

Efficiency Increasing of No-Reference Image Quality Assessment in UAV Applications

Oleg Ieremeiev^a, Vladimir Lukin^a, Krzysztof Okarma^b and Karen Egiazarian^c

^a National Aerospace University, Chkalova 17, Kharkiv, 61070, Ukraine

^b West Pomeranian University of Technology in Szczecin, al. Piastów 17, Szczecin, 70-310, Poland

^c Tampere University of Technology, Kalevantie 4, Tampere, FIN 33101, Finland

Abstract

Unmanned aerial vehicle (UAV) imaging is a dynamically developing field, where the effectiveness of imaging applications highly depends on quality of the acquired images. No-reference image quality assessment is widely used for quality control and image processing management. However, there is a lack of accuracy and adequacy of existing quality metrics for human visual perception. In this paper, we demonstrate that this problem persists for typical applications of UAV images. We present a methodology to improve the efficiency of visual quality assessment by existing metrics for images obtained from UAVs, and introduce a method of combining quality metrics with the optimal selection of the elementary metrics used in this combination. A combined metric is designed based on a neural network trained to utilize subjective assessments of visual quality. The metric was tested using the TID2013 image database and a set of real UAV images with embedded distortions. Verification results have demonstrated the robustness and accuracy of the proposed metric.

Keywords

image quality assessment, no-reference metric, visual quality, UAV images, correlation analysis, artificial neural network

1. Introduction

A scope of applications of drones and other unmanned aerial vehicles (UAVs) has expanded rapidly in recent few decades. Since most of UAVs contain cameras, there is a growing interest in analysis and processing of visual data. UAVs mainly use optical band cameras, thus, the existing digital image processing solutions are applicable [1-2]. However, the mobility and autonomy of these systems can impose significant restrictions and one must consider all these factors.

The key problems of applying digital image processing in UAV applications are as follows:

1. UAVs require an adaptive integrated approach to suppress the present noise, motion blur, and other typical distortions, which can only partially be compensated by a camera stabilization.
2. For data transmission from drones, a wireless connection is used. The range and reliability of data transmission determine one of the key characteristics of UAVs – the flight range. In this sense, the efficiency of processing and compression of high-resolution data for transmission over a radio channel with limited bandwidth is decisive.
3. The data obtained at the end device, in addition to storage and more complex post-processing, can be used in various applications. Among them, high level vision tasks based on machine learning, such as detection, recognition, etc., become more widespread [1, 3, 4].

Common to all above mentioned challenges, there is a need to accurately assess image quality and measure distortion parameters, which will be used in image reconstruction methods, to enhance the

The Sixth International Workshop on Computer Modeling and Intelligent Systems (CMIS-2023), May 3, 2023, Zaporizhzhia, Ukraine

EMAIL: o.ieremeiev@khai.edu (O. Ieremeiev); v.lukin@khai.edu (V. Lukin); okarma@zut.edu.pl (K. Okarma);

karen.egiazarian@tuni.fi (K. Egiazarian)

ORCID: 0000-0001-7865-0570 (O. Ieremeiev); 0000-0002-1443-9685 (V. Lukin); 0000-0002-6721-3241 (K. Okarma);

0000-0002-8135-1085 (K. Egiazarian)



© 2023 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
CEUR Workshop Proceedings (CEUR-WS.org) Proceedings

visual quality of the acquired images. In addition, an effective lossy compression is required. Certain results of UAV image processing have already been reported [5, 6, 7]. Nevertheless, robust methods that can accurately assess the visual component and determine the optimal parameters for subsequent image processing methods are required.

Image quality assessment (IQA) is usually applied by visual quality metrics. To improve their accuracy, some features of human perception are employed. There are two main classes of visual quality assessment methods. Full-reference (FR) visual quality metrics are widely used to verify image processing methods by evaluating the relative changes in image quality. No-reference (NR) metrics assess the quality based on the characteristics of the image itself and can be applied as a tool in many UAV applications [8, 9].

There are many developed NR IQA methods, but their common problem is a low accuracy, due to only limited amount of information available for analysis, and these metrics inability to accurately separate image elements (textures, borders, gradients, etc.) from distortions (noise, blur, etc.) [10, 11].

To design and verify visual quality metrics, special test image databases [11] are used. They contain images distorted by certain types of distortions. For each image, a visual quality score (mean opinion score (MOS)) is formed based on the results of a large number of subjective experiments with volunteers. Correlation analysis between metric values and MOS serves as a quantitative indicator of its compliance with human vision. Considering the most universal and large test image databases with tens of distortion types such as TID2013 [12], the efficiency of no-reference metrics usually does not exceed 0.5, according to the Spearman rank order correlation coefficient (SROCC).

Fortunately, one can increase accuracy of IQA using existing methods through their joint use, e.g., using methods presented in [13, 14]. In this paper, we propose a method of combining no-reference visual quality metrics based on an artificial neural network (ANN) that is focused on solving various problems of processing UAV images. Since many tasks with UAVs require the mobility of computing devices, the priority of this work is to ensure high accuracy of visual quality estimation while maintaining acceptable performance.

2. The efficiency of metrics for UAV purposes

Drones can lead to a significant amount of various distortions for an image during its acquisition, processing, compression and transmission over a communication channel. In this regard, the design of a combined metric requires the presence of test image databases that allow simulating such situations. As a result of the analysis of many image databases [11], we have chosen TID2013.

A distinctive feature of this image database is that it contains 24 types of various distortions, including such unique ones as bit errors in the transmission of compressed images. TID2013 contains 25 reference images that have been distorted by 24 types of distortion at 5 levels of intensity, for a total of 3000 test images. A complete list of distortions and their applicability to solving the current problem is given in Table 1.

Let us analyze the distortions listed in Table 1 and their relation to imaging from UAVs:

- Additive Gaussian noise (##1-2) is the basic model for representing most of the physical processes that cause noise. It is more pronounced in low light conditions.
- Spatially correlated noise (#3) is a characteristic of optical images due to the use of the Bayer filter or its modifications on sensors. It significantly increases with digital zoom.
- Impulse noise (#6) may be a manifestation of dead pixels and a lot of other causes such as coding/decoding artifacts.
- Quantization noise (#7) may occur during image acquisition and transformations.
- Blurring (#8) is one of the most relevant distortions due to the motion and vibrations of the UAV.
- Denoising (#9) is a manifestation of the noise reduction built into most cameras.
- Compression (##10-11) is a typical stage in the image processing chain to reduce data redundancy.
- Transmission errors (##12-13) are typical for wireless communication channels, especially over long distances.

- Changes in brightness, contrast and saturation (## 16-18) allow simulating changes in lighting conditions at different time instances of a day and weather conditions.
- Multiplicative noise (#19) is relevant because sensor noise is mostly signal-dependent.
- Noise (#20) allows the simulation of some artifacts of image processing and compression.
- Lossy compression of noisy images (#21) is a typical example of a real situation where an image with some noise is compressed.
- Chromatic aberration (#23) is a result of the refraction of light in the camera's optics.

