https://doi.org/10.2352/issn.2694-118X.2021.LIM-5

Portrait Quality Assessment using Multi-Scale CNN

Chahine Nicolas, Belkarfa Salim; DXOMARK Image Labs; Boulogne-Billancourt, France

Abstract

In this paper, we propose a novel and standardized approach to the problem of camera-quality assessment on portrait scenes. Our goal is to evaluate the capacity of smartphone front cameras to preserve texture details on faces. We introduce a new portrait setup and an automated texture measurement. The setup includes two custom-built lifelike mannequin heads, shot in a controlled lab environment. The automated texture measurement includes a Region-of-interest (ROI) detection and a deep neural network. To this aim, we create a realistic mannequins database, which contains images from different cameras, shot in several lighting conditions. The ground-truth is based on a novel pairwise comparison technology where the scores are generated in terms of Just-Noticeable-differences (JND). In terms of methodology, we propose a Multi-Scale CNN architecture with random crop augmentation, to overcome overfitting and to get a low-level feature extraction. We validate our approach by comparing its performance with several baselines inspired by the Image Quality Assessment (IQA) literature.

Introduction

With the rapid rise of camera technology and social media photography, short videos and selfie photos gained a huge interest between users, especially young adults and teenagers. Most of these shots are captured with the front/selfie camera of a smartphone, requiring manufacturers to enhance the quality of their cameras and pushing forward the competition on the social media market. For that, brands compete to deliver the best rendering of portraits, which constitute the main subject of a selfie camera.

The quality of an image can be evaluated on several attributes: target exposure, dynamic range, color (saturation and white balance), texture, noise and different artifacts [1]. For portrait images, the main focus is to evaluate the render quality of the face. On this matter, characteristics like skin tone, bokeh, texture details and skin smoothness are mostly of interest. These attributes should be measured in a way that reflects the human perception. In this study, we aim at evaluating camera capabilities to preserve fine texture details on the face. This study falls within a more general topic that we refer to as *portrait quality assessment*.

When evaluating a camera device, the analysis usually focuses on its capabilities in low-light, zoom or shallow depth-of-field simulation. One standard way to evaluate the quality difference between camera devices is to compare their output in a controlled environment and on the same visual content, namely a *chart*. Using a chart to evaluate the camera quality has both benefits and drawbacks over uncontrolled scenes. On the one hand, it ensures a repeatable measurement and ease of interpretation of the comparison. Since it provides a consistent visual content, environment variance and bias can be eliminated, thus associating the score inconsistency to the camera's capability alone. On the other hand, one can argue that charts are not very good in reflecting the human perception. In fact, lab experiments do not reflect real-world conditions and the visual content of the chart may differ from the content that the final user is interested in.

Cameras have been traditionally evaluated with techniques based on an explicit estimation of the Modulation Transfer Function (MTF) corresponding to the optical system. These approaches can be employed only in the case of synthetically generated visual charts. MTF-based methods suffer from important drawbacks. These methods were originally designed for optical systems that can be modeled as linear. Consequently, non-linear processing such as multi-image fusion or deep learning-based image enhancement, may lead to inaccurate quality evaluation [2]. Moreover, these methods assume that the norm of the device transfer function is a reliable measure of texture quality. However some recent works have shown that the magnitude of image transformations do not always coincide with the perceived impact of the transformations [3]. Therefore, we claim that comparing camera devices on a synthetic visual content is not sufficient to capture the complex behaviour of modern imaging systems.

We favour an explicit contribution of human judgment in the texture quality measurement process. On this matter, deep learning approaches have been proposed recently [4]. In this paper, the authors introduced a convolutional neural network that was trained on a custom database, composed of an arranged collection of objects with different textures and annotated by image quality experts. That said, the focus of this database is to simulate a natural scene environment and not portraits. Also, since it only contains charts, it has strict framing conditions, which does not respond to the natural framing of selfies and human portraits.

