Received: 23 February 2021 | Revised: 31 March 2021 | Accepted: 31 March 2021

DOI: 10.1002/col.22675

RESEARCH ARTICLE

COLOR RESEARCH AND APPLICATION WILEY

Illumination estimation challenge: The experience of the first 2 years

Egor Ershov[1,2] | **Alex Savchik**[1] | **Ilya Semenkov**[1,2] | **Nikola Banić**[3] | **Karlo Koščević**[4] | **Marko Subašić**[4] | **Alexander Belokopytov**[1] | **Arseniy Terekhin**[1] | **Daria Senshina**[1] | **Artem Nikonorov**[5] | **Zhihao Li**[6] | **Yanlin Qian**[7,8] | **Marco Buzzelli**[9] | **Riccardo Riva**[9] | **Simone Bianco**[9] | **Raimondo Schettini**[9] | **Jonathan T. Barron**[10] | **Sven Lončarić**[4] | **Dmitry Nikolaev**[1]

[1]Institute for Information Transmission Problems, Moscow, Russia

[2]The Higher School of Economics, National Research University, Moscow, Russia

[3]Gideon Brothers, Zagreb, Croatia

[4]Faculty of Electrical Engineering and Computing, University of Zagreb, Zagreb, Croatia

[5]Image Processing Systems Institute of RAS, Samara, Russia

[6]Nanjing University, Nanjing, China

[7]Tampere University, Tampere, Finland

[8]Honor Camera Team, Shenzhen, China

[9]Department of Informatics Systems and Communication, University of Milano—Bicocca, Milan, Italy

[10]Google Research, San Francisco, California, USA

Correspondence
Egor Ershov, Institute for Information Transmission Problems, RAS, Bol'shoi Karetnyi per. 19, Moscow, Russia.
Email: e.i.ershov@gmail.com

Funding information
Hrvatska Zaklada za Znanost, Grant/Award Number: IP-06-2016-2092; Russian Science Foundation, Grant/Award Number: 20-61-27089

Abstract

Illumination estimation is the essential step of computational color constancy, one of the core parts of various image processing pipelines of modern digital cameras. Having an accurate and reliable illumination estimation is important for reducing the illumination influence on the image colors. To motivate the generation of new ideas and the development of new algorithms in this field, two challenges on illumination estimation were conducted. The main advantage of testing a method on a challenge over testing it on some of the known datasets is the fact that the ground-truth illuminations for the challenge test images are unknown up until the results have been submitted, which prevents any potential hyperparameter tuning that may be biased. The First illumination estimation challenge (IEC#1) had only a single task, global illumination estimation. The second illumination estimation challenge (IEC#2) was enriched with two additional tracks that encompassed indoor and two-illuminant illumination estimation. Other main features of it are a new large dataset of images (about 5000) taken with the same camera sensor model, a manual markup accompanying each image, diverse content with scenes taken in numerous countries under a huge variety of illuminations extracted by using the SpyderCube calibration object, and a contest-like markup for the images from the Cube++ dataset. This article focuses on the description of the past two challenges, algorithms which won in each track, and the conclusions that were drawn based on the results obtained during the first and second challenge that can be useful for similar future developments.

KEYWORDS

challenge, color constancy, illumination estimation, mixed illumination, multiple illumination, white balancing

1 | INTRODUCTION

In the modern world of technology, a lot of devices, including mobile phones, tablets, and laptops are equipped with digital cameras. One of the essential parts of the image processing pipelines of these cameras is to remove the influence of the scene illumination on the image colors. For that to be done properly, it is first required to perform an accurate illumination estimation in an image or a video, and then using this information for the actual color correction step.[1] Human color perception system performs a similar task by means of a feature known as color constancy[2] that allows it to recognize object coloration regardless of the scene illumination. While the mechanism of color constancy in human color perception is not fully understood, there are numerous proposed implementations of the analogous feature in digital cameras known as auto white balance (AWB), which is supposed to achieve computational color constancy. Some of the simplest illumination estimation methods such as max-RGB[3-5] or Gray-World[6] and its upgrades[7,8] are based on simple image statistics and their main advantages are their speed and simplicity of implementation. Over time, the best accuracy used to be achieved by learning-based methods such as Bayesian learning,[9] spatio-spectral learning,[10] illumination solution space restriction,[11-13] using color moments,[14] regression trees with simple features from color distribution statistics,[15] spatial localizations,[16,17] convolutional neural networks,[18-21] genetic algorithms,[22] etc. Thousands of other scientific papers about the problem of illumination estimation have been written, but it is still not solved to a satisfactory level due to the following reasons:

- Volumes of publicly available scientific datasets are not large enough to cover all various cases of illumination.
- Images in existing datasets have ground-truth scene illumination, but information about the scene is usually not provided. This excludes or complicates the process of a deeper study of the problem and its division into subtasks. For example, an efficient illumination estimation process can vary considerably at different times of the day (Figure 1).
- Considering oversimplified formulation by evaluating only a single dominant light source (there are only few publicly available datasets with multiple light sources). The uniform illumination assumption is true for images taken on a cloudy day, or in a room with a single lamp and white walls. However, this assumption is not valid on a sunny day when there are at least two light sources: the sun and the sky. Another example includes a street at night and a closed room.