Table 1

List of TID2013 distortions and their relevance for UAV purposes

##	Distortion type	Relevance for UAV imaging
1	Additive Gaussian noise	+
2	Additive noise (more intensive in color components)	+
3	Spatially correlated noise	+
4	Masked noise	-
5	High-frequency noise	-
6	Impulse noise	+
7	Quantization noise	+
8	Gaussian blur	+
9	Image denoising	+
10	JPEG compression	+
11	JPEG2000 compression	+
12	JPEG transmission errors	+
13	JPEG2000 transmission errors	+
14	Non-eccentricity pattern noise	-
15	Local block-wise distortions of different intensity	-
16	Mean shift (intensity shift)	+
17	Contrast change	+
18	Change of color saturation	+
19	Multiplicative Gaussian noise	+
20	Comfort noise	+
21	Lossy compression of noisy images	+
22	Image color quantization with dither	-
23	Chromatic aberrations	+
24	Sparse sampling	-

The listed 18 distortions comprehensively allow a use of the vast majority of noise types and distortions that can occur in UAV images or be the result of weather conditions. These distortion types give together 2250 test images from the TID2013 dataset that will be used in the paper.

Let us analyze the performance of the existing NR metrics on this subset of images. Since our task is to ensure high accuracy of estimation, the maximum possible number of different metrics is included. The SROCC values for the entire TID2013 database and the selected subset are given in Table. 2.

As it can be seen from the results in Table 2, the best performance is demonstrated by the ILNIQE metric, but its SROCC values (equal to 0.492 for all and 0.529 for the selected 18 UAV distortions) are relatively (inappropriately) low. It should be noted that Table 2 shows the absolute SROCC values because the metrics have been developed using different image databases that can evaluate the visual quality (MOS values) in two ways: as a higher value for better quality, or vice versa - a higher value as a larger difference from the perfect quality.

Table 2

SROCC values of no-reference IQA on TID2013 subsets

##	Metric	SROCC (All)	SROCC (UAV)	##	Metric	SROCC (All)	SROCC (UAV)
1	ILNIQE [15]	0.492	0.529	23	DIQU [34]	0.240	0.251
2	CORNIA [16]	0.435	0.521	24	SDQI [35]	0.224	0.248
3	HOSA [17]	0.471	0.515	25	DIPIQ [36]	0.140	0.209
4	C-DIIVINE[18]	0.373	0.448	26	MLV [37]	0.201	0.195
5	BLIINDS2 [19]	0.395	0.425	27	FISHBB [38]	0.145	0.152
6	BRISQUE[20]	0.367	0.416	28	JNBM [39]	0.141	0.152
7	BIQI [21]	0.405	0.409	29	DESIQUE[40]	0.069	0.150
8	SSEQ [22]	0.341	0.406	30	GMLOG [41]	0.109	0.139
9	NIQE [23]	0.313	0.403	31	NIQMC [42]	0.113	0.124
10	QAC [24]	0.372	0.379	32	ARISM [43]	0.145	0.109
11	SISBLIM_SM [25]	0.318	0.360	33	CPBDM [44]	0.112	0.109
12	LPSI [26]	0.395	0.357	34	LSSn [31]	0.168	0.105
13	LPC-SI [27]	0.323	0.354	35	PSS [31]	0.022	0.087
14	SISBLIM_SFB[25]	0.336	0.348	36	LSSs [31]	0.114	0.084
15	DIIVINE [28]	0.344	0.343	37	ARISM _c [43]	0.138	0.081
16	BIBLE [29]	0.281	0.333	38	PSI [45]	0.001	0.075
17	OG-IQA [30]	0.276	0.327	39	SMETRIC[46]	0.097	0.074
18	BPRI [31]	0.229	0.313	40	FISH [38]	0.052	0.041
19	TCLT [32]	0.233	0.308	41	NR-PWN [47]	0.016	0.039
20	SISBLIM_WFB [25]	0.293	0.301	42	NMC [48]	0.054	0.033
21	MSGF-PR [33]	0.244	0.274	43	BLUR [49]	0.008	0.020
22	SISBLIM_WM [25]	0.239	0.265	44	NJQA [50]	0.100	0.007

3. The problem of metrics selection

It is possible to increase the accuracy of image quality assessing by combining several metrics. Successfully selected metrics are able to complement each other and provide a comprehensive analysis of the image taking into account various types of distortions. As it was shown in [13], the greatest efficiency is achieved through multi-parameter optimization using artificial neural networks. Combining the listed 44 metrics can potentially give the best accuracy of visual quality assessment. However, most of these metrics can make a low contribution requiring significant computing resources. High mobility and minimal computing costs are among the key requirements for UAV applications. Therefore, it is necessary to reduce the number of metrics without a significant decrease in the accuracy of IQA. Several possible solutions can be employed for the correct choice of elementary metrics (listed in Table 2) as inputs of an ANN, but not all of them are feasible or give an effective solution:

1. A complete enumeration of options is not possible in practice, since even for 5 or 10 incoming metrics, it will be necessary to calculate 1.6×10^8 and 2.7×10^{16} combinations, respectively.
2. The choice of the best metrics with high SROCC rates or the exclusion of similar metrics with high cross-correlation values has shown insufficient efficiency in [51].
3. “Intelligent” selection of appropriate metrics. As a possible solution, the approach of using regularization was tested in [13] and proven to be effective. Lasso (least absolute shrinkage and selection operator) regularization is widely used in machine learning to reduce the model complexity and prevent overfitting. As a result of introducing restrictions, it allows determining the least important input features (corresponding metrics) and excludes them by setting zero weight coefficients. This approach can be applied to reduce the number of metrics.

To display the influence of the number of elementary metrics used on the accuracy of the trained ANN, we employ several of their combinations defined using Lasso in the range of values from the minimum 3-5 to all 44 metrics. The Lasso parameters were selected in such a way as to obtain non-zero weights for a given number of the metrics. Totally, 10 dimensions are considered in the paper: 4, 5, 7, 10, 16, 20, 25, 30, 35, and 44. Metric combinations with 16 metrics and less, which are focused on, are presented in Table 3.

Table 3
List of the metrics, defined by Lasso

Metrics' number	Metrics' names
4	ARISM, CORNIA, DIPIQ, ILNIQE
5	Above 4 + LPCSI
7	Above 5 + MLV, NIQMC
10	Above 7 + MSGF-PR, NIQE, PSS
16	Above 10 + C-DIIVINE, GMLOG, HOSA, JNBM, PSI, TCLT

4. Preliminary results

Despite the popularity of neural networks, their use in the field of image quality assessment has some limitations.

First, there are limited variety and size of datasets, because only image databases containing MOS values can be applied. It should be noted that due to the limited number of distortion levels and the variety of reference images, it can be assumed that some test images have unique properties and their distribution into training or test subset may affect the accuracy of the trained neural networks. Therefore, it is impossible to choose exactly which images should be in each of these sets. To ensure a result approaches the optimal one, for each ANN configuration over 100 repetitions with a random distribution of images on training (70%) and testing (30%, respectively) subsets have been completed.

Second, the choice of the type of ANN can have a significant impact on the final efficiency. Two types of networks are considered: feed-forward and cascade networks, which have a non-linear relationship between layers since the resulting value of each layer, including the input one, affects all subsequent layers.