Responding to the limitations of the previous techniques, we follow a new and standardized approach for assessing portrait quality preservation. Our contributions are threefold. First, we introduce a novel laboratory setup based on custom-built lifelike mannequins to evaluate the portrait quality preservation of cameras. Using this setup, we establish a portrait database of mannequin images, shot in a controlled lab environment with non-strict framing and annotated by image quality experts. That said, the data we use in the experiments are not available along with this paper. Second, we adopt a deep learning approach based on a multi-scale convolutional neural network (MSCNN) to measure the portrait quality. The scores should reflect the subjective quality judgment; thus, we employ a recent approach of generating subjective image quality scores based on pairwise comparisons to build our ground-truth. Finally, we conduct an extensive study that shows that our mannequin setup combined with the learning-based method can standardize portrait quality assessment of cameras on specific attributes, like details preservation, and performs better than existing methods for general purpose texture quality evaluation.

Related work

In this section, we review the existing work on texture quality assessment and separate it into two categories: MTF-based and learning-based methods. We also review existing IQA datasets. We also review approaches that have previously used portraits as a subject for image quality assessment.

MTF-based methods

MTF-based methods suppose that the camera can be modeled as a linear system that produces an image y as a convolution of the point spread function h and the incoming radiant flux x. In the frequency domain, $Y(f) = H(f)X(f) + N(f)$, where we also consider additive noise N. The modulation transfer function, $MTF(f) = |H(f)|$ is commonly used to characterize an optical acquisition device.

These MTF-based methods assume that the noise-free content, referred to as x, is available in order to estimate the transfer function of the system. This implies that they work the best with synthetic visual charts and cannot be employed for real-world images. Early methods use charts containing a blur spot or a slanted edge for this computation. In [5], Loebich et al. propose a method using the Siemens-Star. Cao et al. propose to use the Dead-Leaves model [6], and introduce an associated method in [7], which is shown to be more appropriate to evaluate fine detail preservation since the texture is more challenging for camera devices. In the rest of this paper, we refer to this chart as the *Dead-Leaves* chart (DL). In this chart, the reference image consists of occluding disks generated with a random center location, radius and grey-scale value.

Importantly, digital camera systems present high-frequency noise, which affects the MTF estimation by dominating signal in the higher frequencies. Consequently, estimating the noise power spectral density (PSD) is key to obtain an accurate acutance evaluation, and this task is not easily performed on the textured region. One approach is to estimate the noise PSD on a uniform patch with a known reference color, which then can be used to estimate the MTF. In modern digital cameras, this approach is affected by denoising algorithms, which not only interfere with the noise PSD, but also perform differently between textured and uniform regions. To bypass this limitation, other approaches [8, 9] propose to compute the MTF using the cross-power spectral density between the reference and target image.

As a conclusion, MTF-based methods might give a good estimation of texture quality. However, it has been shown that MTF does not always reflect the human perception of qualilty. This observation pushes for more modern solutions, mainly learning-based methods that aim at estimating texture quality as perceived by human subjects.

Learning-based methods

In opposition to the MTF-based approach, learning-based methods usually require annotated datasets and can be separated into classical and deep learning approaches. The first automated approaches [10, 11] were based on a combination of hand-crafted features and a regression technique such as support vector regression (SVR). However, these methods were surpassed later by convolutional neural networks (CNN) [12] that proved their superiority in image processing and computer vision applications. Consequently, we focus on deep learning solutions and divide them into two categories, based on their main usage: image quality oriented and camera quality oriented.

Image quality oriented methods

Early datasets (LIVE [13], CSIQ [14], and TIDs [15, 16]) consist of noise-free images processed with several artificial distortions and annotated based on subjective preference. These distortions aim to describe compression or transmission scenarios and most of them are not relevant to the problem of camera evaluation. Besides, they fail to capture the complexity of modern camera systems with non-linear processing pipelines. Other datasets such as KonIQ10k [17] and LIVE In the Wild [18] consist of media-gathered images with unknown distortions. The latter claims that the images were captured using a representative variety of modern mobile devices. For both datasets, annotations were collected using subjective preference crowdsourcing, which we refer to as subjective image quality assessment.

These datasets are large enough to conduct a deep learning solution for image quality evaluation [19, 20, 21]. However, because of their wild nature, uncontrolled environment and their focus on evaluating any input image, they do not form a strong background to evaluate the quality of camera devices, which we are most interested in.