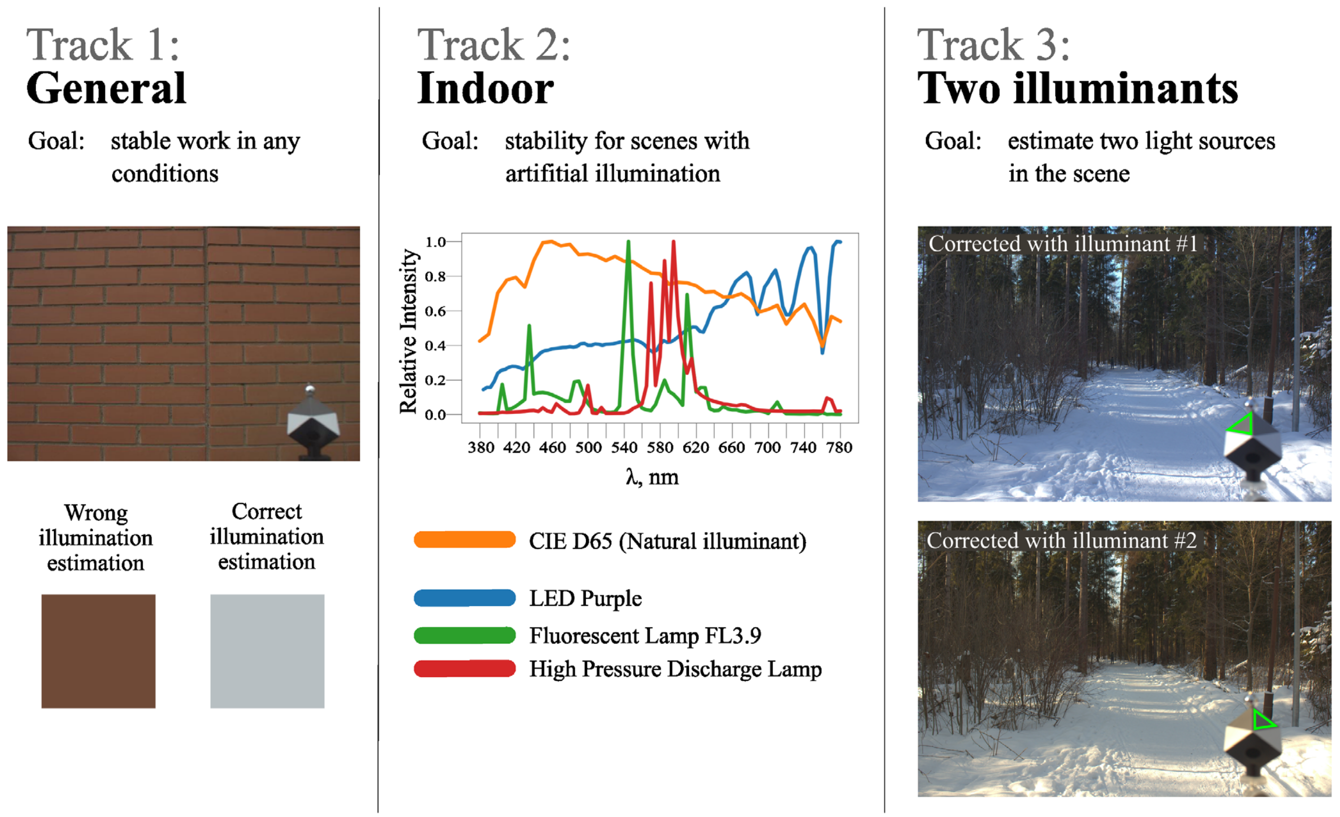

FIGURE 1 Ideas of second illumination estimation challenge illustration

To take a step forward toward overcoming these three obstacles, two challenges for the task of illumination estimation were organized. The First Illumination Estimation Challenge (IEC#1) was organized as a part of the 11th International Symposium on Image and Signal Processing and Analysis (ISPA 2019, September 23-25, 2019, Dubrovnik, Croatia). A year later, the Second Illumination Estimation Challenge (IEC#2) was organized as a part of the 13th International Conference on Machine Vision (ICMV 2020, November 2-6, 2020, Rome, Italy). A large illumination estimation dataset was created as a key component of the challenges. The dataset contains around 5000 images captured with two Canon cameras (Canon 600D and Canon 550D), which share the same image sensor model. For each image, ground-truth illuminations with up to two light sources per image were extracted. Additionally, various information about the scene content was provided for each image. Later collected images have been additionally filtered, processed, and finally published as Cube++ dataset.[23]

For both challenges, the ground-truth illuminations for the images from the test set were made publicly available only after the authors have submitted their solutions. This was done in order to prevent possible hyperparameter tuning that could compromise the testing fairness. Such testing with previously unknown ground-truth illumination can be seen as a potential advantage over testing on publicly available datasets.

The article is structured as follows: Section 2 describes IEC#1, Section 3 describes IEC#2, Section 4 gives an overview of the obtained challenge results, the results and the conclusions that were drawn based on them are discussed in Section 5.

2 | THE FIRST ILLUMINATION ESTIMATION CHALLENGE

As a new checkpoint in experimental results in the field of illumination estimation, the First Illumination estimation challenge (IEC#1) was conducted as part of the 11th International Symposium on Image and Signal Processing and Analysis (ISPA 2019) that was held in Dubrovnik, Croatia. The challenge gave the opportunity to researchers to benchmark new or existing illumination estimation methods on a newly created image dataset for which the ground-truth illumination would be released only after the challenge. The goal was to check the performance of various methods when the ground-truth for the test set was not available immediately. The participants were assigned the task to estimate a vector representing the color of the dominant illumination source in the new given dataset.

The challenge dataset was split into train and test parts. For method training, the participants were provided with the publicly available Cube+ dataset,[24] which includes both images and corresponding ground-truth illuminations. The test set consisted of 363 images taken with the same camera that was used to capture images of the Cube+ dataset. The images in the test data were made public on the last day of the challenge by providing the password to decrypt the test archive that was already available earlier. The ground-truth illuminations corresponding to the images in the test set were made public in a similar manner after the challenge ended. By hiding the test set until the end of the challenge, it was intended to prevent or significantly reduce problems such as potential data manipulation and method overfitting.