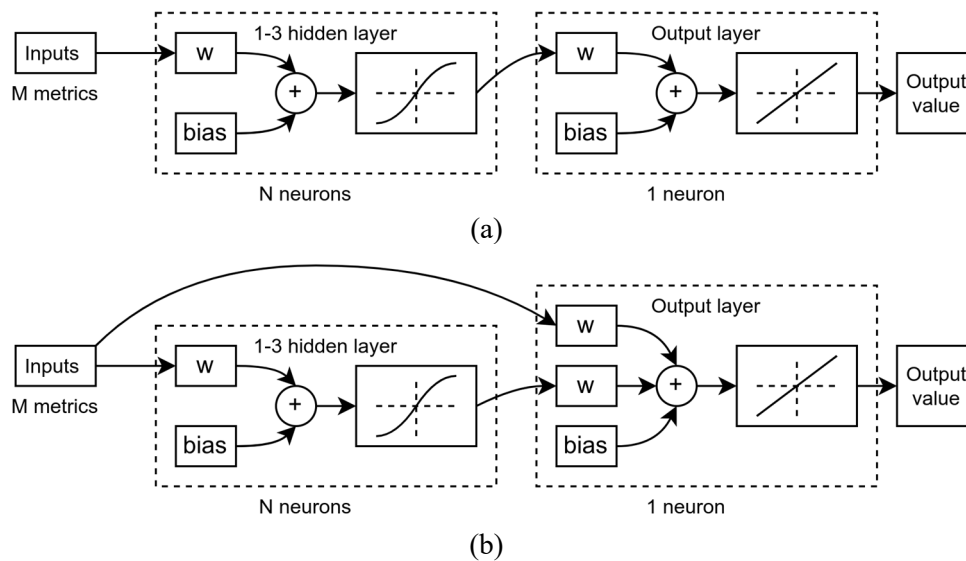


Figure 1: Generalized schemes of the used feed-forward (a) and cascade (b) networks

Further, the efficiency of ANN is also determined by its structure (the number of hidden layers and the number of neurons in each of them). Since a significant number of factors affecting the efficiency

of the final neural network have already been indicated, several basic configurations are used at the preliminary stage of the analysis. A more precise configuration of the ANN will be determined at the final stage of creating the combined metric. At this stage, variants of the neural network structure with 1-3 hidden layers are used. For each of them, there are two options for the number of neurons N in each layer: 1) in all layers, it is equal to the number of input metrics M ($N = M$), and 2) each next layer starting from the second one the number is divided by two ($N_1 = M, N_2 = M/2, N_3 = M/4$). There are only 5 options totally because for a single-layer network they are identical.

As the activation function, a sigmoid function is used, which allows, regardless of the value ranges of the used metrics, to obtain, after the 1st hidden layer, the values in the fixed range $[0,1]$. This procedure allows us to implement the built-in function fitting and value normalization. This stage involves the construction of 10,000 variants of ANN (2 types \times 5 configs \times 100 repetitions \times 10 metric combinations). All calculations were performed using the MatLab software.

Let us analyze the results obtained after training all these ANNs. The main dependence is that the accuracy of the combined metric grows with the number of elementary metrics used. The maximum is achieved for all 44 metrics. The graph of SROCC dependence on the number of metrics is shown in Fig. 2 for ANNs with maximum SROCC rates among repetitions of each configuration for the feed-forward network.

Based on these results, several conclusions can be drawn. Thus, the use of an ANN for metrics combination is an effective solution for UAV applications, since even the minimal number of them (4) significantly exceeds in accuracy the maximum result among elementary metrics (SROCC = 0.53). The current 5 configurations of the ANN structures give similar indicators, their comparison will be carried out in more detail later. This graph allows making some recommendations for choosing the structure of an ANN depending on the requirements and constraints of the problem solved. For example, if it is necessary to ensure maximum performance, the desired choice would be a combined metric of 5 elementary ones, its result reaches SROCC = 0.74, which is much higher than for 4 metrics, but a further increase of accuracy with the number of input parameters is slow. Nevertheless, if accuracy or balance with performance is a priority, then the options of 10 or 16 elementary metrics can be useful. Their accuracy reaches 0.82 – 0.84 of SROCC. Further, the accuracy at the level of 0.85 is practically independent of the number of metrics. Considering that one of the requirements of this study is to maintain acceptable performance with high accuracy, we will use a combined metric consisting of 10 elementary metrics.

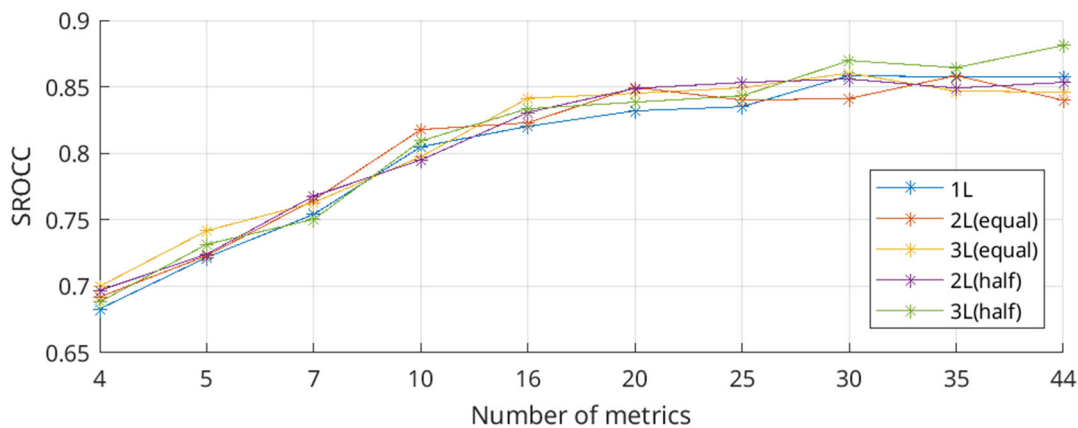


Figure 2: Dependence of SROCC on the number of elementary metrics selected by the Lasso criterion

To display the main statistical indicators and some problems, Fig. 3 shows a box chart for 4 (full graph and limited range higher than 0.5 under it), 5 (similarly to the previous one), 10, and all 44 metrics. Its advantage is the ability to display simultaneously the median, the lower (0.25) and upper (0.75) quartiles, any outliers (computed using the interquartile range), and the minimum and maximum values that are not outliers. From these graphs, it can be noted that with a small number of metrics (4 and 5), the complexity of the neural network (number of neurons) is not enough for proper training, as a result of which anomalous results were obtained - incorrectly trained neural networks with indicators below individual metrics. This is also the problem of multilayer neural networks with

fewer neurons in each layer. For 10 and more metrics, this problem is no longer observed. The highest values for each presented network configuration are already denoted in Fig 1. Quantitative indicators of the best neural networks for the feed-forward network from Fig. 1 and Fig. 2 are given in Table 4, where M means the number of elementary metrics.

