NoR-VDPNet++: Efficient Training and Architecture for Deep No-Reference Image Quality Metrics

Francesco Banterle
ISTI-CNR
Pisa, Italy

Alessandro Artusi
DeepCamera, CYENS Ltd.
Nicosia, Cyprus

Alejandro Moreo
ISTI-CNR
Pisa, Italy

Fabio Carrara
ISTI-CNR
Pisa, Italy

ABSTRACT

Efficiency and efficacy are two desirable properties of the utmost importance for any evaluation metric having to do with Standard Dynamic Range (SDR) imaging or High Dynamic Range (HDR) imaging. However, these properties are hard to achieve simultaneously. On the one side, metrics like HDR-VDP2.2 are known to mimic the human visual system (HVS) very accurately, but its high computational cost prevents its widespread use in large evaluation campaigns. On the other side, computationally cheaper alternatives like PSNR or MSE fail to capture many of the crucial aspects of the HVS. In this work, we try to get the best of the two worlds: we present NoR-VDPNet++, an improved variant of a previous deep learning-based metric for distilling HDR-VDP2.2 into a convolutional neural network (CNN). In this work, we try to get the best of the two worlds: we present NoR-VDPNet++, an improved version of a deep learning-based metric for distilling HDR-VDP2.2 into a convolutional neural network (CNN).

CCS CONCEPTS

• **Computing methodologies → Supervised learning**; **Computational photography**.

KEYWORDS

Perceptual metrics, Neural networks, High Dynamic Range imaging

ACM Reference Format:
Francesco Banterle, Alessandro Artusi, Alejandro Moreo, and Fabio Carrara. 2021. NoR-VDPNet++: Efficient Training and Architecture for Deep No-Reference Image Quality Metrics. In *Special Interest Group on Computer Graphics and Interactive Techniques Conference Talks (SIGGRAPH '21 Talks), August 09-13, 2021.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3450623.3464636

1 INTRODUCTION AND RELATED WORK

In Computer Graphics, Computer Vision, and Imaging, the quality of synthetic images is commonly assessed either through user studies or through objective metrics. Although the former typically produce more reliable results, user studies tend to be cumbersome to run since a considerable amount of time is often required (e.g., from weeks to months in some cases) given that a large number of participants and images should be involved. Therefore, objective metrics are generally preferred in real applicative scenarios. Objective metrics are nonetheless fairly reliable too, and that is especially true for implementations rooting on the simulation of (complex) aspects of the human visual system (HVS). HDR-VDP2.2 [Narwaria et al. 2015] is a good example and has indeed become fairly popular in the field of high dynamic range (HDR) and standard dynamic range (SDR) imaging, and amongst the standardization community thanks to its high reliability. Unfortunately, its high computational cost precludes HDR-VDP2.2 from being employed in several quality assessment scenarios such as standardization, where large databases of high-resolution videos are customarily involved. Moreover, metrics like HDR-VDP2.2 are bounded by the availability of a so-called *reference* image (a ground truth image against to which the quality of a processed image is to be assessed), which might in some cases be difficult, or even impossible, to obtain. This altogether motivates the need for efficient, yet effective, objective metrics able to predict visual significant differences of test images and that do so (i) in real-time, and (ii) in the absence of a ground truth reference whenever this reference is missing. Recently, convolutional neural networks (CNN) based metrics have been proposed both for full-reference (DIQM) [Artusi et al. 2019] and no-reference scenarios (NoR-VDPNet) [Banterle et al. 2020]. These methods showed how objective metrics like HDR-VDP2.2[Narwaria et al. 2015] and DRIIM [Aydın et al. 2008] could be accurately distilled, yielding similar results in real-time. In this work, we present NoR-VDPNet++, an improved variant of [Banterle et al. 2020] that achieves higher accuracy using fewer parameters and preserving real-time performance.

2 DISTILLING IMAGE QUALITY METRICS AND ULTRA-AGED DISTILLATIONS

NoR-VDPNet [Banterle et al. 2020] accomplishes the conversion of HDR-VDP2.2 [Narwaria et al. 2015] into a no-reference model encoded as a CNN. This is achieved by training a CNN architecture (see top architecture in Figure 1) using a medium-large dataset (i.e., more than 70,000 examples with/without reference) of SDR and HDR images for different scenarios such as SDR distortions detection (blur, noise, quantization, etc.), JPEG-XT compression artifacts, etc. Each example pair consists of a *distorted image* and

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

SIGGRAPH '21 Talks, August 09-13, 2021, Virtual Event, USA
© 2021 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-8373-8/21/08.
https://doi.org/10.1145/3450623.3464636

the *quality value* that HDR-VDP2.2 calculates using its *reference*. Note that the key for distilling HDR-VDP2.2 into a no-reference metric comes down to omitting the reference during training.

We improve the training stability by applying Batch Normalization [Ioffe and Szegedy 2015] to the data flow, thus re-centering and re-scaling the data and effectively reducing the covariate shift between layers. This allowed us to train the network for a much larger number of epochs (up to 1000) than was done before (only 75 in [Banterle et al. 2020]) and this resulted in a reduction of the error prediction. The addition of Batch Normalization, however, slightly penalizes the time required for performing a *forward pass* through the network layers —in other words, the whole network becomes slower in a production deployment. We managed to compensate for this by removing the last two convolutional layers of the network without sacrificing performance. The resulting network achieves a lower prediction error than the original NoR-VDPNet but still preserves real-time performance. Figure 1 shows NoR-VDPNet before (upper) and after (bottom) these changes.