The quality of a solution was based on the reproduction angular error,[25] which is defined as

$$R(\mathbf{g},\mathbf{a}) = \arccos\left\langle \mathbf{1}, \frac{\mathbf{g}\oslash\mathbf{a}}{\|\mathbf{g}\oslash\mathbf{a}\|}\right\rangle, \tag{1}$$

where $\mathbf{g}$ and $\mathbf{a}$ are three-dimensional vectors representing ground truth and estimated illumination, respectively, $\mathbf{1}$ is the vector of perfectly corrected white color, that is, $\mathbf{1} = \frac{1}{\sqrt{3}}(1,1,1)$, $\oslash$ denotes element-wise division, $\langle\cdot\rangle$ denotes the scalar product of two vectors, and $\|\cdot\|$ denotes the Euclidean norm.

Nine solutions were submitted and the final ranking is shown in Table 1. For the sake of transparency the submitted illumination estimations were also made available on the challenge website.[*,†] Submitted solutions were evaluated in terms of three averages: the arithmetic mean, the median, and the trimean (the arithmetic mean of the two quartiles and the median, wherein the median's influence is doubled). These three statistical average measures were selected on the basis of their ubiquity in the color constancy literature, and the median was selected as the measure by which submitted solutions were to be ranked.

Upon reflection, using the median as the primary evaluation metric for this task may be suboptimal. The median is a robust average measure, in that it is invariant to extremely large or small inputs. Indeed, the median of a set of errors is just a function of 1 (or 2) of those constituent errors, and as such is extremely sensitive to the performance of the "average" image but is entirely invariant to performance on "difficult" images. Since the majority of images in any given dataset are, by definition, common, the median is much more likely to reflect performance on "easy" images than hard images. Thus, the winning method of IEC#1, after hyperparameter optimization, ignored the 20% of the worst and only 1% of the best mini-batch samples on every training step.[26] Hence,

Authors	Median	Mean	Trimean
Alex Savchik, Egor Ershov, and Simon Karpenko	**1.513**	**2.652**	**1.649**
Jonathan T. Barron and Yun-Ta Tsai	*1.590*	*2.489*	*1.731*
Yanlin Qian, Ke Chen, and Huanglin Yu	***1.638***	***2.934***	***1.773***
Ke Chen, Huanglin Yu, and Yanlin Qian	1.692	2.611	1.843
Yanlin Qian, Ke Chen, and Huanglin Yu	1.709	2.491	1.763
Yanlin Qian, Ke Chen, and Huanglin Yu	2.096	6.869	2.495
Viktor Vuk and V. N. Karazin	2.142	3.338	2.327
Simon Karpenko, Egor Ershov, and Alex Savchik	4.577	6.683	5.071
Hassan Ahmed Sial and Maria Vanrell I Martorell	5.906	7.289	6.180

TABLE 1 The results of the first illumination estimation challenge ranked by the median of the reproduction angular errors calculated from the illumination estimations reported by the authors for the test images; the error statistics are reported in degrees (°) and a lower error is equivalent to better estimation performance

the method achieved a good median error, while simultaneously not paying attention to the worst errors, which may be detrimental for the overall performance and the worst-case performance.

This is in contrast to the standard practice among camera manufacturers, where the primary goal is to avoid *large* errors, under the assumption that a grossly incorrect white balance is objectionable to a photographer or a viewer, but that small errors are often not noticeable. That is, it is acceptable if the subject of a photo is rendered slightly warmer or cooler than the objective true color, but it is unacceptable to render grass as white or the ocean as red. These priorities can be observed in the documentation for the DxOMark benchmark (the independent image quality assessment commonly used to evaluate and rank cameras). To quote the DxOMark website[‡]: "good and acceptable color manifests on a continuum rather than having a single fixed value." DxOMark therefore explicitly adopts a scoring system in which small or modest errors have little to no significance, and large errors are penalized. This priority can also be observed in recent academic publications, such as the work from Liba et al[27] documenting the technology behind Google's low-light photography systems on the Pixel smartphones. The measure white balance performance using the mean of the 25% largest errors in the benchmark — a metric that is so heavily focused on large errors that it is completely insensitive to both the median and the quartiles of the errors. Therefore, in the second iteration of our challenge, during evaluation we shifted our emphasis to metrics that more strongly reflect performance on difficult images, such as the aforementioned "mean 25%" metric. We also use mean *squared* error in future challenges, which functionally emphasizes larger errors more than smaller errors in comparison to "mean error".

3 | THE SECOND ILLUMINATION ESTIMATION CHALLENGE

3.1 | Dataset

For the Second Illumination Estimation Challenge, the new dataset Cube++[23] was used. This dataset is the extension of the Cube+[24] dataset, which was used as a train set in IEC#1. The Cube++ provides a lot of new images captured in various conditions, including indoor images and images with two illumination sources present in the scene. Such a dataset sets the ground for a broader range of illumination estimations research as indicated with multiple tracks of IEC#2. To make the Cube++ compliant with its predecessor Cube+, the same camera sensor model and calibration object were used during the image acquisition.

For ground-truth illumination extraction, the SpyderCube color target was used. Due to its cuboid shape, ground-truth illuminations can be extracted from two different directions. This is beneficial in two ways. First, by comparing such two ground-truths, it can be verified whether the scene illumination is uniform. Second, when needed, the ground-truth for two-illuminant estimation can be extracted.

Along with the ground-truth illuminations, each image in the dataset is also accompanied by the metadata related to the conditions during image capturing procedure (ISO, exposure time, etc.), and manually labeled semantic information such as whether the image is captured indoor or outdoor, at what time of day was the image captured, is the image sharp, etc. Manually labeled semantic data was not available for the test set during the time of the challenge.