Camera quality oriented methods

Since evaluating camera quality focuses on fixed visual content comparisons, there is a need to define a different training setting, relying on specific datasets. Tworski et al. [4] introduced, to the best of our knowledge, the first of its kind deep learning camera quality evaluation solution. They adopt a regression formulation and train a network to estimate the camera capacity to preserve texture, by comparing images of a common perceptual chart taken with different devices. They introduce a mechanism to identify the chart regions that are the most suited to evaluate quality. Nevertheless, this solution assumes an effective registration of the chart, thus posing a risk of distorting the visual content. Besides, they try to estimate texture preservation quality on a simulated natural scene, which does not respond to our problem of evaluating texture quality on portraits.

Portrait quality assessment methods

Despite the lack of methods that solely focus on evaluating portrait quality preservation of cameras, there exist many attempts to evaluate face quality to help face detection, or even utilize portraits as a subject of image quality assessment. For instance, in [22], the authors propose a method to define a standard portrait image, which can be later used to evaluate colour rendering and other attributes between cross-media. In this work, we can see a first attempt to use a standard portrait as a subject for image quality assessment. Many other approaches to evaluate the face quality as a support for face recognition exist in the literature. In [23], Patrick Grother et al. from the National Institute of Standards and Technology (NIST), intend to support accurate face recognition by establishing specifications for face image quality assessment algorithms as well as evaluate the performance of these algorithms. Other papers [24, 25] propose different methods to evaluate the face quality in the context of face recognition.

Although these solutions deal directly with the problem of face quality, they try to solve a distinct problem, which is not related to camera evaluation. That said, they can still be used in the future as a support for portrait quality preservation measurement, if we need to extend the protocol.

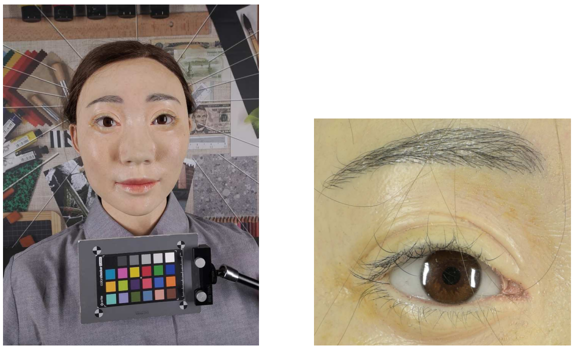
Responding to the limitations and lack of similar work in the literature, we introduce the first standardized texture preservation quality evaluation on portraits, based on deep learning.

Texture quality assessment on portraits

Dataset

Realistic mannequins

In this work, we introduce a new portrait quality database with the main focus on repeatability and reliability. We are primarily interested in texture quality, although it is possible to extend to other attributes like noise, skin tone, and sharpness. To



(a) Realistic mannequin (b) Texture ROI

Figure 1. *(a) Female realistic mannequin head; (b) Eyebrows region for texture evaluation.*

ensure a repeatable measurement, it is important to have a consistent visual content, specifically a chart. However, using a chart similar to what has been used previously, as for example a printed pristine portrait image, will not capture complex behaviour of camera devices. This is especially true when rendering the face's shape and hair's texture, which are not well reflected in a printed image. For this reason, we employ a dataset of custom-built lifelike mannequins, referred to as *realistic mannequins* (RM).

The dataset contains images of silicone heads, shot at a fixed distance, in different lighting conditions, and with several smartphones' front camera. Apart from the strict distancing, the framing is loose and is not fixed to a strict angle, contrarily to the usual chart setups. This configuration helps to simulate a real self-portrait (selfie) scenario as shown in Figure 1a.

Ground truth & region of interest

Many attributes can be extracted from the realistic mannequins, in particular texture details preservation. Practically, to ensure a reliable measurement, we need to focus on a specific region of interest (ROI). In the case of texture measurement, we focus on the eye and eyebrows region, as shown in figure 1b. We choose the highest resolution in the dataset and upscale the other crops to match it, keeping the same aspect ratio.

To obtain a ground truth for the dataset, we need to collect subjective preference evaluations. However, evaluating the image quality in an absolute manner is not a simple task, and is still a hot topic in research. One of many approaches is pairwise comparison, where one can infer quality scores out of a comparison matrix. The main problem with this approach is its quadratic growth, which means that the cost and difficulty increase in a quadratic manner with the size of the dataset. Following ITU-T [26] and ITU-R [27] recommendations, a minimum of 15 full comparisons $\mathcal{O}(n^2)$ (i.e. 15 annotators to compare all $n(n-1)/2$ pairs) is required to generate reliable results, which can be very costly and time consuming. Usually it is not possible to achieve a complete design (a full passage on all data points for each observer), which implies the need for a sampling strategy or *active sampling* as it's referred to in the literature.