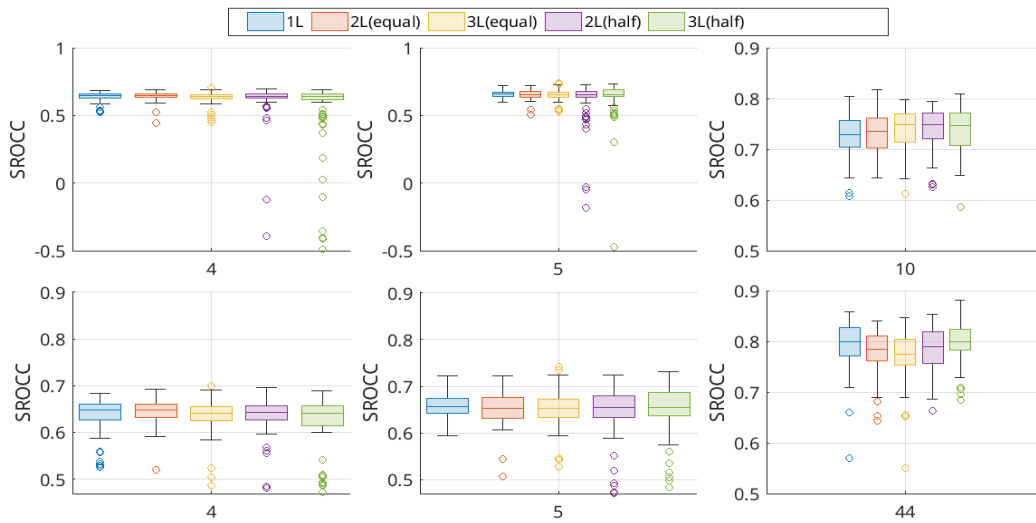


Figure 3: Box charts of the results of the obtained neural networks for 4, 5, 10, and 44 input metrics.

Table 4

Results of the best feed-forward networks for different numbers of inputs (4, 5, 10, and 44)

NN config	Description (in NN layers)	SROCC			
		M = 4	M = 5	M = 10	M = 44
1	[M]	0.683	0.722	0.805	0.858
2	[M, M]	0.692	0.723	0.818	0.840
3	[M, M, M]	0.700	0.742	0.797	0.846
4	[M, M/2]	0.697	0.724	0.795	0.853
5	[M, M/2, M/4]	0.688	0.732	0.809	0.881

5. Final network modifications

In the first phase of experiments, when forming a neural network for 10 input metrics, the following configurations were used for neural networks with 1-3 hidden layers: [10], [10, 10], [10, 10, 10], [10, 5] and [10, 5, 2].

The general trend in Fig. 2 shows that the number of neurons in layers less than 10 may not be enough. Therefore, additional configurations with a number of neurons up to 20 per layer ($\times 2$ compared to the number of input metrics) were additionally built. More than 30 configurations for both network types have been used and the best results of the ANN for each number of hidden layers and some statistics are partly shown in Table 5. It shows lists of neural network configurations, both the best 2 from the initial five and additionally trained for 10 input metrics (50 repetitions).

To evaluate the effectiveness of each configuration and both types of networks, some statistical indicators are given: the maximum (best neural network) and minimum value, median, skewness, and quartiles 0.75 and 0.95. Skewness is a measure of the asymmetry of the data around the sample mean. If skewness is positive, the data spread out more to the higher values. The skewness of the normal distribution (or any perfectly symmetric distribution) is zero.

The maximum performance for both types of networks in Table 5 has been achieved by configuration #8. Despite the random learning process, in general, for a feed-forward network, an increase in the number of neurons to 20 leads to an increase in accuracy. This is also confirmed by the values of the quartiles 0.75 and 0.95. A further increase in the number of neurons does not provide a significant improvement. According to skewness values, it can be noted that there is a slight tendency

toward obtaining neural networks with low performance, and in the worst cases they differ a little from elementary metrics (SROCC can be less than 0.6). Cascade neural networks do not provide any advantages demonstrating somewhat lower performance for almost all configurations. This network shows the advantage in terms of maximum SROCC for configurations with a small number of neurons (#2 and #4), therefore, it is presumably the most effective for solutions with a small amount of input data and simpler layer structures.

According to the results of Table 5, the ANN with the maximum Spearman correlation coefficient of 0.8307 was chosen as a combined metric for visual quality assessment tasks. The list of metrics used in it and a visual comparison of its effectiveness for elementary metrics is shown in Fig. 4. This metric is available at <https://github.com/OlegIeremeiev/CNNM-UAV.git>.

Table 5
Results of the best feed-forward networks for 10 inputs

NN config	Description (in NN layers)	SROCC					
		Max	Min	Median	Skewness	0.75 Quartile	0.95 Quartile
Feed-forward network							
1	[10] [10]	0.8178	0.6433	0.7351	-0.0261	0.7621	0.8033
2	[10] [5]	0.7949	0.6256	0.7487	-0.9202	0.7716	0.7899
3	[20]	0.8111	0.6475	0.7483	-0.5799	0.7704	0.8041
4	[20] [10]	0.8074	0.6312	0.7563	-0.7113	0.7862	0.8024
5	[20] [20]	0.8280	0.6787	0.7625	-0.2461	0.7906	0.8209
6	[15] [10] [10]	0.8214	0.6430	0.7452	-0.3190	0.7726	0.8116
7	[10] [15] [20]	0.8050	0.6383	0.7386	-0.3225	0.7608	0.8009
8	[20] [15] [10]	0.8307	0.6500	0.7607	-0.5689	0.7806	0.8161
9	[10] [20] [15]	0.8195	0.5945	0.7387	-0.6796	0.763	0.8120
Cascade network							
1	[10] [10]	0.8010	0.6074	0.7507	-1.2935	0.7698	0.7902
2	[10] [5]	0.8213	0.5437	0.7435	-1.6415	0.7643	0.7844
3	[20]	0.7989	0.5954	0.7521	-1.0453	0.7685	0.7966
4	[20] [10]	0.8176	0.5606	0.7582	-1.5616	0.7811	0.8098
5	[20] [20]	0.8080	0.6108	0.7577	-1.5696	0.7784	0.8000
6	[15] [10] [10]	0.8126	0.6253	0.7487	-0.9623	0.7739	0.8040
7	[10] [15] [20]	0.8185	0.6326	0.7618	-0.9489	0.7821	0.8041
8	[20] [15] [10]	0.8214	0.6193	0.7726	-1.3170	0.7895	0.8053
9	[10] [20] [15]	0.8177	0.6681	0.7572	-0.2515	0.7797	0.8122

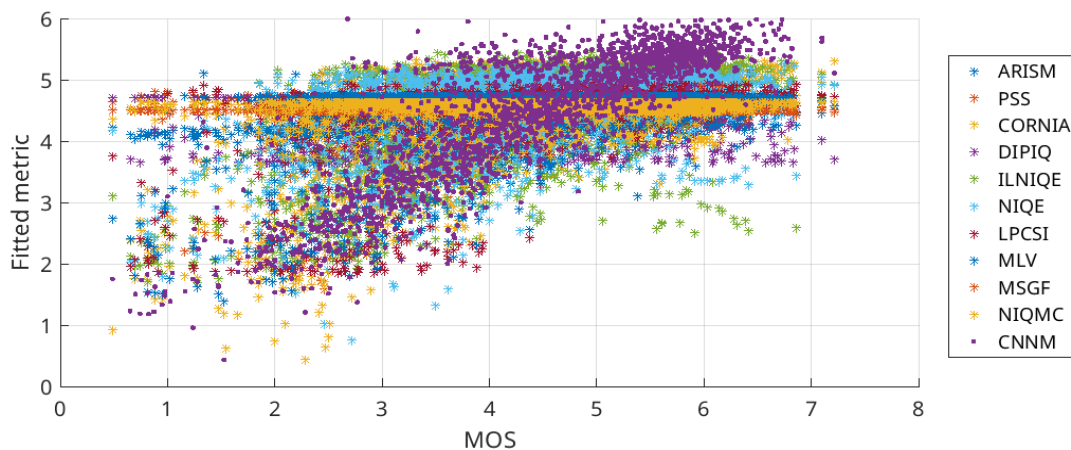


Figure 4: Scatter plot of the 10 elementary metrics and the combined metric (CNNM) including them.

