https://doi.org/10.2352/EI.2022.34.14.COIMG-219

Transfer learning for no-reference image quality metrics using large temporary image sets

Mykola Ponomarenko, Sheyda Ghanbaralizadeh Bahnemiri and Karen Egiazarian
Tampere University, Tampere, Finland

Abstract

One of the main problems of neural network-based no-reference metrics design for image visual quality assessment is small size of image databases with mean opinion scores (MOS). For large networks which can memorize key features of several thousands of images, usage of the databases for metrics training may lead to overlearning. Since data augmentation for image quality assessment is limited by a horizontal image flipping only, the main way to decrease overlearning is to use transfer learning which can significantly speed up training process. In theis paper, we propose a new technique of transfer learning between networks of different architectures using a large set of images without MOS. We implemented the technique for transfer learning between pre-trained KonCept512 metric and a IMQNet metric proposed in this paper. An effectiveness of the transfer learning is estimated in a numerical analysis. It is shown that the trained IMQNet metric provides significantly better correlation with KonCept512 metric (0.89) than other modern metrics. It is also shown that IMQNet pre-trained by the proposed transfer learning shows better correlation with MOS of KonIQ-10k database (0.86) than IMQNet pre-trained using directly the MOS of KonIQ10k (0.73).

Introduction

No-reference image quality assessment (NR-IQA) is the area of intensive research during last decades [1-2]. A NR-IQA metric with a good correspondence to human perception can be useful in many practical applications: in-camera image visual quality index, automatic selection of best image settings in acquisition process, image indexing in large image databases and search engines, custom loss function for neural networks training.

For training and verification of no-reference metrics many specialized image databases with mean opinion scores (MOS) are proposed. The largest are FLIVE (40 000 images) [3], SPAQ (11 000 images) [4], KonIQ-10k (10 000 images) [2], HTID (2880 images) [5], Live-in-the-Wild (WILD, 1200 images) [6].

Many NR-IQA metrics are designed till the moment. During the last decade, convolutional neural networks (CNN) show great progress in solving the task of NR-IQA. KonCept512 [2] metric can be mentioned as providing relatively good Spearman rank order correlation coefficient (SROCC) values with MOS on most databases. However, no one of the metric shows enough correspondence to human perception to be used in practice.

Design of a database with MOS is a very time-consuming task. To obtain MOS value for an image, one must collect up to 100 judgments of image visual quality [2]. Because of this, image databases for NR-IQA contain only thousands of images, whereas databases for image classification contain millions of images. Authors of FLIVE have collected MOS of 40 000 images due to decreasing a number of judgments to 20 per image. However, as a result, MOS of FLIVE has much worse quality than MOS of other databases.

Large CNNs show good efficiency in tasks of image classification [7], object detection [8], noise [9] and blur [10] parameters estimation. Potentially they can show a good correspondence to human perception in NR-IQA as well. The main problem here is overlearning. Image databases with MOS are relatively small and large CNNs are able to memorize key features of all images in the databases. One needs at least hundreds of thousands of images for training of large CNN based NR-IQA metrics without overlearning. Usage of data augmentation for NR-IQA is limited by a horizontal image flipping only, because addition of noise, blur, cropping and arbitrary rotations can unpredictably decrease image visual quality.

In [11], an efficient technique of merging of MOS of several databases into one large database is proposed. It decreases the overlearning problem, but does not eliminate it.

Authors of KonCept512 metric used the transfer learning approach to mitigate the overlearning problem [2]. Architecture of KonCept512 metric is based on the architecture of Inception-ResNet-v2 network [12]. KonCept512 was trained using preliminary transfer learning from Inception-ResNet-v2 pre-trained for image classification.

However, such transfer learning can be applied only between networks with the same architecture (except of several final layers). For a network with a new architecture this approach is not applicable. Also, it is not possible to add additional input layers to existing architecture without destroying a possibility of usage of classical transfer learning from the pretrained network.

In this paper, we propose a new approach to transfer learning for NR-IQA metrics using a large temporary image set. This approach allows to transfer knowledge between NR-IQA metrics of totally different architectures without the overlearning effect.

Proposed transfer learning scheme

A structural scheme of the proposed approach is drawn in Fig. 1. Learning is transferred from a good pre-trained NR-IQA metric A to NR-IQA metric B in the following way:

1. A large image set TMP without MOS is collected. TMP should include at least several hundreds of thousands of images to prevent overlearning.
2. Predictions of image quality of TMP images are calculated using metric A. Let us call the array of predictions as PA.
3. Metric B is pretrained on TMP database to predict PA values.