The dataset contains files of three types: JPG images for preview; PNG images for illumination estimation; ground-truth illuminations and metadata in JSON format.[§]

3.2 | Tracks

The second illumination estimation challenge was organized in three tracks: general, indoor, and two-illuminant. For each track different metrics, appropriate for the given task, were used, but all based on the reproduction angular error.[25] For reproducibility, the source code for each metrics was made public. During the challenge, the reproduction angular error was mistakenly calculated with the wrong order of arguments,¶ resulting in a slight difference in results, but nevertheless with the same ranking of methods in the challenge. All tables related to IEC#2 in this article are recalculated with the corrected metrics (Figure 2).

3.2.1 | General track

Usually, the quality of an illumination estimation algorithm is measured using a statistic of an error measure computed for all images in the test dataset. Very often, including IEC#1, the median of per image angular error values is used, which is reasonable in the case when the dataset itself has errors in markup. Nevertheless, as it was mentioned before, it is much more important not to have extremely wrong illumination estimations than to have the best values for only a single statistic. Moreover, the problem of errors in markup should be solved not by metric selection, but by careful data labeling. Therefore, the general track of the challenge was devoted to robust illumination estimation algorithms.

For the general track **dataset**, all images with angular difference less than 2° between the ground-truth chromaticities extracted from the left and right SpyderCube gray faces were selected from Cube++. For each image, the dominant light source was chosen manually.

The **metric** for the general track was the mean of the 25% largest reproduction angular errors (1).

In Table 2, the leaderboard for the general track is shown. The first place was achieved by Z. Li from Nanjing University with the CAUnet algorithm. This algorithm achieved the lowest mean reproduction error of the 25% of the worst estimations, which was 4.087°.

3.2.2 | Indoor track

One of the stand-alone photography categories is indoor photography. Very often, in indoor conditions, the illumination in the scene is quite complicated. In different places of the scene there may be various sources of illumination such as light coming through the window, incandescent lamps, LED lamps, etc. Under these conditions, the determination

FIGURE 2 Examples of general track images

TABLE 2 The results of the second illumination estimation challenge general track; the error statistics are reported in degrees (°) and a lower error is equivalent to better estimation performance

Team	Algorithm	Mean (worst 25%)	Worst RE	Mean	Median	Trimean
Z. Li	**CAUnet**	**4.087**	**16.571**	**1.606**	**0.969**	**1.085**
Z. Li	CAUnet	4.334	18.307	1.727	1.073	1.207
X. Xing et al	AL-AWB (sub. # 2)	4.413	17.838	1.82	1.192	1.32
X. Xing et al	AL-AWB (sub. # 1)	4.646	17.838	1.887	1.236	1.349
Y. Qian	***sde-awb***	***4.944***	***18.81***	***1.906***	***1.169***	***1.27***
Y. Qian	sde-awb	5.097	18.019	1.948	1.15	1.298
Y. Qian	sde-awb	5.351	26.692	2.028	1.15	1.285
J. Qiu et al	illumGAN	10.327	26.359	4.766	3.635	3.914
BASELINE	GreyWorld	10.565	24.223	4.539	3.312	3.575
BASELINE	Constant	15.871	35.427	6.704	3.881	5.062
J. Qiu et al	*illGAN1.0*	*20.905*	*41.987*	*17.536*	*17.138*	*17.186*
J. Qiu et al	illGAN1.0	21.104	47.809	17.615	17.132	17.177
J. Qiu et al	illGAN1.0	21.469	54.426	17.666	17.146	17.247

of the dominant light source in the scene may turn out to be a rather difficult task, mostly because of inter-reflections.

The indoor track **dataset** contains exclusively images for which it was manually determined that they are captured in indoor spaces. Additionally, the angle between ground-truth illuminations extracted from SpyderCube's left and right gray faces had to be less than 2° to include an image into the dataset (Figure 3).

As a **metric** for the indoor track, the mean reproduction error given in (1) was used.

In Table 3, the final indoor track ranking is shown. The first place was achieved by Y. Qian from Huawei Multimedia Team with an algorithm sde-awb which achieved the mean reproduction error of 2.511°.

3.2.3 | Two-illuminant track

In everyday life, there are rarely situations where there is really only one source of illumination in the scene. Even during the daytime, it is customary to divide the illumination into two sources—the sun and sky. The main question of this track is whether it is possible to reliably extract more information about illumination using a single image. For these purposes, a dataset was assembled using a volumetric color target (SpyderCube) whose faces are illuminated by different sources (Figure 4).

Two-illuminant track **dataset** includes images for which the angle between ground-truth illuminations extracted from SpyderCube's left and right gray faces is greater than or equal to 2°. This helps to ensure that illuminations that are reflected from SpyderCube's faces are different enough to considered them originating from two different light sources. In this track, it was required to estimate two vectors, each representing one light (Table 4).

The first place was achieved by Y. Qian (Huawei MultiMedia Team) with the sde-awb algorithm. The final squared sum of two angular reproduction errors was 31.026217 for this algorithm.

As a **metric**, the sum of squared reproduction angular errors:

$$E(\mathbf{g_1},\mathbf{g}_2,\mathbf{a}_1,\mathbf{a}_2)=\min\left(R^2(\mathbf{g_1},\mathbf{a_1})+R^2(\mathbf{g_2},\mathbf{a_2}),R^2(\mathbf{g_1},\mathbf{a_2})+R^2(\mathbf{g_2},\mathbf{a_1})\right), \tag{2}$$

where g_1, g_2 are ground-truth chromaticity vectors and a_1, a_2 are algorithm estimations. Such a strong metric allows to penalize an algorithm with a single answer for two close estimations.

3.3 | Challenge rules for all tracks

Registration for the challenge was held up to 13:00 (GMT + 3) on July 31, 2020. Participants were required to list both names and affiliations for all team members.

For each track, a corresponding dataset was prepared. Each team could have submitted up to three solutions for each track resulting in up to nine submissions in total for three tracks. Participants were required to run their model on the test part and to submit their prediction files. Submission was available until 13:00 (GMT + 3) on 31 July 2020.