A visual representation of the effectiveness of assessing the quality of certain types of distortions is shown in the graph in Fig. 5. The numbers of distortions correspond to the serial number of the distortions selected for analysis (see Table 1). It can be seen that the combined metric provides consistently high results with a decrease in accuracy at distortions #12 (mean shift) and #14 (change of color saturation), these distortions are problematic for all the metrics used in the paper.

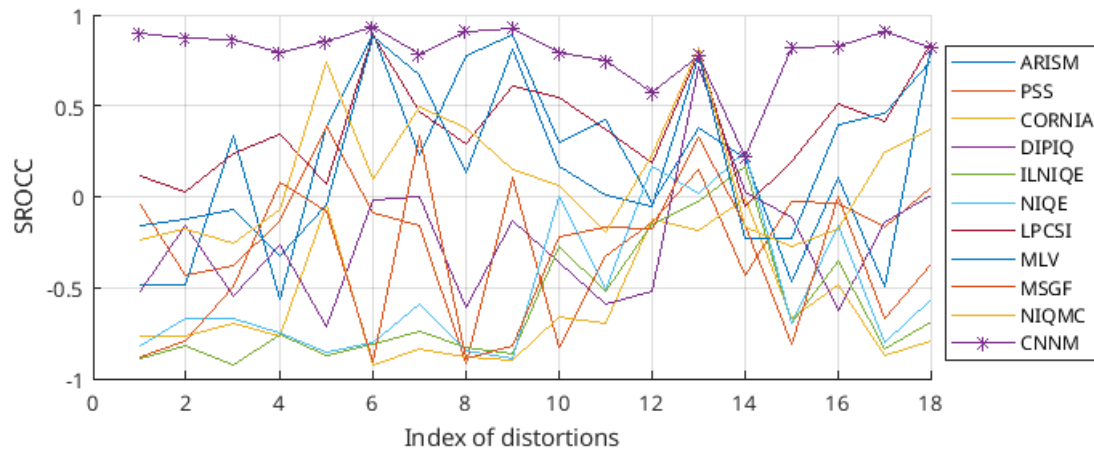


Figure 5: Dependency of the metrics' SROCC values on the type of distortion (absolute values)

6. Combined metric analysis

The purpose of creating a combined metric was to improve the accuracy of the visual quality assessment of images in various UAV tasks. However, there is a limitation: general-purpose color image database TID2013 with the corresponding MOS values was taken to train the neural network. Therefore, it is necessary to analyze the effectiveness of the obtained metric in practice for real images.

It should be noted that the application area and its inherent types of distortion significantly affect the results obtained. Thus, in [14], the visual quality metric was proposed for the assessment of remote sensing images. Its SROCC value reached the level of 0.8813. At the same time, verification on UAV-related distortions from TID2013 showed significantly worse results - SROCC has decreased to 0.7083. The reason lies in the different sets of distortions. In particular, transmission errors are rare in remote sensing practice, since these systems operate in more static and predictable conditions. Distortions in brightness and contrast as a factor of weather and daylight conditions changing were also not taken into account in the design in [14]. This confirms the fact that individual metrics are often not enough for application areas with unique features and the combined approach based on neural networks allows for an increase of 50% or more.

The practicality and applicability of the proposed solution can only be assessed on the basis of real images from the UAV. At the same time, this approach has significant limitations: the absence of MOS values and the complexity of obtaining images with all the considered distortions and needed combinations. Taking this into account, a number of assumptions and simplifications have been made, and the results obtained are mostly illustrative.

1. Verification of visual metrics requires MOS, which values can only be obtained from a significant amount of subjective experiments and require considerable time. The first simplification is that the missing MOS values can to some extent be replaced by objective indicators, the accuracy of which significantly exceeds the analyzed metrics. For a comparative analysis of the combined and individual metrics, this may be sufficient. Such a condition can be provided by full-reference quality metrics - the accuracy of some of them reaches $SROCC = 0.9$ for the entire TID2013 and more than 0.96 for certain types of distortions and significantly exceeds SROCC for existing no-reference metrics.

2. It is technically difficult to ensure the presence of real test images with the considered distortions, therefore, it is proposed to artificially simulate their presence by adding the distortions under the interest of different intensities to the selected images.

3. The level of distortion should preferably have a wide range of intensities from inconspicuous to significant.

To verify the metrics, real images from UAVs were used. As a basis, some images of the UAVDT (Unmanned Aerial Vehicle Benchmark Object Detection and Tracking) dataset [52] were taken, examples of which are shown in Fig. 6. The dataset contains more than 40,000 images with a resolution of 1080×540 pixels. Of these, 16 images were selected with different terrain, daylight, and weather conditions.



Figure 6: Examples of reference images of the UAV test set.

Creation of test images with the necessary types of distortion requires special skills. TID2013 distortions were generated in accordance with a certain strategy, however, their generation code is not available. Therefore, our mechanisms for generating distortions are used in the paper, and from the list of selected types of distortions, 9 main ones are taken into account:

- Gaussian white noise;
- Multiplicative noise;
- Gaussian blur;
- Denoising (applying BM3D filter to images with Gaussian white noise);
- JPEG and JPEG2000 compression;
- Brightening, darkening, and mean shift (darkening and lightening).

According to the variety of intensities, 9 different levels were chosen for a more accurate gradation of distortion, in contrast to 5 levels for TID2013. Their intensity varies from inconspicuous to significant. The distribution of peak signal-to-noise ratio (PSNR) values is shown in Fig. 7.

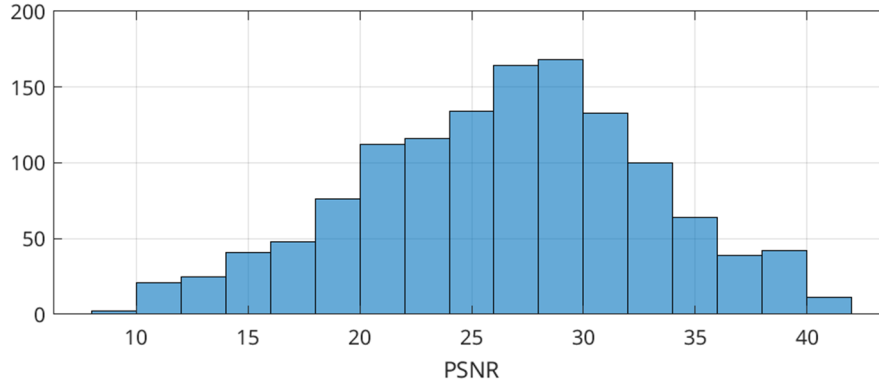


Figure 7: Histogram of PSNR values of the test images

As a result, the verification test set based on real UAV images consists of 1296 images (16 images x 9 distortions x 9 intensity levels).

