DR2S: Deep Regression with Region Selection for Camera Quality Evaluation

Marcelin Tworski, Stéphane Lathuilière, Salim Belkarfa, Attilio Fiandrotti and Marco Cagnazzo

International Conference on Pattern Recognition 2020, 10-15 Janvier 2021

Abstract

In this work, we tackle the problem of estimating a camera capability to preserve fine texture details at a given lighting condition. Importantly, our texture preservation measurement should coincide with human perception. Consequently, we formulate our problem as a regression one and **we introduce a deep convolutional network to estimate texture quality score**. At training time, we use ground-truth quality scores provided by expert human annotators in order to obtain a subjective quality measure. In addition, we propose a **region selection** method to identify the image regions that are better suited at measuring perceptual quality. Finally, our experimental evaluation shows that our learning-based approach outperforms existing methods and that our region selection algorithm consistently improves the quality estimation.

Details preservation attribute

Several attributes are important to evaluate an image: target exposure and dynamic range, color (saturation and white balance) texture, noise and various artifacts that can affect the quality of the final image. In this work, **we aim at evaluating camera capabilities to preserve fine texture details**.

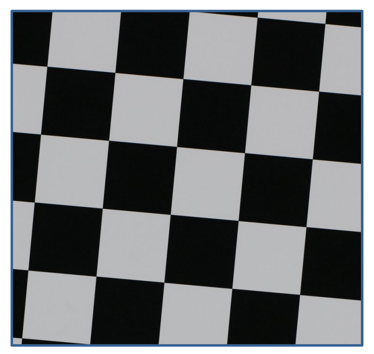
State of the art

A typical way to evaluate the quality of a set of cameras consists of comparing shots of the same visual content in a controlled environment.

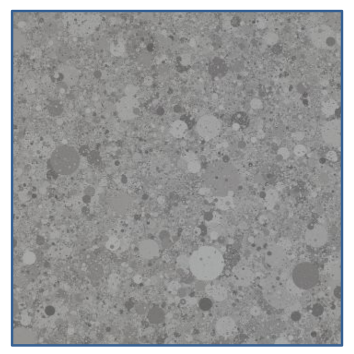
Relation between content X and the resulting image Y can be expressed with the function H for each frequency : $Y(f) = H(f)X(f)$

The modulus of H is called the Modulation Transfer Function (MTF) [1]Human sensitivity to different frequencies are determined with the Contrast Sensitivity Function (CSF) for a viewing condition setting, that allow us to derive a single quality metric :

$$\text{Acutance} = \frac{A}{A_r} \quad \text{with} \quad A = \int_0^\infty MTF(f) \cdot \text{CSF}(f) \cdot df \quad \text{and} \quad A_r = \int_0^\infty \text{CSF}(f) \cdot df$$



Slanted Edges



Dead-Leaves Pattern [2]

Limitations :