The test dataset was published a 2 weeks before submission day in an encrypted archive, the SpyderCube instance

FIGURE 3 Examples of indoor track images

TABLE 3 The results of the second illumination estimation challenge indoor track; the error statistics are reported in degrees (°) and a lower error is equivalent to better estimation performance

Team	Algorithm	Mean	Median	Trimean	Mean (worst 5%)	Worst RE
X. Xing et al	AL-AWB (sub. # 2)	2.511	2.251	2.208	12.369	12.369
Y. Qian	**sde-awb**	**2.512**	**1.74**	**1.922**	**10.31**	**10.31**
X. Xing et al	AL-AWB (sub. # 1)	2.85	2.251	2.403	13.318	13.318
M. Buzzelli et al	*MCGAN*	*3.185*	*2.354*	*2.349*	*18.125*	*18.125*
J. Qiu et al	illumGAN	3.239	2.623	2.649	12.672	12.672
M. Buzzelli et al	PCGAN	3.272	2.354	2.39	18.125	18.125
BASELINE	GreyWorld	4.098	3.681	3.594	15.769	15.769
BASELINE	Constant	13.872	13.532	14.108	26.586	26.586

was masked out from the image scenes, the identifiers were shuffled to avoid finding similarities in close images with close identifiers, and there were no ground-truth illuminations nor manually annotated properties given.

4 | METHOD DESCRIPTION

4.1 | CAUnet

We utilize an encoder-decoder based Unet[28] combining global illumination with white patches as our primary model. The network's deep features are usually too generic so that it is clear that some channel features are more significant than others in different scenes. In order to make the Unet pay more attention to these crucial channels, we employ channel attention-convolutions (CA-Convs) blocks inherited from W-Net[29] to specify the various scenes' response. CA-Convs block first uses global pooling to extract spatial information from convolutional features, and then transforms them via fully connected layers (FC), ReLU, and sigmoid. At last, it multiplies the convolutional features with sigmoid's output, which represented the channel attention's weights. CA-Convs significantly enhances the network's robustness by paying attention to different channel features in diverse scenes. The whole

FIGURE 4 Examples of two-illuminant track images

Team	Algorithm	Mean squared	Mean	Median	Trimean
Y. Qian	**sde-awb (sub. # 1)**	**31.801**	**2.764**	**2.284**	**2.312**
Y. Qian	sde-awb (sub. # 2)	32.009	2.727	2.206	2.327
X. Xing et al	AL-AWB (sub. # 2)	32.157	2.64	1.876	2.101
Y. Liu et al	3du-awb	36.071	2.827	2.494	2.477
X. Xing et al	AL-AWB (sub. # 1)	42.34	2.925	2.121	2.323
BASELINE	GreyWorld	83.143	4.153	3.506	3.675
BASELINE	Constant	128.531	5.053	3.362	3.727

TABLE 4 The results of the second illumination estimation challenge two-illuminant track; the error statistics are reported in degrees (°) and a lower error is equivalent to better estimation performance

network is shown in Figure 5. All activation functions are implemented with 0.2 negative slope's parametric rectified linear units (PReLU).[30] We normalize the global average pooling's output to one at the output layer, which is the same as the white balance coefficient's constraint.

We design our model using Pytorch and train it with 4 NVIDIA V100 GPUs. We use Adam with a learning rate of 5e-5 as our optimizer. The CA-Unet was trained for 30 epochs with batch size 96. And then, we use the plateau scheduler to auto reduce the learning rate. For more details of CAUnet, see our article.[31]

4.2 | Sde-awb

For the sake of robust performance in all three tracks in IEC 2, sde-awb is designed with portable modules, easy-to-follow structure and plain training strategy. Shown in Figure 6, A pre-trained Squeeze-Net backbone is firstly applied as baseline model (can be cascaded up to three times), with a differential 2D chroma histogram layer stacked to boost further the performance. The first offers semantic feature while the latter describes the distribution of the visible color. A shallow MLP is then added to accommodate image capturing metadata (eg, exposure, shutter) contained in Exif. On top of model A, following,[32] three SqueezeNets are cascaded in order to realize a coarse-to-fine illumination regression.

Adam optimizer is adopted in our net to minimize the mean reproduction angular error. Validated on three challenge tracks, all proposed modules show their effectiveness, making sde-awb a top-one or top-two solution in varying use cases (Figure 7).