In the role of MOS values for no-reference metrics verification, the best full-reference quality metrics are used. The SROCC values of some well-known FR IQA for all TID2013 images and UAV-related test set are given in Table 6. Since their problems and solutions are similar to those solved in the article, a combined full-reference metric was formed to improve the accuracy. It uses the metrics listed in Table 6 as input and consists of a two-layer neural network (marked as C_MOS) with the number of neurons [16, 8] and all other parameters listed above. Since its SROCC for the task considered is almost 0.04 higher than for the best of elementary metrics, this combined metric has been chosen as the analog of MOS for UAV test images.

Table 6

SROCC values of the full-reference visual metrics on the TID2013 image dataset

Metric	VSI [53]	PSIM [54]	MDSI [55]	HaarPSI [56]	UNIQUE [57]	CVSSI [58]	IQM2 [59]	ADM [60]	C_MOS
SROCC	0.8967	0.8926	0.8897	0.8730	0.8599	0.8090	0.7955	0.7861	0.9107
SROC(UAV)	0.8274	0.8519	0.8873	0.8811	0.8496	0.8478	0.8507	0.8075	0.9261

The results of the verification of the combined and elementary no-reference metrics are shown in Table 7. In addition to the overall assessment, the SROCC values for individual types of distortions are also shown. The two best results for each type of distortion are highlighted in bold.

Table 7

The results of verification of the no-reference metrics on the UAV test set

Distortion	ARISM	PSS	CORNIA	DIPIQ	ILNIQE	NIQE	LPCSI	MLV	MSGF	NIQMC	CNNM
All	0.069	0.307	0.581	0.109	0.548	0.616	0.330	0.024	0.499	0.074	0.659
AWGN	0.804	0.202	0.861	0.470	0.890	0.886	0.511	0.239	0.794	0.180	0.874
Multiplicative noise	0.702	0.257	0.784	0.053	0.722	0.701	0.557	0.313	0.624	0.309	0.903
Blur	0.942	0.898	0.904	0.254	0.869	0.900	0.966	0.949	0.784	0.200	0.956
Denoise	0.360	0.072	0.186	0.196	0.004	0.153	0.020	0.181	0.228	0.118	0.154
JPEG	0.769	0.958	0.868	0.534	0.713	0.728	0.166	0.087	0.866	0.367	0.787
JP2k	0.134	0.378	0.738	0.337	0.240	0.632	0.054	0.486	0.028	0.433	0.764
Brighten	0.524	0.070	0.027	0.027	0.166	0.035	0.183	0.372	0.197	0.378	0.474
Darken	0.359	0.044	0.057	0.575	0.341	0.301	0.342	0.460	0.021	0.052	0.281
Mean shift	0.515	0.328	0.027	0.429	0.335	0.442	0.169	0.391	0.253	0.054	0.520

From the obtained results, it can be seen that despite the limitations of the approximate MOS values, the combined metric provides the maximum overall accuracy and is one of the best for most of the indicated types of distortions, providing the best balance between various distortions. It should be noted that these results have been obtained for the most common types of distortions, which are commonly used in the design of elementary metrics. Considering the types of distortions used in TID2013, but not modeled in this set (e.g. transmission errors, etc.), it can be expected that the combined metric can have additional benefits by providing more stable visual quality estimation.

7. Conclusions

The paper is devoted to visual quality assessment of UAV images, which is actual for automating the image processing and improving image quality for UAV applications. A list of more than 40 known no-reference visual quality metrics is considered. To analyze the effectiveness of visual quality metrics, the TID2013 image database and a subset with actual types of distortions have been selected. The verification of existing visual quality metrics has shown an accuracy of less than 0.53 for the best one and less than 0.3 for most metrics according to SROCC. Therefore, the method of combining visual quality metrics using the neural network has been proposed to improve the accuracy of visual quality assessment. The problem of the optimal choice of elementary metrics for reducing the redundancy and rational use of computing resources has been considered and the solution based on the Lasso regularization method has been proposed, which determines the weight coefficient equal to 0 for the excluded and least important metrics. Training the neural networks of different types and their configurations has been carried out, taking into account the limitations of the test image database used in experiments. The analysis of the effectiveness of this approach, which reaches a result of about 0.85 for 20 metrics or more, has been carried out, and the dependence on the number of metrics used in the paper together with the main statistics is shown. For 10 metrics, as the optimal solution for high accuracy and performance, the results have been refined with the training of additional configurations of the structure of neural networks. It is shown that the accuracy of the final combined metric reaches $SROCC = 0.83$.

To evaluate the effectiveness of the metric on real images, a test image database of almost 1300 images was formed. As an alternative to the missing MOS values, a combined full-reference metric has been created. Its accuracy reaches 0.926 for the used TID2013 distortion set and is significantly higher than the values of any no-reference metric, which is acceptable for their comparison. It is shown that, on this test set, the obtained metric provides the best result.

In the future, research in this area can be expanded by adding new distortions typical for UAV images and new neural network models including deep-learning models of limited complexity.

8. References

- [1] R. C. Gonzalez, R. E. Woods, Digital Image Processing, 4th ed., Pearson, New York, NY, 2018.
- [2] W. Burger, M.J. Burge, Principles of Digital Image Processing, Springer, New York, NY, 2009.
- [3] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning, 2nd ed., Springer, New York, NY, 2009. doi: 10.1007/978-0-387-84858-7.
- [4] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT Press, 2016.
- [5] R. Wang, X. Xiao, B. Guo, Q. Qin, R. Chen, An Effective Image Denoising Method for UAV Images via Improved Generative Adversarial Networks, *Sensors* 18 (2018) 1–23. doi: 10.3390/s18071985.
- [6] T. Sieberth, R. Wackrow, J. H. Chandler, UAV image blur - its influence and ways to correct it, *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XL-1/W4 (2015) 33–39*. doi: 10.5194/isprsarchives-XL-1-W4-33-2015.
- [7] V. S. Alfio, D. Costantino, M. Pepe, Influence of Image TIFF Format and JPEG Compression Level in the Accuracy of the 3D Model and Quality of the Orthophoto in UAV Photogrammetry, *J. Imaging* 6 (2020) 1–22. doi: 10.3390/jimaging6050030.