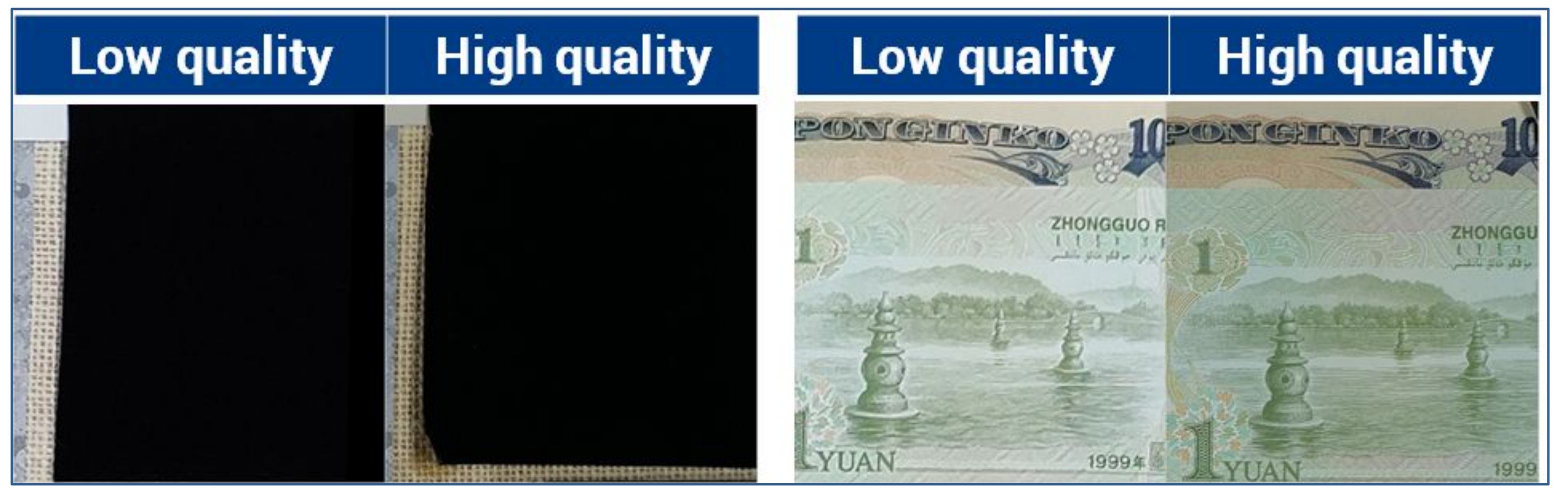
- Too simplistic.
- Highly unnatural details.
- Does not explicitly measure perception.

Dataset

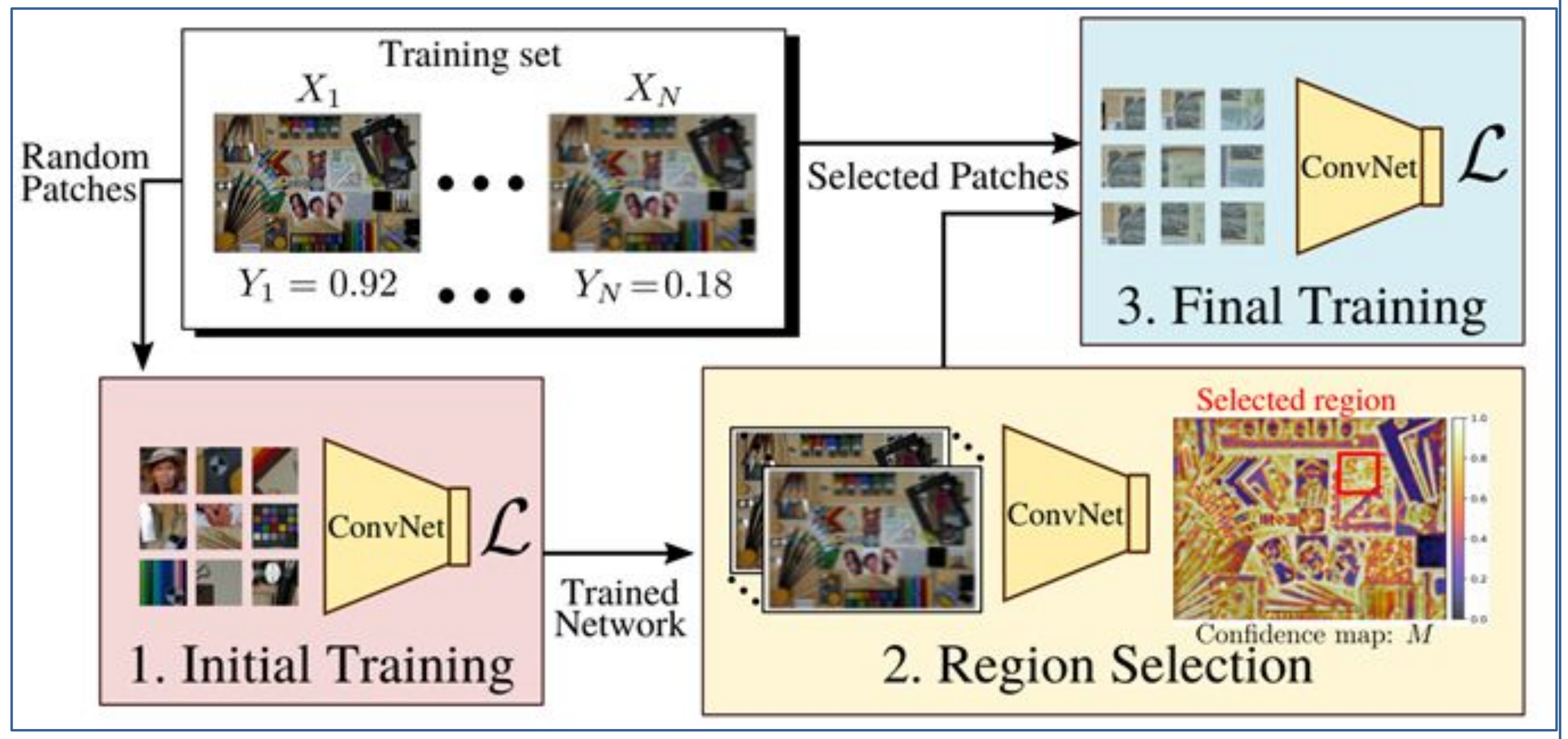
Still-Life chart: 140 devices x 5 lightning conditions
Expert annotations
Details are more natural.
Dataset is registered to a higher resolution using AKAZE [3] keypoint and SCRAMSAC [4] model selection, using bicubic interpolation