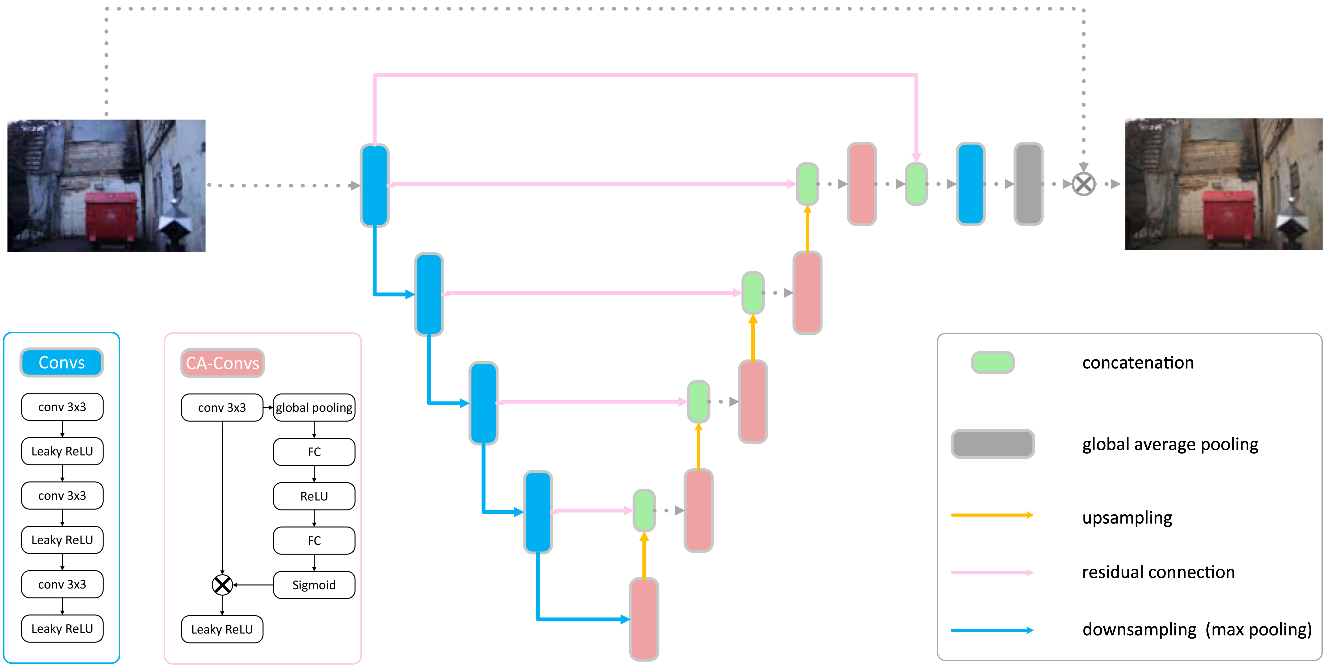

FIGURE 5 Channel attention (CA)-Unet's architecture

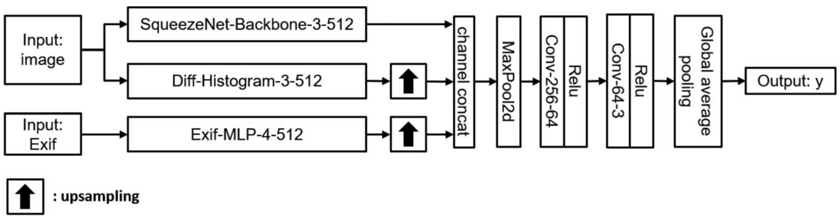

FIGURE 6 The architecture of sde-awb (model A). "LayerName-*x*-*y*" denotes a 2D layer with *x* input channels and *y* output channels where the layer is either a standard convolution layer, a backbone network (eg, SqueezeNet) or a 2D LSTM. "Upsampling" does the upsampling to match the spatial dimensions of the output of SqueezeNet backbone. "Channel concat" concatenates three features maps along the feature channel axis. **y** is the illumination color vector after the last ReLU layer. Model B has the similar architecture without the Diff-Histogram and the Exif-MLP sub-brunches

In our solution we demonstrate a generic illumination estimation—sde-awb, can perform well for general, indoor and two-illuminant tracks. With specific module design and combination, it obtains first place in both indoor (mean error 1.763) and two-illuminant (mean error 2.751) tracks and second place in general track (mean error 1.914). Our used modules, Squeeze-Net backbone, differential 2D chroma histogram layer and a shallow MLP show their effectiveness on a cross-validated ablation study. Our future plan is to explore more elegant architecture to merge cross-domain information like Exif metadata. For more details of sde-awb, see our article.[33]

4.3 | PCGAN, MCGAN

The Predicted-Consensus Generative Adversarial Network (PCGAN) and Median-Consensus Generative Adversarial Network (MCGAN) are based on the concept of generating a spatially-varying illumination map from the input image, and subsequently reducing this map to three RGB coefficients.

The overall process is illustrated in Figure 8: starting from an input image, its color-balanced version is directly created through a generative model. Then, the pixel-by-pixel ratio between the input image and the generated image is computed, in order to obtain a first proposal for a spatially varying illuminant map. From this, a properly-defined consensus strategy is eventually applied, with the goal of deriving a global-estimation illuminant, which follows the von Kries model.

The generative model is based on the pix2pix architecture,[35] trained with a set of input-target image pairs. Specifically, the input is the raw-unbalanced image, from which the black level is first subtracted. The target is the balanced image, obtained dividing the

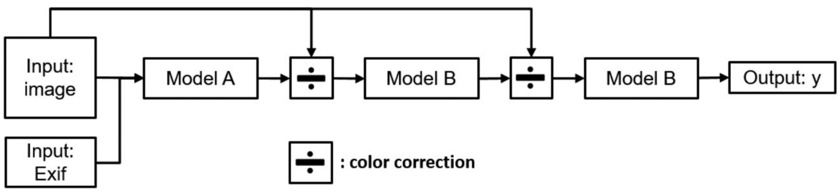

FIGURE 7 The architecture of sde-awb (model C). "Color correction" takes the predicted illumination vector from the layer above and do color correction on the input image, resulting into a color-corrected image. Note that model A takes 2 inputs while model B takes only 1 input

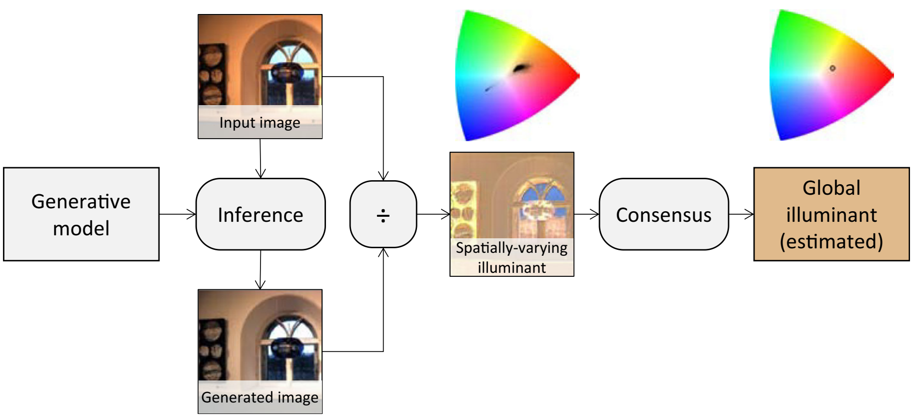

FIGURE 8 Schematic representation of the PCGAN/MCGAN solution for illuminant estimation. The global estimation is defined as the consensus over the spatially-varying illuminant coming from a generative model. Illuminant distributions are visualized in Angle-Retaining Chromaticity[34]

input image channel-by-channel by the provided ground truth illuminants. This configuration was determined after preliminary experiments against alternative representations, including a fixed white balancing preprocessing, and a de-linearization of the input images.