In the case of enough representativity of TMP database, metric B will be pre-trained to do the same work as the metric A. Let us note that metrics A and B can have different network architectures.

There are three main requirements for collected TMP set.

First, TMP set should be representative enough including images with all kinds of image quality factors (distortions) typical for existing image databases with MOS.

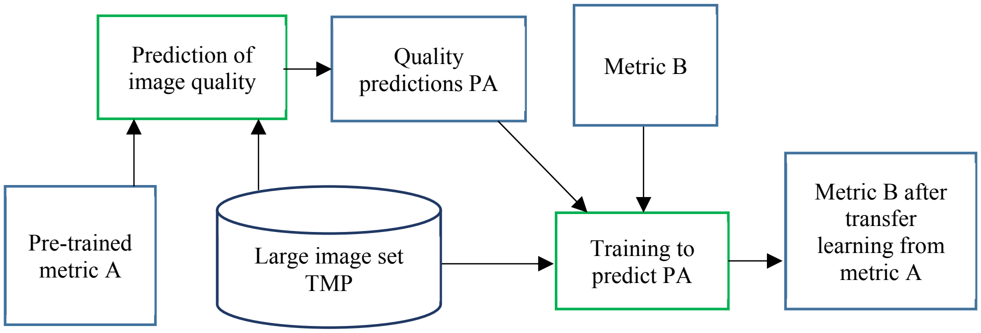


Fig. 1. Structure scheme of the proposed algorithm of transfer learning between metrics with different network architectures

Second, it should contain images with the same range of PA values as MOS of databases used for training of metric A.

Third, images of TMP should be difficult enough for PA prediction.

Fitting of the second and third requirements can be analyzed statistically. We will show it in the paper on a practical example.

Selection of pre-trained metric for transfer learning

Table 1 shows Spearman rank order correlation coefficient (SROCC) values for five no-reference databases with MOS: KonIQ-10k, FLIVE, WILD, NRTID [1] and HTID. The last column contains weighted aggregated SROCC calculated as

$$\mathrm{SROCC_W} = |\ 0.25\ \mathrm{S_{KonIQ\text{-}10k}} + 0.15\ \mathrm{S_{FLIVE}} + 0.2\ \mathrm{S_{NRTID}} + 0.25\ \mathrm{S_{WILD}} + 0.15\ \mathrm{S_{HTID}}\ |, \quad (1)$$

where S stands for SROCC values.

Table 1. SROCC values for five databases and weighted aggregated SROCC

	Metric	Koniq10	FLIVE	NRTID	WILD	HTID	$\mathrm{SROCC_w}$
1	KonCept512 [2]	0.83	0.45	0.69	0.79	0.66	0.71
2	PaQ-2-PiQ [3]	0.72	0.52	0.76	0.71	0.26	0.63
4	Otroshi [13]	0.84	0.32	0.56	0.61	0.49	0.60
3	UIQA [14]	0.62	0.21	0.77	0.57	0.60	0.57
6	FISH [15]	0.61	0.26	0.6	0.53	0.12	0.46
7	Smetric [16]	0.61	0.23	0.71	0.33	0.25	0.45
5	C-DIIVINE [17]	0.48	0.19	0.58	0.49	0.37	0.44
8	ilniqe [18]	0.54	0.19	0.38	0.44	0.57	0.44
9	DIIVINE [17]	0.5	0.17	0.48	0.46	0.38	0.42
10	DB-CNN [19]	0.54	0.2	0.28	0.43	0.57	0.41

As the metric A for testing of the proposed transfer learning algorithm we select the metric KonCept512 providing good SROCCs for all five databases and state-of-the-art SROCCw. In this paper, we will use KonCept512 after transfer learning from Inception-ResNet-v2 and training on the merged MOS [11] of six databases: KonIQ-10k, FLIVE, WILD, NRTID, HTID and SPAQ. MOS of the databases after merging [5] are given in the Fig. 2.

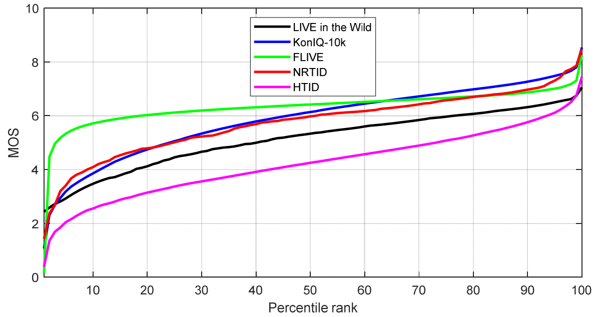


Fig.2. Ranges of merged MOS of six image databases

Collecting images for TMP image set

To test the proposed methodology, we have randomly selected 360000 images from Google Open Images dataset [20]. From each image a central part with the size 1024x768 was cropped.