Method

Discriminant zone issue :



On the left, this zone present similar content, while on the right details preservation differences are visible

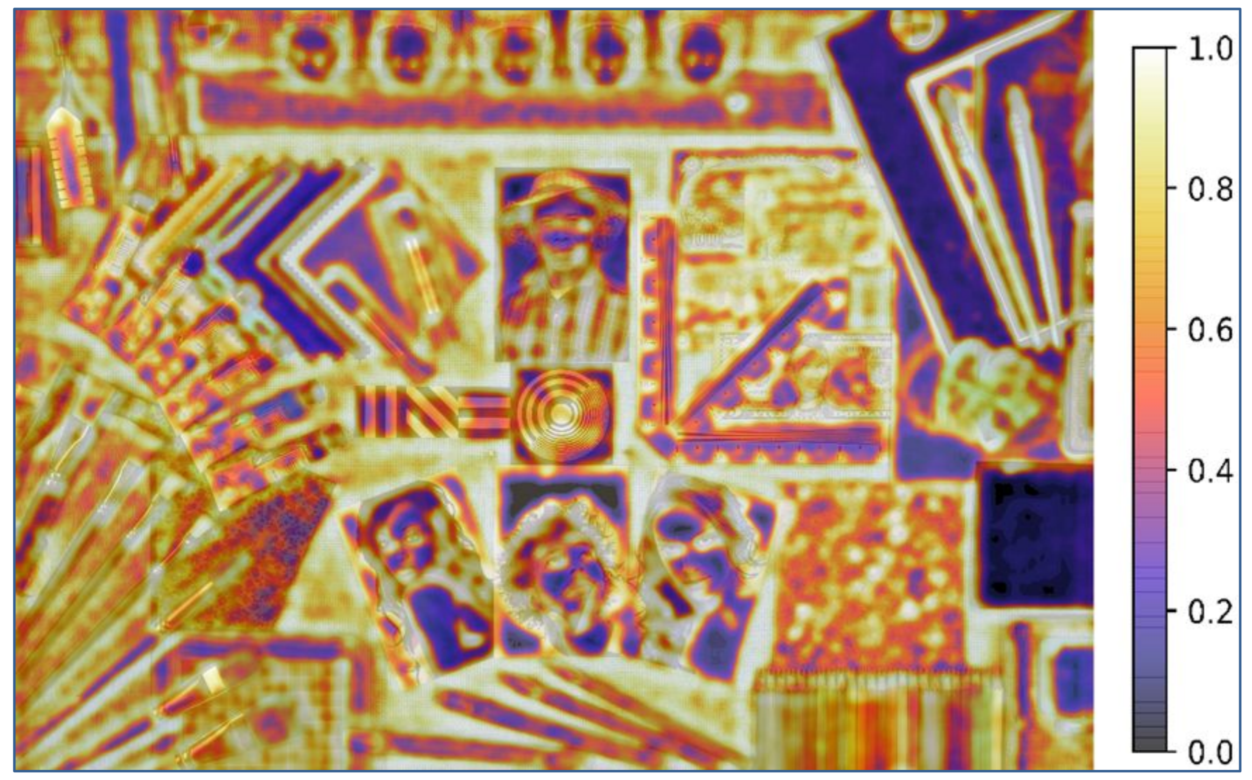


Method Overview

First we train naively a Convolutional Neural Network with patches selected randomly across images. With this naively-trained network on the whole chart, we compute quality score maps, using Class-Activation Map-like technique [5] adapted to the regression setting.

We then estimate a **confidence score map** that indicates the discriminability of each chart region . This map is defined at every location as the variance of the estimated quality scores over the training set.

Discriminant Regions



Ablation Study and SOTA Comparison

- Random Patch: Random patches are selected from the whole chart at both training and testing time.
- Random Region: We restrict the random patch extraction to a single zone, chosen randomly (Average over 5 runs)
- Selected Region: Full pipeline

Number of devices	20	60	100	140	20	60	100	140
	SROCC				**KROCC**			
Random Patch	0.626	0.818	0.784	0.806	0.433	0.617	0.588	0.613
Random Region	0.795	0.863	0.866	0.879	0.606	0.680	0.682	0.700
Selected Region (Full model)	**0.830**	**0.912**	**0.890**	**0.900**	**0.638**	**0.740**	**0.716**	**0.728**

We also compare our method to the MTF-based method and a ResNet trained on the Dead Leaves chart.

Number of devices		20	60	100	140	20	60	100	140
Method	**Chart**	**SROCC**				**KROCC**			
RR Acutance [6]	*Gray-DL*	**0.704**	0.794	0.747	0.788	**0.533**	0.595	0.592	0.592
ResNet [7]	*Gray-DL*	0.641	0.795	0.792	**0.824**	0.464	0.598	0.592	0.630
DR^2S (Ours)	*Still-Life*	**0.830**	**0.912**	**0.890**	**0.900**	**0.638**	**0.740**	**0.716**	**0.728**

While comparisons between results obtained using different charts must be interpreted with care, this result clearly shows that a learning-based approach can be intrinsically better than acutance-based methods using the exact same input images. Finally our DR2S method on the Still-Life chart leads to the best results according to both metrics and for every number of devices.

Conclusion

Our results also suggest that, if enough training samples are available, learning-based methods outperform MTF-based methods. A limitation of our method is that we select only a single region. However, texture quality is known to be multi-dimensional. Consequently, as future work, we plan to extend our method to multiple regions in order to highlight several complementary discriminant features and better measure the intrinsic qualities of a device.

References

[1] G. Boreman, “Modulation Transfer Function in Optical and Electro-optical Systems”. Society of Photo Optical, 2001.
[2] Y. Gousseau and F. Roueff, “Modeling occlusion and scaling in natural images,” Multiscale Modeling & Simulation, 2007.
[3] P. F. Alcantarilla and T. Solutions, “Fast explicit diffusion for accelerated features in nonlinear scale spaces,” IEEE Trans. Pattern. Anal. Mach. Intell, 2011.
[4] Sattler, Torsten, Leibe, and Kobbelt. "SCRAMSAC: Improving RANSAC's efficiency with a spatial consistency filter." 2009 IEEE 12th ICCV. 2009.
[5] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, “Learning deep features for discriminative localization,” in IEEE CVPR, 2016.
[6] F. Cao, F. Guichard, and H. Hornung, “Measuring texture sharpness of a digital camera,” in Digital Photography V. International Society for Optics and Photonics, 2009.
[7] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016.