At inference time, a consensus strategy is applied in order to map the spatially-varying illuminant into a global estimated illuminant. Several alternatives were investigated:

- Mean: each channel of the spatially-varying illuminant map is independently averaged.
- Median: the median is computed for each channel, as a more-robust statistic with respect to outliers.
- Weighted mean: each pixel is weighted in the overall average as a function of its luminance, penalizing low luminance for numerical instability, and high luminance due to sensor non-linearities.
- Meta strategy: a neural network is trained to predict the best consensus strategy for each image, thus treated as an image classification problem.

All solutions were trained on the Cube+ dataset, originally composed of 1707 images collected with the same Canon 550D camera used for part of the challenge images. Developmental experiments were validated on the challenge training set, specifically the "indoor" track, and lead to the eventual submission of two configurations for the challenge: PCGAN (Predicted-Consensus GAN, based on the meta strategy) and MCGAN (Median-Consensus GAN). These solutions were fine-tuned on the challenge training set after the initial Cube+ training. For more details of PCGAN and MCGAN see our article.[36]

5 | DISCUSSION

The main goal for a participant in the described challenges was to create an algorithm that predicts the ground-truth value for each image with the least error. Thus, the best algorithm is the one that most accurately predicts the answer proposed by the organizers. However, would such an algorithm be the best one for estimating the illumination in a scene? The answer to this question is decidedly negative.

Firstly, when using this type of datasets, the task is implicitly substituted from "Estimate illumination parameters in the scene" to "Predict the color of the white patch in the scene." Considering that the color of a white flat object in the scene changes significantly with a change in its orientation, it turns out that the winning algorithm solves the problem of predicting the color of a white reference object in some way located better than others. In other words, the winning algorithm is good at predicting one of many averaged illumination estimation in the scene, but not all of them and it may not be the best one. The experience gained from the two-illuminant track suggests that even in simple daytime scenes where the only light source is the sun, the difference in chromaticity between different sides of the SpyderCube can go up to more than 10°. In that context, the statement "Algorithm X works with an accuracy of less than 1°" becomes negligible, since the scatter of estimates of correct answers in some images can be even 10 times greater.

Secondly, the errors in the dataset cannot be neglected. In the first challenge, the median of errors was used to overcome this. In this case, large error values induced by incorrectly labeled data have less effect on the final accuracy. However, as shown by the results of the challenge, such an error allows the solution to be forgiven for too many gross errors, which, combined with the specific structure of the dataset (more than half of the images were taken outdoors during the day), will lead to inadequate ranking of algorithms. For this reason, in the second challenge, much stricter metrics were used, and the data was marked up and processed with special attention.

These arguments have led us to a logical question: "What is the best way to measure accuracy using such datasets?" According to the organizers of the challenge, the answer to this question is not trivial, and it depends on the ultimate goal for which a technical solution is being developed. In the case of the physical task of estimating illumination parameters in a scene, it is worth noting that the type of datasets used is not the best one: even a three-dimensional achromatic calibration object (gray ball) allows collecting a very small amount of information about the lighting in the scene. This is probably why in 2020, scientists from Simon Fraser University published a new dataset[37] in which the illumination for one scene was determined at once at many different points using flying drone. In the tasks of forming digital photographs, where white balancing algorithms are still used, the stability of the algorithm is much more important than the average accuracy of its operation. Thus, in this problem, it is more appropriate to use the mean of the angular error over some percentage (1%, 5%, etc.) of the worst responses. On the one hand, this kind of metric still allows operating with angular values, on the other hand, to control the worst cases of the algorithm, which is critical from the point of view of the end user.

ACKNOWLEDGMENTS

The work under data collection was supported by the Croatian Science Foundation under Project IP-06-2016-2092, and the work under data processing was supported by Russian Science Foundation under Grant 20-61-47089. Authors also grateful to Irina Zhdanova for helping in preparing beautiful illustrations.

ENDNOTES

* https://www.isispa.org/illumination-estimation-challenge/leaderboard

† https://archive.is/CE8WH

‡ https://www.dxomark.com/dxomark-selfie-how-we-test-smartphone-front-camera-still-image-quality/

§ Detailed description of the dataset is available on GitHub https://github.com/Visillect/CubePlusPlus and IEC# 2 website http://chromaticity.iitp.ru

¶ See commit https://github.com/Visillect/CubePlusPlus/commit/eb37e9fe7b17eeb8bdb8fa71cb109febe92630a7

DATA AVAILABILITY STATEMENT

The contents of the manuscript are shared upon request.

ORCID

Egor Ershov https://orcid.org/0000-0001-6797-6284

REFERENCES

[1] Gijsenij A, Gevers T. Color constancy using natural image statistics and scene semantics. *IEEE Trans Pattern Anal Mach Intell*. 2011;33(4):687-698.

[2] Ebner M. *Color Constancy*. The Wiley-IS&T Series in Imaging Science and Technology. Chichester, West Sussex PO19 8SQ, England: Wiley; 2007.

[3] Land EH. The retinex theory of color vision. *Sci Amer*. 1977;237:108-128.