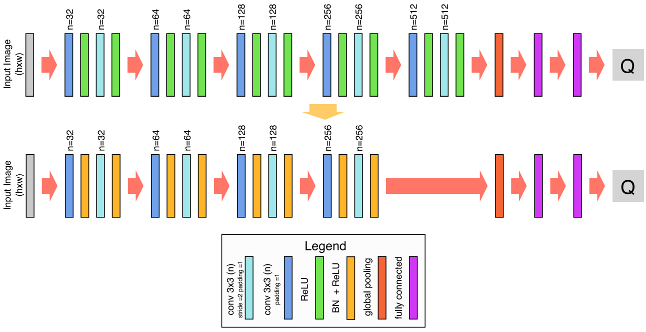

Figure 1: The new NoR-VDPNet++ architecture: At the top, the previous architecture [Banterle et al. 2020]. At the bottom, the new architecture where Batch Normalization is added before ReLU and the last convolutional layers have been removed.

We trained the new NoR-VDPNet++ for Scenario 1 (detection of distortions in SDR images) and Scenario 2 (detection of compression artifacts in HDR images), both from the DIQM dataset [Artusi et al. 2019]. We used Adam as the optimizer with default parameters and learning rate initialized to 1e-5 and halved whenever a plateau is reached. We trained this new model for 1000 epochs for each scenario.

To assess the quality of the predictions that our new (no-reference) model delivers, we compared the Mean Squared Error (MSE) of the predictions against the (fully-reference) target quality values (as produced by HDR-VDP2.2) for the test datasets of Scenario 1 and Scenario 2. The differences in performance between NoR-VDPNet++ and NoR-VDPNet, as measured in terms of MSE are statistically significant at a high confidence level ($\alpha = 0.005$) with p-value=8.17e-3 (μ_{new}=8.83e-4 and μ_{old}=1.10e-3, respectively) for Scenario 1, and p-value=3.18e-5 (μ_{new}=3.96e-5 and μ_{old}=6.48e-5, respectively) for Scenario 2. Regarding the time efficiency, the new architecture is 4% slower than the original one (due to the addition of Batch Normalization) as clocked on the very same machine[1], but still maintains real-time performance; e.g., NoR-VDPNet++ manages to evaluate 4MPixel images in no more than 32ms. Being a real-time metric allows NoR-VDPNet++ to be employed in several optimization-based applications for imaging; see Figure 2. Furthermore, the method can be used to assess the quality of renderings to determine how many path-samples to shoot after a few iterations.

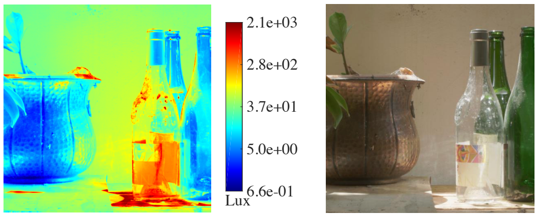

Figure 2: An application of NoR-VDPNet++, the parameters selection of the tone mapping operator for compressing HDR images using JPEG-XT. On the left side, a false-color HDR image. On the right side, the tone-mapped version with parameters selected by NoR-VDPNet++, which are optimal for JPEG-XT compression.

3 DISCUSSION AND CONCLUSIONS

We have presented NoR-VDPNet++, an improved variant of [Banterle et al. 2020] that achieves more reliable quality scores while keeping real-time performance. This allows NoR-VDPNet++ to be employed in any real-time constrained application such as optimization processes for parameter selections as those required for JPEG-XT compression, denoising, or inverse tone mapping, to name a few, that remained out of reach for expensive metrics like HDR-VDP2.2 up to now.

ACKNOWLEDGMENTS

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), No 813170 (EVOCATION) for Dr. Banterle respectively. The funding for the work of Dr. Artusi are also complemented by the Government of the Republic of Cyprus through the Directorate General for European Programmes, Coordination and Development.

REFERENCES

Alessandro Artusi, Francesco Banterle, Alejandro Moreo, and Fabio Carrara. 2019. Efficient Evaluation of Image Quality via Deep-Learning Approximation of Perceptual Metrics. *IEEE Transactions on Image Processing* 29 (oct 2019), 1843–1855. http://vcg.isti.cnr.it/Publications/2019/ABMC19

Tunç Ozan Aydın, Rafał Mantiuk, Karol Myszkowski, and Hans-Peter Seidel. 2008. Dynamic Range Independent Image Quality Assessment. *ACM Transactions on Graphics (TOG)* 27, 3, Article 69 (2008).

Francesco Banterle, Alessandro Artusi, Alejandro Moreo, and Fabio Carrara. 2020. NoR-VDPNet: A No-Reference High Dynamic Range Quality Metric Trained on HDR-VDP 2. In *IEEE International Conference on Image Processing (ICIP)*. IEEE. http://vcg.isti.cnr.it/Publications/2020/BAMC20

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

Manish Narwaria, Rafal K. Mantiuk, Mattheiu Perreira Da Silva, and Patrick Le Callet. 2015. 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).

[1] A Linux machine (Ubuntu 18.04) equipped with an Intel CPU Core i7-7800X (3.50 GHz) with 64 GB of memory and an NVIDIA GeForce GTX 1080 GPU with 8 GB of memory. For implementing NoR-VDPNet++ we modified the publicly available code of NoR-VDPNet that uses PyTorch 1.3.1 deep-learning framework.