K. Mantiuk, Thomas Richter, and Alessandro Artusi. Fine-tuning JPEG-XT compression performance using large-scale objective quality testing. In 2016 IEEE International Conference on Image Processing (ICIP), pages 2152–2156, 2016.

[29] Belén Masiá, Ana Serrano, and Diego Gutierrez. Dynamic range expansion based on image statistics. Multim. Tools Appl., 76(1):631–648, 2017.

[30] Anish Mittal, Anush Krishna Moorthy, and Alan C. Bovik. No-Reference Image Quality Assessment in the Spatial Domain. IEEE Trans. on Image Processing, 21(12):4695–4708, Dec 2012.

[31] Anish Mittal, Rajiv Soundararajan, and Alan C. Bovik. Making a "Completely Blind" Image Quality Analyzer. IEEE Signal Process. Lett., 20(3):209–212, 2013.

[32] Manish Narwaria, Rafał K. Mantiuk, Mattheiu Perreira Da Silva, and

Patrick Le Callet. HDR-VDP-2.2: A calibrated method for objective quality prediction of high dynamic range and standard images. Journal of Electronic Imaging, 24(1), 2015.
[33] Marcel Santana Santos, Tsang Ing Ren, and Nima Khademi Kalantari. Single Image HDR Reconstruction Using a CNN with Masked Features and Perceptual Loss. ACM Trans. Graph., 39(4), Jul 2020.
[34] Taimoor Tariq, O. Tursun, Munchurl Kim, and P. Didyk. Why are deep representations good perceptual quality features? In ECCV, 2020.
[35] Zhou Wang, Alan C. Bovik, and B. L. Evan. Blind measurement of blocking artifacts in images. In Proceedings 2000 International Conference on Image Processing (Cat. No.00CH37101), volume 3, pages 981–984, Sep. 2000.
[36] Feng Xiao, Jeffrey M. DiCarlo, Peter B. Catrysse, and Brian A. Wandell. High Dynamic Range Imaging of Natural Scenes. In The Tenth Color Imaging Conference: Color Science and Engineering Systems, Technologies, Applications, CIC 2002, Scottsdale, Arizona, USA, November 12-15, 2002, pages 337–342. IS&T - The Society for Imaging Science and Technology, 2002.
[37] Hojatollah Yeganeh and Zhou Wang. Objective Quality Assessment of Tone-Mapped Images. IEEE Trans. Image Process., 22(2):657–667, 2013.
[38] Richard Zhang, Phillip Isola, Alexei A. Efros, E. Shechtman, and O. Wang. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 586–595, 2018.
[39] Hancheng Zhu, Leida Li, Jinjian Wu, Weisheng Dong, and Guangming Shi. MetaIQA: Deep Meta-Learning for No-Reference Image Quality Assessment. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 14143–14152, Jun 2020.
[40] Xiang Zhu and Peyman Milanfar. A no-reference sharpness metric sensitive to blur and noise. In 2009 International Workshop on Quality of Multimedia Experience, pages 64–69, Jul 2009.

FRANCESCO BANTERLE, PH.D. is a full-time researcher at the Visual Computing Laboratory at ISTI-CNR, Italy. He received a Ph.D. in Engineering from Warwick University in 2009 when he developed Inverse Tone Mapping that bridges the gap between SDR and HDR Imaging. He is the first author of the book "Advanced High Dynamic Range Imaging", a reference book for HDR imaging research, and co-author of the book "Image Content Retargeting". His main research interests are in the field of HDR imaging, Computer Graphics (image-based lighting), Computer Vision, and Deep Learning.

ALESSANDRO ARTUSI, PH.D. received a Ph.D. in Computer Science from the Vienna University of Technology in 2004. He is currently an Research Associate Professor and the Managing Director of the DeepCamera Lab at CYENS (Cyprus) who recently has joined, as a founding member, the Moving Picture, Audio and Data Coding by Artificial Intelligence (MPAI), a not-for-profit standards organization established in Geneva. He is the recipient, for his work on the JPEG-XT standard, of the prestigious BSI Award. He recently has been appointed as Head of Delegation for the Cypriot National Body for the ISO/IEC/JCT 1 SC29 committee, and in the past he has been BSI member of the IST/37 committee and UK representative for the JPEG and MPEG standardization committees. His research interests include visual perception, image/video processing, HDR technology, objective/subjective imaging/video evaluation, deep-learning, computer vision and color science.

ALEJANDRO MOREO received a PhD in Computer Sciences and Information Technologies from the University of Granada in 2013. He is a researcher at the Artificial Intelligence for Multimedia and Humanities Laboratory (AIMH, http://aimh.isti.cnr.it/) of the ISTI-CNR, Italy and a member of the Text Learning group. His research interests include deep learning and representation learning, with particular focus on quantification and transfer learning for text classification.

FABIO CARRARA is a researcher at the Artificial Intelligence for Multimedia and Humanities Laboratory (AIMH, http://aimh.isti.cnr.it/) of ISTI-CNR, Italy. He received a Master's Degree and a Ph.D. in Computer Engineering from the University of Pisa (Italy), respectively, in 2015 and 2019. His research interests include deep learning for multimedia data with a focus on visual perception, image classification, content-based and cross-media image retrieval and analysis.

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