- [8] M. Kedzierski, D. Wierzbicki, Radiometric quality assessment of images acquired by UAV's in various lighting and weather conditions, *Measurement* 76 (2015) 156–169. doi: 10.1016/j.measurement.2015.08.003.
- [9] G. Koretsky, J. Nicoll, M. Taylor, A Tutorial on Electro-Optical/ Infrared (EO/IR) Theory and Systems, IDA Document D-4642, 2013.
- [10] W. Lin, C.-C. Jay Kuo, Perceptual Visual Quality Metrics: A Survey, *Journal of Visual Communication and Image Representation* 22 (2011) 297–312. doi: 10.1016/j.jvcir.2011.01.005.
- [11] Y. Niu, Y. Zhong, W. Guo, Y. Shi, P. Chen, 2D and 3D Image Quality Assessment: A Survey of Metrics and Challenges, *IEEE Access* 7 (2018) 782–801. doi: 10.1109/ACCESS.2018.2885818.
- [12] N. Ponomarenko, L. Jin, O. Ieremeiev, V. Lukin, K. Egiazarian, etc, Image database TID2013: Peculiarities, results and perspectives, *Signal Processing: Image Communication* 30 (2015), pp. 57–77. doi: 10.1016/j.image.2014.10.009.
- [13] O. Ieremeiev, V. Lukin, K. Okarma, K. Egiazarian, Full-Reference Quality Metric Based on Neural Network to Assess the Visual Quality of Remote Sensing Images, *Remote Sensing* 12 (2020) 1–31. doi: 10.3390/rs12152349.
- [14] A. Rubel, O. Ieremeiev, V. Lukin, J. Fastowicz, K. Okarma, Combined No-Reference Image Quality Metrics for Visual Quality Assessment Optimized for Remote Sensing Images, *Applied Sciences* 12 (2022) 1–19. doi: /10.3390/app12041986.
- [15] L. Zhang, L. Zhang, A.C. Bovik, A Feature-Enriched Completely Blind Image Quality Evaluator, *IEEE Trans. Image Processing* 24 (2015) 2579–2591. doi: 10.1109/TIP.2015.2426416
- [16] P. Ye, J. Kumar, L. Kang, D. Doermann, Unsupervised Feature Learning Framework for No-Reference Image Quality Assessment, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR '2012, Providence, RI, USA, 2012*, pp. 1098–1105. doi: 10.1109/CVPR.2012.6247789.
- [17] J. Xu, P. Ye, Q. Li, H. Du, Y. Liu, D. Doermann, Blind Image Quality Assessment Based on High Order Statistics Aggregation, *IEEE Trans. Image Process.* 25 (2016) 4444–4457. doi: 10.1109/TIP.2016.2585880.
- [18] Y. Zhang, A.K. Moorthy, D.M. Chandler, A.C. Bovik, C-DIIVINE: No-Reference Image Quality Assessment Based on Local Magnitude and Phase Statistics of Natural Scenes, *Signal Process. Image Commun.* 29 (2014) 725–747. doi: 10.1016/j.image.2014.05.004.
- [19] M. A. Saad, A. C. Bovik, C. Charrier, Blind Image Quality Assessment: A Natural Scene Statistics Approach in the DCT Domain, *IEEE Trans. Image Process.* 21 (2012) 3339–3352. doi: 10.1109/TIP.2012.2191563.
- [20] A. Mittal, A. K. Moorthy, A. C. Bovik, No-Reference Image Quality Assessment in the Spatial Domain, *IEEE Trans. Image Process.* 21 (2012) 4695–4708. doi: 10.1109/TIP.2012.2214050.
- [21] A. K. Moorthy, A.C. Bovik, A Two-Step Framework for Constructing Blind Image Quality Indices, *IEEE Signal Process. Lett.* 17 (2010) 513–516. doi: 10.1109/LSP.2010.2043888.
- [22] L. Liu, B. Liu, H. Huang, A.C. Bovik, No-Reference Image Quality Assessment Based on Spatial and Spectral Entropies, *Signal Process. Image Commun.* 29 (2014) 856–863. doi: 10.1016/j.image.2014.06.006.
- [23] A. Mittal, R. Soundararajan, A. C. Bovik, Making a “Completely Blind” Image Quality Analyzer, *IEEE Signal Process. Lett.* 20 (2013) 209–212. doi: 10.1109/LSP.2012.2227726.
- [24] W. Xue, L. Zhang, X. Mou, Learning without Human Scores for Blind Image Quality Assessment, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR '2013, Portland, OR, USA, 2013*, pp. 995–1002. doi: 10.1109/CVPR.2013.133.
- [25] K. Gu, G. Zhai, X. Yang, W. Zhang, Hybrid No-Reference Quality Metric for Singly and Multiply Distorted Images, *IEEE Trans. Broadcast.* 60 (2014) 555–567. doi: 10.1109/TBC.2014.2344471.
- [26] Q. Wu, Z. Wang, H. Li, A Highly Efficient Method for Blind Image Quality Assessment, in: *Proceedings of the IEEE Int. Conf. on Image Processing, ICIP '2015, Quebec City, QC, Canada, 2015*, pp. 339–343, 2015. doi: 10.1109/ICIP.2015.7350816.
- [27] R. Hassen, Z. Wang, M. M. A. Salama, Image Sharpness Assessment Based on Local Phase Coherence, *IEEE Trans. Image Process.* 22 (2013) 2798–2810. doi: 10.1109/TIP.2013.2251643.