We adopt the research of Mikhailiuk et al. [28] in order to efficiently utilize observers' time and obtain the most possible accurate scale. In this algorithm, we select the next comparison to deliver the most information, i.e., the one that has the highest impact on the posterior distribution of the scores. For that, separately for each image, we obtained the distribution of quality scores after every performed comparison using Expectation Propagation (EP), assuming Thurstone case V [29], and then estimated the distribution of the scores assuming every possible future comparison. The comparison maximizing the Kullback-Leibler divergence between the current distribution and the result after every possible comparison was chosen to be performed next.

The analysis and pairwise comparisons were conducted in a controlled environment with a fixed viewing condition. We adopt our setting so that our viewing condition is aligned with that of a human eye, with a cutoff frequency ν_{cut} = 30cpd. Hence, we use a 32" 4k monitor with a pixel pitch of 0.185 , and we fix the eye to screen distance to 65cm. We use a calibrated display (D65 whitepoint with luminance $\geq 75cd/m^2$ with no direct illumination of the screen and a background illumination with a lighting panel set to D65 / 15% for reducing eye stress.

Method

In this section, we detail the proposed method for estimating texture quality on portraits. We formulate this task as a regression problem and suppose that a perfect registration of the images is not needed.

ROI extraction

The first step in evaluating the texture quality on portraits is to correctly detect the ROI. Usually, the ROI is fixed and its detection assumes an effective registration of the chart. However, because of the inaccurate framing of portraits, registration is a bad choice, since it can introduce face distortions and alter the quality of the image. Therefore, using a manually extracted reference ROI, with a sufficiently high resolution, we could determine the crop dimensions in the other images. Thoroughly, we detect a set of facial landmarks to localise the eyes region. Then, we fix the width of the crop to cover the ROI horizontally and compute the height using the same reference aspect ratio. Finally, we upscale the crop to the reference resolution.

Batch creation with random crops

In our problem, the dataset is content-specific, meaning we have the same content in every image (same face). Consequently, there is a high possibility of overfitting when using a complete image. To avoid this, we use random crops as a data augmentation technique. We create a batch of N images, where every image is represented by n patches, assigned with the same score as the source image; the total batch size is then $N * n$, where each patch is considered as an independent image. Also, By using random crops, we can fix the input size without the need to resize the original image, which maintains the quality and solves the problem of limited memory. Moreover, with random crops there is a possibility of advanced extensions, as variance-based random crops and attention-based crops, as in [30, 31]. Nonetheless, random crops can increase the variance and limit the training capacity, since some of the crops, especially skin crops, are not relevant and can alter the average score. To tackle this problem, we use the Huber loss as described in [4].

Multi-scale CNN

This paper aims to provide an automatic, all-in-one protocol to measure the portrait quality preservation capabilities of cameras. Therefore, we do not intend to introduce state-of-the-art neural network architectures, but we prefer to adopt proven working solutions.

Inspired by the work of Chen et al. [32], we adopt a multi-scale architecture for our convolutional neural network model, referred to as MSCNN. This technique aims to solve a problem with previous random patch methods, by paying equal attention to global and local features. The model consists of three convolutional blocks. Each convolutional block extracts specific features from one scale. Features are then concatenated into one vector and fed to a fully connected block (FC) with four layers. The final output is a single float value, representing the final quality value of the input list.

Since we have limited resources, we are not able to train a model that big from the ground up. For instance, the large parameter set can be easily trained on our relatively small dataset, however, the model will overfit because it is pretty complex for a small dataset. An intuitive solution to the problem is fine-tuning. We use the convolutional backbone of a well known CNN architecture that was pretrained on a large dataset, to extract relevant features from the image. The fine-tuning process significantly decreases the time required to train and optimize a new model, especially for a smaller task, as it is already trained on a larger dataset, and is assured to give robust results. Finally, since we consider random crops as augmentation of the data, we optimize the loss over individual patch scores instead of the average image score.