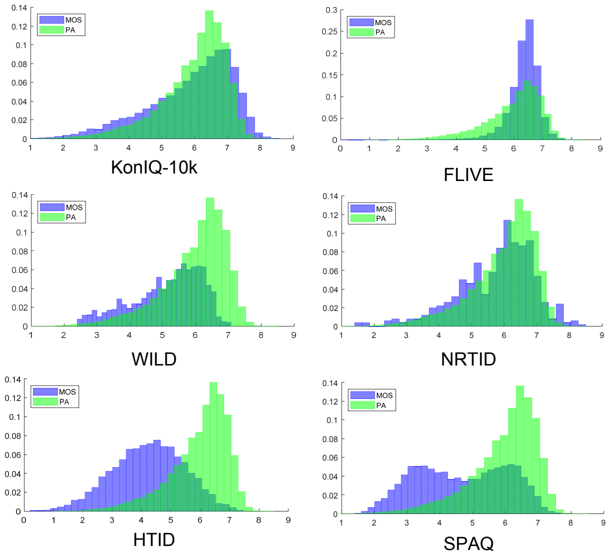


Fig. 3. Histogram of PA in comparison with histograms of six image databases with MOS

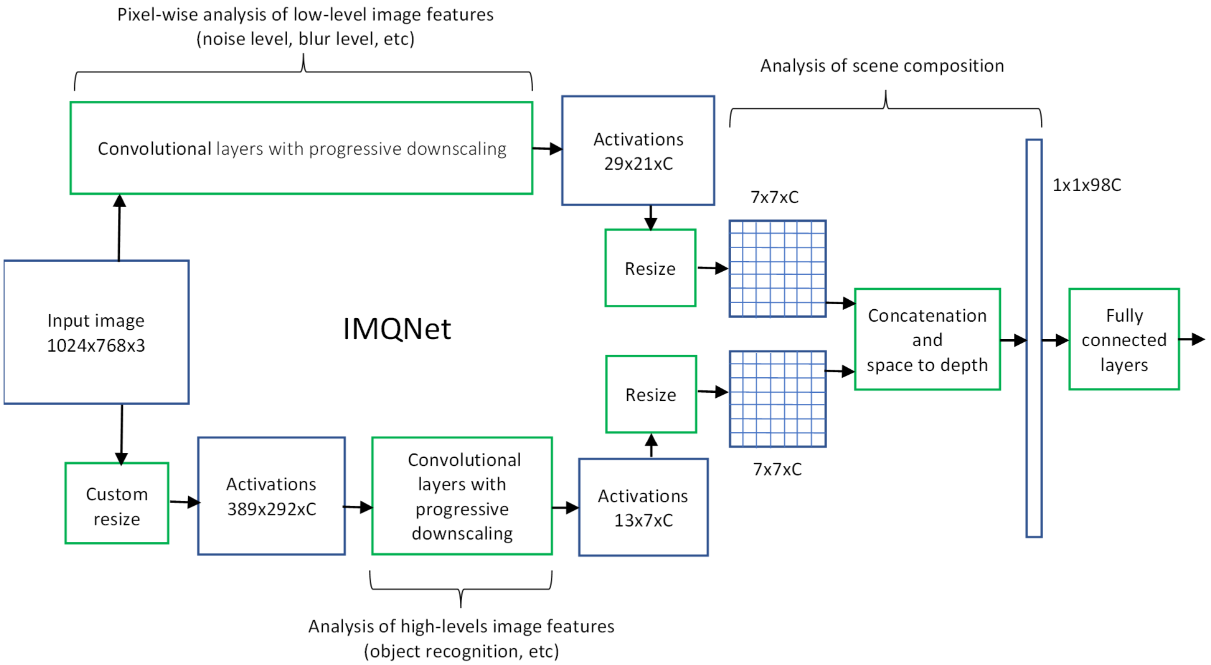


Fig. 4. Structural scheme of IMQNet

We will call a set of collected images as TMP set.

Array of PA values was calculated as predictions of visual quality of images of TMP set using KonCept512 metric. Fig. 3 shows histogram of PA values combined with histograms of image databases used for KonCept512 training.