- [28] A. K. Moorthy, A. C. Bovik, Blind Image Quality Assessment: From Natural Scene Statistics to Perceptual Quality, *IEEE Trans. Image Process* 20 (2011) 3350–3364. doi: 10.1109/TIP.2011.2147325.
- [29] L. Li, W. Lin, X. Wang, G. Yang, K. Bahrami, A. C. Kot, No-Reference Image Blur Assessment Based on Discrete Orthogonal Moments, *IEEE Trans. Cybern.* 46 (2016) 39–50. doi: 10.1109/TCYB.2015.2392129.
- [30] L. Liu, Y. Hua, Q. Zhao, H. Huang, A. C. Bovik, Blind Image Quality Assessment by Relative Gradient Statistics and Adaboosting Neural Network, *Signal Process. Image Commun.* 40 (2016) 1–15. doi: 10.1016/j.image.2015.10.005.
- [31] X. Min, K. Gu, G. Zhai, J. Liu, X. Yang, C.W. Chen, Blind Quality Assessment Based on Pseudo-Reference Image, *IEEE Trans. Multimed.* 20 (2018) 2049–2062. doi: 10.1109/TMM.2017.2788206.
- [32] Q. Wu, H. Li, F. Meng, K.N. Ngan, B. Luo, C. Huang, B. Zeng, Blind Image Quality Assessment Based on Multichannel Feature Fusion and Label Transfer, *IEEE Trans. Circuits Syst. Video Technol.* 26 (2016) 425–440. doi: 10.1109/TCSVT.2015.2412773.
- [33] Q. Wu, H. Li, F. Meng, K. N. Ngan, S. Zhu, No Reference Image Quality Assessment Metric via Multi-Domain Structural Information and Piecewise Regression, *J. Vis. Commun. Image Represent.* 32 (2015) 205–216. doi: 10.1016/j.jvcir.2015.08.009.
- [34] L. Li, Y. Yan, Z. Lu, J. Wu, K. Gu, S. Wang, No-Reference Quality Assessment of Deblurred Images Based on Natural Scene Statistics, *IEEE Access* 5 (2017) 2163–2171. doi: 10.1109/ACCESS.2017.2661858.
- [35] M. Rakhshanfar, M. A. Amer, Sparsity-Based No-Reference Image Quality Assessment for Automatic Denoising, *Signal Image Video Process.* 12 (2018) 739–747. doi: 10.1007/s11760-017-1215-3.
- [36] K. Ma, W. Liu, T. Liu, Z. Wang, D. Tao, DipIQ: Blind Image Quality Assessment by Learning-to-Rank Discriminable Image Pairs, *IEEE Trans. Image Process.* 26 (2017) 3951–3964. doi: 10.1109/TIP.2017.2708503.
- [37] K. Bahrami, A. C. Kot, A Fast Approach for No-Reference Image Sharpness Assessment Based on Maximum Local Variation, *IEEE Signal Process. Lett.* 21 (2014) 751–755. doi: 10.1109/LSP.2014.2314487.
- [38] P. V. Vu, D. M. Chandler, A Fast Wavelet-Based Algorithm for Global and Local Image Sharpness Estimation, *IEEE Signal Process. Lett.* 19 (2012) 423–426. doi: 10.1109/LSP.2012.2199980.
- [39] R. Ferzli, L. J. Karam, A No-Reference Objective Image Sharpness Metric Based on the Notion of Just Noticeable Blur (JNB), *IEEE Trans. Image Process.* 18 (2009) 717–728. doi: 10.1109/TIP.2008.2011760.
- [40] Y. Zhang, D. M. Chandler, No-Reference Image Quality Assessment Based on Log-Derivative Statistics of Natural Scenes, *J. Electron. Imaging* 22 (2013) 043025. doi: 10.1117/1.JEI.22.4.043025.
- [41] W. Xue, X. Mou, L. Zhang, A. C. Bovik, X. Feng, Blind Image Quality Assessment Using Joint Statistics of Gradient Magnitude and Laplacian Features, *IEEE Trans. Image Process.* 23 (2014) 4850–4862. doi: 10.1109/TIP.2014.2355716.
- [42] K. Gu, W. Lin, G. Zhai, X. Yang, W. Zhang, C.W. Chen, No-Reference Quality Metric of Contrast-Distorted Images Based on Information Maximization, *IEEE Trans. Cybern.* 47 (2017) 4559–4565. doi: 10.1109/TCYB.2016.2575544.
- [43] K. Gu, G. Zhai, W. Lin, X. Yang, W. Zhang, No-Reference Image Sharpness Assessment in Autoregressive Parameter Space, *IEEE Trans. Image Process.* 24 (2015) 3218–3231. doi: 10.1109/TIP.2015.2439035.
- [44] N. D. Narvekar, L. J. Karam, A No-Reference Perceptual Image Sharpness Metric Based on a Cumulative Probability of Blur Detection, in *Proceedings of the International Workshop on Quality of Multimedia Experience, QoMEX '2009, San Diego, CA, USA, 2009*, pp. 87–91. doi: 10.1109/QOMEX.2009.5246972.
- [45] C. Feichtenhofer, H. Fassold, P. Schallauer, A Perceptual Image Sharpness Metric Based on Local Edge Gradient Analysis, *IEEE Signal Process. Lett.* 20 (2013) 379–382. doi: 10.1109/LSP.2013.2248711.

- [46] N. N. Ponomarenko, V. V. Lukin, O. I. Ereemeev, K. O. Egiazarian, J. T. Astola, Sharpness Metric for No-Reference Image Visual Quality Assessment, *SPIE* 8295 (2012) 829519. doi: 10.1117/12.906393.
- [47] T. Zhu, L. Karam, A No-Reference Objective Image Quality Metric Based on Perceptually Weighted Local Noise, *EURASIP J. Image Video Process.* (2014) 1–5. doi: 10.1186/1687-5281-2014-5.
- [48] Y. Gong, I. F. Sbalzarini, Image Enhancement by Gradient Distribution Specification, *Lecture Notes in Computer Science* 9009 (2015) 47–62. doi: 10.1007/978-3-319-16631-5_4.
- [49] F. Crété-Roffet, T. Dolmiere, P. Ladret, M. Nicolas, The Blur Effect: Perception and Estimation with a New No-Reference Perceptual Blur Metric, In *Proceedings of the Human Vision and Electronic Imaging XII, HVEI '2007, San Jose, CA, USA, 2007*, pp. 649201. doi: 10.1117/12.702790.
- [50] S. A. Golestaneh, D. M. Chandler, No-Reference Quality Assessment of JPEG Images via a Quality Relevance Map, *IEEE Signal Process. Lett.* 21 (2014) 155–158. doi: 10.1109/LSP.2013.2296038.
- [51] O. Ieremeiev, V. Lukin, K. Okarma, K. Egiazarian, B. Vozel, On properties of visual quality metrics in remote sensing applications, in: *Proceedings of the IS&T Int'l. Symp. on Electronic Imaging: Image Processing: Algorithms and Systems, IPAS '2022, San Francisco, CA, USA, 2022*, pp. 354-1-354-6. doi: 10.2352/EL.2022.34.10.IPAS-354.
- [52] D. Du, Y. Qi, H.g Yu, Y. Yang, K. Duan, G. Li, W.g Zhang, Q. Huang, Q. Tian, The Unmanned Aerial Vehicle Benchmark: Object Detection and Tracking, in: *Proceedings of the European Conference on Computer Vision, ECCV'2018*. doi: 10.48550/arXiv.1804.00518.
- [53] L. Zhang, Y. Shen, H. Li, VSI: A Visual Saliency-Induced Index for Perceptual Image Quality Assessment, *IEEE Trans. Image Process* 23 (2014) 4270–4281. doi:10.1109/TIP.2014.2346028.
- [54] K. Gu, L. Li, H. Lu, X. Min, W. Lin, A Fast Reliable Image Quality Predictor by Fusing Micro- and Macro-Structures, *IEEE Trans. Ind. Electron.* 64 (2017) 3903–3912. doi:10.1109/TIE.2017.2652339.
- [55] H. Ziaei Nafchi, A. Shahkolaei, R. Hedjam, M. Cheriet, Mean Deviation Similarity Index: Efficient and Reliable Full-Reference Image Quality Evaluator, *IEEE Access* 4 (2016) 5579–5590. doi:10.1109/ACCESS.2016.2604042.
- [56] R. Reisenhofer, S. Bosse, G. Kutyniok, T. Wiegand, A Haar wavelet-based perceptual similarity index for image quality assessment, *Signal Process. Image Commun.* 61 (2018) 33–43. doi:10.1016/j.image.2017.11.001.
- [57] D. Temel, M. Prabhushankar, G. AlRegib, UNIQUE: Unsupervised Image Quality Estimation. *IEEE Signal Process. Lett.* 23 (2016) 1414–1418. doi:10.1109/LSP.2016.2601119.
- [58] H. Jia, L. Zhang, T. Wang, Contrast and Visual Saliency Similarity-Induced Index for Assessing Image Quality, *IEEE Access* 6 (2018) 65885–65893. doi:10.1109/ACCESS.2018.2878739.
- [59] E. Dumic, S. Grgic, M. Grgic, IQM2: new image quality measure based on steerable pyramid wavelet transform and structural similarity index, *Signal Image Video Process.* 8 (2014) 1159–1168. doi:10.1007/s11760-014-0654-3.
- [60] S. Li, F. Zhang, L. Ma, K. N. Ngan, Image Quality Assessment by Separately Evaluating Detail Losses and Additive Impairments, *IEEE Trans. Multimed.* 13 (2011) 935–949. doi: 10.1109/TMM.2011.2152382.