Experiments

In this section, we perform an extensive experimental study of the proposed solution. The study consists of two tests: first, we test the performance of the measure on a mannequin database, which we create separately from the training database, referred to as the ground-truth. Second, we compare the measure with an MTF-based general purpose texture quality measure. All the data are internally collected and might not be available outside of the working team.

Training details

We implemented our deep learning model with PyTorch, which was then trained on a 48 devices dataset of realistic mannequin images, shot in three lighting conditions, a total of 144 images. All the images are in fact ROI crops that were upscaled to the same reference resolution with a fixed aspect ratio, using bicubic upsampling. We choose the reference resolution as the highest crop resolution in our dataset. We adopt a transfer learning technique using a MobileNetV3 [33] pre-trained on ImageNet [34]. We employ Adam optimizer, with an initial learning rate of $1 \cdot 10^{-5}$ and a decay of 10% every 5 epochs for a total of 50 epochs. To limit the effects of noise, contrast, exposure, saturation and over sharpening on the texture score, we add 7 different augmentations that slightly alter these attributes. After visually checking many examples of the augmented data, we conclude that the small alteration will not change the texture quality, thus we keep the same score for each augmentation. In addition to that, we add horizontal and vertical flips as well as a grayscale variation. In total, our training database contains 1584 images.

Test datasets

To ensure that the new method performs better than previous solutions, we test its performance on a well-established ground truth, which was carefully annotated by human observers. Also, we compare the correlation of our scores with the MTF based methods; we aim to conclude whether the MTF-Based methods are sufficient to evaluate the texture preservation quality of a selfie camera, or do we absolutely need a solution that is inspired by and respects human observations.

As there is no well-established reference dataset for our problem, we collected and annotated our own data on two different charts/visual content.

New-Resolution

First, we create a reference dataset of realistic mannequin images, shot at a fixed distance, in two different lighting conditions. We collect data from 23 smartphone front cameras, a total of 46 images. This dataset is referred to as *New-Resolution* (NR). The procedure to collect and annotate the images is similar to what has been done for the training dataset (Section Dataset). We choose to separate the annotations of the RM and the NR datasets, in order to ensure that no hidden correlation exists between the scores. Additionally, no device has been shared between both datasets.

To ensure a good variability in our texture evaluation, we shoot mannequins at two distances: 30cm and 55cm. For the 30cm, we only provide female mannequins, while for the 55cm we only provide male mannequins with a consistent beard. To evaluate the texture, we use eyebrow crops for the female, and eyebrow and beard crops for the male. Finally, for each setup, we evaluate the models that were trained on its respective training set, as well as evaluate the average outputs of both setups.

Gray-DL

Second, we employ the Dead-Leaves chart proposed in [7] (Section MTF-based methods). In our experiment, we refer to this dataset as Gray-DL. We use the same two lighting conditions and 23 devices as in the *New Resolution* dataset. In the case of the Dead-Leaves charts, since the charts are unnatural images, human perceptual annotation is problematic. Since we do not have a proper ground truth for this chart, we choose to evaluate it on the ground truth of the *New Resolution* dataset, similarly to what we do on our deep learning model.

Nevertheless, there exists one limitation that concerns us when assigning the NR ground truth to the Gray-DL dataset, which we try to explain and solve next. In fact, the Dead-Leaves chart is shot with a specific framing, and depending on the camera's field of view (FOV), the lens to chart distance can lie between 45cm and 55cm. Although the DL chart is theoretically scale invariant and thus the measurement should not depend on the distance, the device's range of focus will drastically alter the results. For instance, one device might feature an *Autofocus* (AF), while the other might have a fixed range of focus. Usually, this range is between 30 and 60cm, but can be freely positioned by the manufacturer. For this reason, only the 55cm distance is relevant to evaluate the Gray-DL scores. To solve this problem, we choose to test on two different device sets. The first set includes all 23 devices (46 images) for the 55cm distance, referred to as the *Full* set, while the second only includes 13 devices (26 images) that have AF and/or a wide range of focus, covering both 30cm and 55cm distances; we refer to this set as the *Limited* set.