One can see that PA values cover a range of values of MOS of existing databases well enough except of small ranges of largest and smallest values. Thus, collected TMP set meets the second requirement of the proposed transfer learning methodology.

Let us check the correspondence of TMP set to the third requirement. Table 2 contains SROCC values between PA and quality predictions of images of TMP set by good metrics PaQ-2-PiC, Otroshi and UIQA.

Table 2. SROCC between PA and metrics values for TMP set

PaQ-2-PiQ	0.77
Otroshi	0.75
UIQA	0.59

The value of 0.77 is not too high, so prediction of PA values is not a trivial task. Thus, images of TMP set fit well the third requirement to the proposed transfer learning algorithm.

Network architecture for verification of the proposed algorithm of transfer learning

To test the proposed approach of transfer learning, we designed a metric architecture IMQNet presented in Fig. 4.

IMQNet analyzes all image features affecting image visual quality. The network takes input images downscaled two times by Matlab's bicubic interpolation and has two branches. The structure in Fig. 4 is given for input image 1024x768 pixels, but the network can work with images of an arbitrary size.

First branch of IMQNet analyzes low-level image features (e.g., noise and blur levels). It starts from the full-size input and has eleven layers blocks which slowly analyze and downscale the input image.

A structure of analyzing or downscaling layers block is given in Fig. 5. The block has three sub-branches containing 2D convolutional layers and ReLu activations. In a downscaling block, the last convolutional layer of each three sub-branches has "Stride" parameter set to 2.

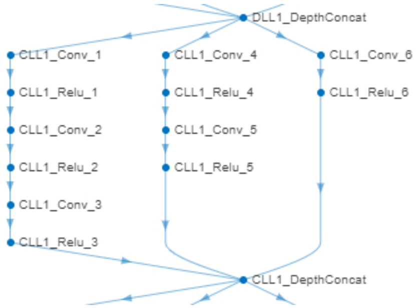


Fig.5. Main analyzing and downscaling layers block of IMQNet

The branch ends by a resize layer resizing activations to the size 7x7xC, where C is a number of channels in activations. Output of this layer contain results of low-level features analysis divided into 7x7 image regions. It gives to IMQNet ability to analyze spatial representation of low-level image features (scene composition).

Second branch of IMQNet is used to detect high-level image features (e.g., face detection, object classification).

It starts from a custom resizing layer keeping input image proportions. Image on the output of the layer always has smaller size equal to 292.

The branch has sixteen layers blocks similar to the blocks of the first branch. The branch ends by a resize layer resizing activations to the size 7x7xC, which gives to IMQNet ability to analyze spatial representation of high-level image features (scene composition).

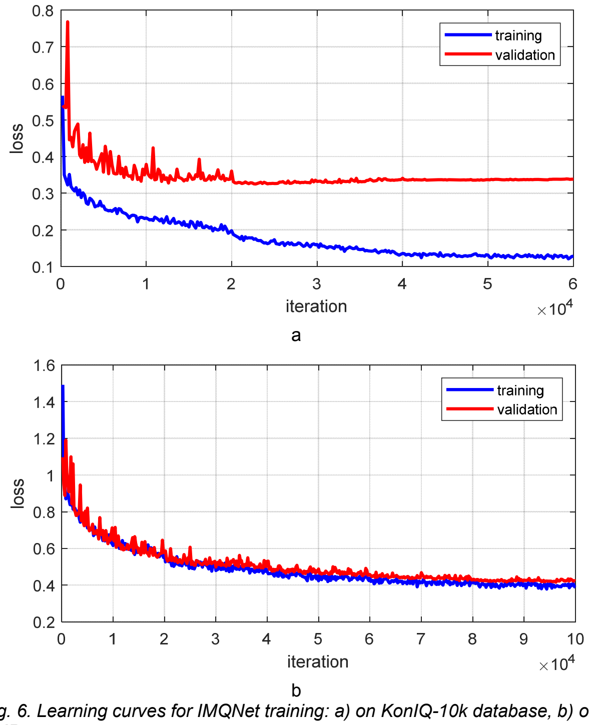
The final part of IMQNet concatenates 7x7 activations of both branches, applies Space-to-Depth layer and analyzes image scene composition in three fully connected layers.

IMQNet training

The training was performed in Matlab's environment in Matlab R2021a. We trained IMQNet on TMP set to predict PA values using 100000 iterations and minibatch size 10. We started from "initialLearnRate" 0.0001 and decreased it 2 times each 20000 iterations.

For comparison and to show the overlearning effect, we also trained IMQNet directly on KonIQ-10k database to predict MOS.

Fig. 6 shows the training curves of IMQNet on KonIQ-10k database and on TMP set. It is clearly seen that for small KonIQ-10k database overlearning affects training almost immediately, while for large TMP set curves for the training and validation sets coincide even after 100000 iterations.



a

b

Fig. 6. Learning curves for IMQNet training: a) on KonIQ-10k database, b) on TMP set

SROCC between the trained IMQNet and PA is 0.89. It significantly surpasses SROCCs in the Table 2 and gives an evidence of suitability of the proposed IMQNet architecture for image quality prediction.

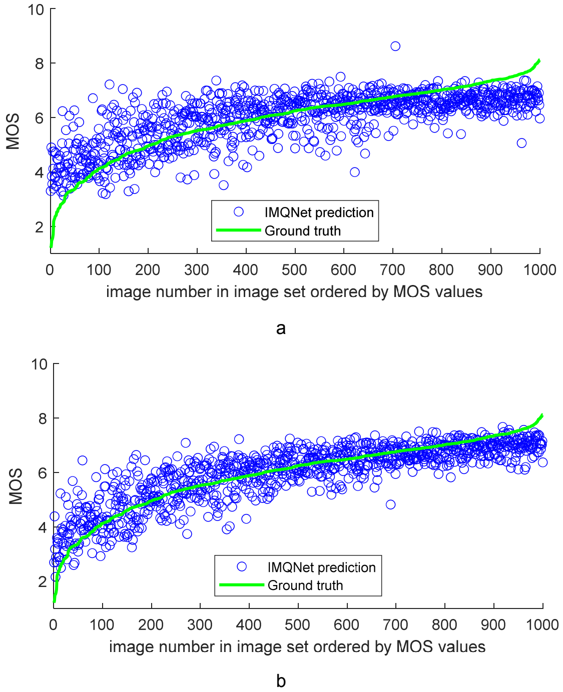
Comparative analysis on KonIQ-10k MOS

Table 3 contains SROCC values between trained IMQNet and MOS of KonIQ-10k. It is clearly seen that the use of the proposed transfer learning algorithm increases SROCC from 0.73 to 0.86 in comparison to the direct learning using MOS of KonIQ-10k database.

Table 3. SROCC between IMQNet and MOS of KonIQ-10k database

IMQNet, trained using KonIQ-10k MOS	0.73
IMQNet, trained using proposed transfer learning algorithm and TMP set	0.86

Fig. 7. Draws scatterplots of IMQNet predictions for both trained models.



a

b

Fig. 7. Scatterplots for predictions of MOS of KonIQ-10k: a) IMQNet, trained to predict MOS of KonIQ-10k, b) IMQNet trained to predict PA of TMP set

It is clearly seen that training using the proposed transfer learning algorithm provides more stable and precise predictions.

Table 4 contains SROCC values for IMQNet trained on TMP set and K5. Here K5 is KonCept512 metric pretrained on the merged MOS of six databases except database used for testing. For example, for testing on KonIQ-10k we used K5 pretrained on merged MOS of FLIVE, WILD, NRTID, HTID and SPAQ.

Table 4. SROCC between metrics values and MOS of image databases

Metric	KonIQ-10k	FLIVE	WILD	NRTID	HTID	SPAQ	Average
K5	0.797	**0.469**	**0.773**	**0.760**	0.612	**0.861**	0.71
IMQNet	**0.860**	0.407	0.699	0.691	**0.732**	0.826	0.70

One can see that the proposed algorithm of transfer learning provides very efficient and stable knowledge transfer between networks of totally different architectures.

Conclusions

A new and efficient algorithm of transfer learning between NR-IQA metrics of different architectures is proposed. Due to the absence of overlearning, the proposed algorithm allows to significantly increase a quality of training in comparison to a direct learning.

A new efficient network architecture IMQNet for estimation of image visual quality is proposed. The network can analyze both low-level and high-level image features, as well as image composition.

IMQNet is pretrained on TMP, details of IMQNet architecture and demo scripts are available in http://ponomarenko.info/imqnet.

Acknowledgments

This research was done with the financial support of Huawei-Tampere University project 3114100158, FlexISP.

References

[1] N. Ponomarenko, O. Eremeev, K. Egiazarian, V. Lukin, “Statistical evaluation of no-reference image visual quality metrics”, in Proceedings of EUVIP, Paris, France, 5p, 2010.