Quantitative study

Metrics

Since we adopt a regression formulation, we need to use relevant metrics that comply with an image quality score. That said, this score is not absolute and is usually relative to the approach of origin. Consequently, MTF-based methods predict a quality score that is not directly comparable to the score provided by human annotators. Also, since we annotate the RM and NR datasets separately, their scores lie on separate scales, which makes a direct comparison irrelavant. One alternative is to explore the linear correlation between the output and the ground truth. However, assuming that the predictions correlate linearly with the ground truth may not hold and bias the experiment. Therefore, we adopt two metrics that are commonly used in the IQA literature, and that are based on the correlation of the rank-order. First, we adopt the Spearman Rank-Order Correlation Coefficient (SROCC) defined as the linear correlation coefficient of the ranks of between two variables. Second, we report the Kendall Rank-Order Correlation Coefficient (KROCC) defined by the difference between concordant and discordant pairs divided by the number of possible pairs. The key advantage of the second metric lies in its robustness to outliers.

Results

We compare the MTF-based measurement on the Dead-Leaves chart with the ground truth of the *New-Resolution* dataset for the 30cm and 55cm distances, as well as for the average score over both datasets. For the realistic mannequin texture evaluation using the MSCNN, we compare the output of each model with its respective NR dataset. Additionally, we compare the average of both outputs to the average score of the NR datasets. All the results are shown in Tables 1, 2.

As we explained in Section Gray-DL, we proceed with two different evaluations based on the focus characteristics of each device. To ensure a correct ground truth, we compute the correlation between the NR-55cm and NR-30cm for the *Limited* set. If the value of this correlation is large enough, we can suppose that we have the same device ranking regardless of the distance. For the *Limited* set with 13 devices, we get **srocc = 0.935** & **krocc = 0.796**, which ensures a similar ranking across both distances.

On both device sets we notice that the MSCNN based methods give larger correlations over the MTF-based methods. Moreover, we can see that the model performs better on the 30cm distance. That said, we should take the results with a grain of salt. Since the *Limited* set's size is relatively small, we cannot extract a statistically significant conclusion. However, when we back these results up with the *Full* set, we have a higher significance to support our new method. Because of the fixed range of focus, the evaluation of texture can change drastically between distances, which explains why we do not include the comparison for the 30cm distance on the *Full* set. However, in average, over the two distances, we can get an idea of the texture preservation capabilities of a smartphone front camera, which is expressed here in the NR-AVG scores.

Because of the ever-growing image quality enhancement technologies, new camera systems incorporate specific pipelines for certain scenes, especially portraits. One advantage of the RM-based approach is that a mannequin face triggers these pipelines, which helps to correctly assess the capabilities of the selfie camera. This scenario is not possible with a normal chart. Therefore, we see that the RM-based method can give a more accurate ranking and a better evaluation of the texture preservation quality of a selfie camera, where portraits are the main subject.

Conclusion

In this paper, we proposed a novel and standardized approach to the problem of camera quality assessment on portrait scenes. We developed a repeatable setup as well as a deep learning-based method to test the texture preservation quality on portraits, inspired by human observations. Our results show that a deep learning approach, can outperform the classical measurements and is more relevant to human preference. One limitation of our method is that it only works on specific mannequins, and not on real human portraits. Consequently, as future work, we plan to extend our database to cover human portraits, and multiple human portraits, in order to highlight several complementary discriminant features and better measure the intrinsic qualities of a smartphone front camera, in a real-life scenario.

Full **set - Comparison of MSCNN trained on realistic mannequins and the MTF-based measurement on the Dead-Leaves chart**

Method	Chart	GT	SROCC	KROCC
Acutance	Gray-DL	NR-55cm	0.748	0.567
MSCNN	RM-55cm	NR-55cm	**0.874**	**0.717**

Limited **set - Comparison of MSCNN trained on realistic mannequins and the MTF-based measurement on the Dead-Leaves chart**

Method	Chart	GT	SROCC	KROCC
Acutance	Gray-DL	NR-30cm	0.760	0.636
Acutance	Gray-DL	NR-55cm	0.779	0.606
Acutance	Gray-DL	NR-AVG	0.769	0.618
MSCNN	RM-30cm	NR-30cm	0.940	0.821
MSCNN	RM-55cm	NR-55cm	0.851	0.698
MSCNN	RM-AVG	NR-AVG	**0.916**	**0.796**

References

[1] Čadík, Martin, et al. ”Image attributes and quality for evaluation of tone mapping operators.” National Taiwan University. 2006.