[2] V. Hosu, H., Lin, T., Sziranyi, D. Saupe, “KonIQ-10k: An ecologically valid database for deep learning of blind image quality assessment”, IEEE Transactions on Image Processing, 29, pp. 4041-4056, 2020.

[3] Z. Ying, H. Niu, P. Gupta, D. Mahajan, D. Ghadiyaram, A. Bovik, “From patches to pictures (PaQ-2-PiQ): Mapping the perceptual space of picture quality”, In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3575-3585, 2020.

[4] Y. Fang, H. Zhu, Y. Zeng, K. Ma, and Z. Wang, “Perceptual quality assessment of smartphone photography”, In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 3677-3686.

[5] M. Ponomarenko, S. Ghanbaralizadeh Bahnemiri, K. Egiazarian, O. Ieremeiev, V. Lukin, V. Peltoketo, J. Hakala, "Color image database HTID for verification of no-reference metrics: peculiarities and preliminary results", EUVIP, 2021, 6p.

[6] D. Ghadiyaram, A. Bovik, “Massive online crowdsourced study of subjective and objective picture quality”, IEEE Transactions on Image Processing, 25(1), 2015, pp. 372-387.

[7] Z. Dai, H. Liu, Q. Le, M. Tan, “CoAtNet: Marrying Convolution and Attention for All Data Sizes”, arXiv preprint arXiv:2106.04803, 2021.

[8] Z. Liu, H. Hu, Y. Lin, Z. Yao, Z. Xie, Y. Wei, J. Ning, Y. Cao, Z. Zhang, L. Dong, “Swin Transformer V2: Scaling Up Capacity and Resolution”, arXiv preprint arXiv:2111.09883, 2021.

[9] K. Zhang, Y. Li, W. Zuo, L. Zhang, L.Van Gool, R. & Timofte, “Plug-and-play image restoration with deep denoiser prior”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 17 p, 2021.

[10] L. Huang, Y. Xia, “Joint blur kernel estimation and CNN for blind image restoration”, Neurocomputing, pp. 324-345, 2020.

[11] A. Kaipio, M. Ponomarenko, K. Egiazarian, "Merging of MOS of large image databases for no-reference image visual quality assessment", MMSP, 2020, 6 p.

[12] C. Szegedy, S. Ioffe, V. Vanhoucke, A. Alemi, "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning." In AAAI, vol. 4, p. 12. 2017.

[13] H. Otroshi-Shahreza, A. Amini, H. Behroozi, "No-Reference Image Quality Assessment using Transfer Learning," in 9th IEEE International Symposium on Telecommunications, 2018.

[14] T. Lu, A. Dooms, "Towards Content Independent No-reference Image Quality Assessment Using Deep Learning," in IEEE 4th International Conference on Image, Vision and Computing, 2019.

[15] P. Vu, D. Chandler, "A fast wavelet-based algorithm for global and local image sharpness estimation", IEEE Signal Processing Letters, pp. 423-426, 2012.

[16] N. Ponomarenko, V. Lukin, O. Eremeev, K. Egiazarian, J. Astola, "Sharpness metric for no-reference image visual quality assessment", Image Processing: Algorithms and Systems X and Parallel Processing for Imaging Applications II. International Society for Optics and Photonics, vol. 8295, 11 p, 2012.

[17] Y. Zhang, A. Moorthy, D. Chandler, A. Bovik, "C-DIIVINE: No-reference image quality assessment based on local magnitude and phase statistics of natural scenes", Signal Processing: Image Communication, vol. 29.7, pp. 725-747, 2014.

[18] L. Zhang, L. Zhang, and A. C. Bovik, “A feature-enriched completely blind image quality evaluator,” IEEE Trans. Image Process., vol. 24, no. 8, pp. 2579-2591, 2015.

[19] W. Zhang, K. Ma, J. Yan, D. Deng, Z. Wang, "Blind image quality assessment using a deep bilinear convolutional neural network", IEEE Transactions on Circuits and Systems for Video Technology, 2018.

[20] Kuznetsova, A., Rom, H., Alldrin, N., Uijlings, J., Krasin, I., Pont-Tuset, J., Kamali, S., Popov, S., Malloci, M., Kolesnikov, A., Duerig, T. & Ferrari, V. The Open Images Dataset V4: Unified image classification, object detection, and visual relationship detection at scale. International Journal of Computer Vision, pp. 1956-1981, 2020.