[2] van Zwanenberg, Oliver, Sophie Triantaphillidou, Robin Jenkin, and Alexandra Psarrou. ”Edge detection techniques for quantifying spatial imaging system performance and image quality.” In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 0-0. 2019.

[3] Zhang, Richard, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. ”The unreasonable effectiveness of deep features as a perceptual metric.” In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 586-595. 2018.

[4] Tworski, Marcelin, Stéphane Lathuilière, Salim Belkarfa, Attilio Fiandrotti, and Marco Cagnazzo. ”DR2S: Deep Regression with Region Selection for Camera Quality Evaluation.” arXiv preprint arXiv:2009.09981 (2020).

[5] Loebich, Christian, Dietmar Wueller, Bruno Klingen, and Anke Jaeger. ”Digital camera resolution measurements using sinusoidal Siemens stars.” In Digital Photography III, vol. 6502, p. 65020N. International Society for Optics and Photonics, 2007.

[6] Gousseau, Yann, and François Roueff. ”Modeling occlusion and scaling in natural images.” Multiscale Modeling & Simulation 6, no. 1 (2007): 105-134.

[7] Cao, Frédéric, Frederic Guichard, and Hervé Hornung. ”Measuring texture sharpness of a digital camera.” In Digital Photography V, vol. 7250, p. 72500H. International Society for Optics and Photonics, 2009.

[8] Kirk, Leonie, Philip Herzer, Uwe Artmann, and Dietmar Kunz. ”Description of texture loss using the dead leaves target: current issues and a new intrinsic approach.” In Digital Photography X, vol. 9023, p. 90230C. International Society for Optics and Photonics, 2014.

[9] Sumner, Robert C., Ranga Burada, and Noah Kram. ”The Effects of misregistration on the dead leaves crosscorrelation texture blur analysis.” Electronic Imaging 2017, no. 12 (2017): 121-129.

[10] Mittal, Anish, Anush Krishna Moorthy, and Alan Conrad Bovik. ”No-reference image quality assessment in the spatial domain.” IEEE Transactions on image processing 21, no. 12 (2012): 4695-4708.

[11] Ye, Peng, Jayant Kumar, Le Kang, and David Doermann. ”Unsupervised feature learning framework for no-reference image quality assessment.” In 2012 IEEE conference on computer vision and pattern recognition, pp. 1098-1105. IEEE, 2012.

[12] Kim, Jongyoo, Hui Zeng, Deepti Ghadiyaram, Sanghoon Lee, Lei Zhang, and Alan C. Bovik. ”Deep convolutional neural models for picture-quality prediction: Challenges and solutions to data-driven image quality assessment.” IEEE Signal processing magazine 34, no. 6 (2017): 130-141.

[13] Sheikh, Hamid R., Muhammad F. Sabir, and Alan C. Bovik. ”A statistical evaluation of recent full reference image quality assessment algorithms.” IEEE Transactions on image processing 15, no. 11 (2006): 3440-3451.

[14] Larson, Eric Cooper, and Damon Michael Chandler. ”Most apparent distortion: full-reference image quality assessment and the role of strategy.” Journal of electronic imaging 19, no. 1 (2010): 011006.

[15] Ponomarenko, Nikolay, Vladimir Lukin, Alexander Zelensky, Karen Egiazarian, Marco Carli, and Federica Battisti. ”TID2008-a database for evaluation of full-reference visual quality assessment metrics.” Advances of Modern Radioelectronics 10, no. 4 (2009): 30-45.

[16] Ponomarenko, Nikolay, Lina Jin, Oleg Ieremeiev, Vladimir Lukin, Karen Egiazarian, Jaakko Astola, Benoit Vozel et al. ”Image database TID2013: Peculiarities, results and perspectives.” Signal processing: Image communication 30 (2015): 57-77.

[17] Hosu, Vlad, Hanhe Lin, Tamas Sziranyi, and Dietmar Saupe. ”KonIQ-10k: An ecologically valid database for deep learning of blind image quality assessment.” IEEE Transactions on Image Processing 29 (2020): 4041-4056.

[18] Ghadiyaram, Deepti, and Alan C. Bovik. ”Massive online crowdsourced study of subjective and objective picture quality.” IEEE Transactions on Image Processing 25, no. 1 (2015): 372-387.

[19] Chen, Diqi, Yizhou Wang, Tianfu Wu, and Wen Gao. ”An Attention-Driven Approach of No-Reference Image Quality Assessment.” arXiv preprint arXiv:1612.03530 (2016).

[20] Varga, Domonkos, Dietmar Saupe, and Tamás Szirányi. ”DeepRN: A content preserving deep architecture for blind image quality assessment.” In 2018 IEEE International Conference on Multimedia and Expo (ICME), pp. 1-6. IEEE, 2018.

[21] Yang, Dan, Veli-Tapani Peltoketo, and Joni-Kristian Kamarainen. ”CNN-based cross-dataset no-reference image quality assessment.” In Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, pp. 0-0. 2019.

[22] Kanafusa, Kunihiko, et al. ”A Standard Portrait Image and Image Quality Assessment.” IS AND TS PICS CONFERENCE. SOCIETY FOR IMAGING SCIENCE & TECHNOLOGY, 2000.

[23] Grother, Patrick, Mei Ngan, and Kayee Hanaoka. ”Face recognition vendor test-face recognition quality assessment concept and goals.” In NIST. 2019.

[24] Terhorst, Philipp, Jan Niklas Kolf, Naser Damer, Florian Kirchbuchner, and Arjan Kuijper. ”SER-FIQ: Unsupervised estimation of face image quality based on stochastic embedding robustness.” In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5651-5660. 2020.

[25] Chen, Jiansheng, Yu Deng, Gaocheng Bai, and Guangda Su. ”Face image quality assessment based on learning to rank.” IEEE signal processing letters 22, no. 1 (2014): 90-94.

[26] ITU-T RECOMMENDATION, P. ”Subjective video quality assessment methods for multimedia applications.” International telecommunication union (1999).

[27] BT, RECOMMENDATION ITU-R. ”Methodology for the subjective assessment of the quality of television pictures.” International Telecommunication Union (2002).

[28] Active Sampling for Pairwise Comparisons via Approximate Message Passing and Information Gain Maximization

[29] Murphy, Kevin P. Machine learning: a probabilistic perspective. MIT press, 2012.

[30] Po, Lai-Man, Mengyang Liu, Wilson YF Yuen, Yuming Li, Xuyuan Xu, Chang Zhou, Peter HW Wong, Kin Wai Lau, and Hon-Tung Luk. ”A novel patch variance biased convolutional neural network for no-reference image quality assessment.” IEEE Transactions on Circuits and Systems for Video Technology 29, no. 4 (2019): 1223-1229.

[31] Chen, Diqi, Yizhou Wang, and Wen Gao. ”No-reference image quality assessment: An attention driven approach.” IEEE Transactions on Image Processing 29 (2020): 6496-6506.

[32] Chen, Peikun, Yuzhen Niu, and Dong Huang. ”No-Reference Image Quality Assessment Based on Multi-scale Convolutional Neural Networks.” In Intelligent Computing-Proceedings of the Computing Conference, pp. 1202-1216. Springer, Cham, 2019.

[33] Howard, Andrew G., Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. ”Mobilenets: Efficient convolutional neural networks for mobile vision applications.” arXiv preprint arXiv:1704.04861 (2017).

[34] Deng, Jia, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. ”Imagenet: A large-scale hierarchical image database.” In 2009 IEEE conference on computer vision and pattern recognition, pp. 248-255. Ieee, 2009.

Author Biography

Nicolas Chahine is a machine learning Ph.D. student. He followed a double degree program between the Lebanese university faculty of engineering and Telecom Paris (2014-2020). He also followed a masters degree in applied mathematics, namely MVA, at the University of Paris Saclay in collaboration with Ecole Normale Superieur (2019-2020). Since December 2020, he is working full time at DXOMARK Image Labs as a Ph.D. student in collaboration with INRIA Paris. His work focuses on automated image quality assessment.

Salim Belkarfa received his Master’s degree in engineering from Telecom Bretagne (2013). He joined DXOMARK Image Labs in 2015 to continue developing state of the art camera quality assessment technology.