[4] Funt B, Shi L. The rehabilitation of MaxRGB. Paper presented at: Color and Imaging Conference. Vol. 2010. Society for Imaging Science and Technology; 2010:256-259.

[5] Banić N, Lončarić S. Improving the White patch method by subsampling. Paper presented at: Image Processing (ICIP), 2014 21st IEEE International Conference on IEEE; 2014:605-609.

[6] Buchsbaum G. A spatial processor model for object colour perception. *J Franklin Inst*. 1980;310(1):1-26.

[7] Finlayson GD, Trezzi E. Shades of gray and colour constancy. Paper presented at: Color and Imaging Conference. Vol. 2004. Society for Imaging Science and Technology; 2004:37-41.

[8] Van De Weijer J, Gevers T, Gijsenij A. Edge-based color constancy. *IEEE Trans Image Process*. 2007;16(9):2207-2214.

[9] Gehler PV, Rother C, Blake A, Minka T, Sharp T. Bayesian color constancy revisited. Paper presented at: IEEE Conference on Computer Vision and Pattern Recognition CVPR; 2008:1-8.

[10] Chakrabarti A, Hirakawa K, Zickler T. Color constancy with spatio-spectral statistics. *IEEE Trans Pattern Anal Mach Intell*. 2012;34(8):1509-1519.

[11] Banić N, Lončarić S. Color cat: remembering colors for illumination estimation. *IEEE Signal Process Lett*. 2015;22(6):651-655.

[12] Banić N, Lončarić S. Using the red chromaticity for illumination estimation. Paper presented at: Image and Signal

Processing and Analysis (ISPA), 2015 9th International Symposium on IEEE; 2015:131-136.
[13] Banić N, Lončarić S. Color Dog: Guiding the Global Illumination Estimation to Better Accuracy. *VISAPP*; 2015:129-135.
[14] Finlayson GD. Corrected-moment illuminant estimation. Paper presented at: Proceedings of the IEEE International Conference on Computer Vision; 2013:1904-1911.
[15] Cheng D, Price B, Cohen S, Brown MS. Effective learning-based illuminant estimation using simple features. Paper presented at: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2015:1000-1008.
[16] Barron JT. Convolutional Color Constancy. Paper presented at: Proceedings of the IEEE International Conference on Computer Vision; 2015:379-387.
[17] Barron JT, Tsai YT. Fast Fourier color constancy. Paper presented at: IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR; 2017. Vol. 1. IEEE.
[18] Bianco S, Cusano C, Schettini R. Color constancy using CNNs. Paper presented at: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops; 2015: 81-89.
[19] Shi W, Loy CC, Tang X. Deep specialized network for illuminant estimation. Paper presented at: European Conference on Computer Vision; Springer; 2016:371-387.
[20] Hu Y, Wang B, Lin S. FC4: Fully convolutional color constancy with confidence-weighted pooling. Paper presented at: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2017:4085-4094.
[21] Oh SW, Kim SJ. Approaching the computational color constancy as a classification problem through deep learning. *Pattern Recogn*. 2017;61:405-416.
[22] Koščević K, Banić N, Lončarić S. Color Beaver: Bounding Illumination Estimations for Higher Accuracy. *VISAPP*; 2019: 183-190.
[23] Ershov E, Savchik A, Semenkov I, et al. The cube++ illumination estimation dataset. *IEEE Access*. 2020;8:227511-227527.
[24] Banic N, Koščević K, Lončarić S. Unsupervised learning for color Constancy. *arXiv:171200436 2019*.
[25] Finlayson GD, Zakizadeh R. Reproduction angular error: an improved performance metric for illuminant estimation. *Perception*. 2014;310(1):1-26.
[26] Savchik A, Ershov E, Karpenko S. Color Cerberus. Paper presented at: 11th International Symposium on Image and Signal Processing and Analysis (ISPA) IEEE; 2019:355-359.
[27] Liba O, Murthy K, Tsai YT, et al. Handheld mobile photography in very low light. SIGGRAPH Asia; 2019.
[28] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. Paper presented at: International Conference on Medical Image Computing and Computer-Assisted Intervention; Springer; 2015: 234-241.
[29] Uhm KH, Kim SW, Ji SW, Cho SJ, Hong JP, Ko SJ. W-Net: two-stage U-Net with misaligned data for raw-to-RGB mapping. Paper presented at: IEEE/CVF International Conference on Computer Vision Workshop (ICCVW) IEEE; 2019: 3636-3642.
[30] He K, Zhang X, Ren S, Sun J. Delving deep into rectifiers: surpassing human-level performance on Imagenet classification. Paper presented at: Proceedings of the IEEE International Conference on Computer Vision; 2015:1026-1034.
[31] Li Z, Ma Z. Robust white balance estimation using joint attention and angular loss optimization. Paper presented at: Thirteenth International Conference on Machine Vision. Vol. 11605. International Society for Optics and Photonics; 2021: 116051E.
[32] Yu H, Chen K, Wang K, Qian Y, Zhang Z, Jia K. Cascading convolutional color constancy. Paper presented at: AAAI Conference on Artificial Intelligence; 2020, Cascading Convolutional Color Constancy.
[33] Qian Y, Feng S, Qian K, Wang M. SDE-AWB: a generic solution for Second International Illumination Estimation Challenge. *Thirteenth International Conference on Machine Vision, International Society for Optics and Photonics, SPIE*. Vol 11605; 2021:168-173. https://doi.org/10.1117/12.2588328.
[34] Buzzelli M, Bianco S, Schettini R. ARC: angle-retaining chromaticity diagram for color constancy error analysis. *J Opt Soc Am A*. 2020;37(11):1721-1730.
[35] Isola P, Zhu JY, Zhou T, Efros AA. Image-to-image translation with conditional adversarial networks. Paper presented at: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2017:1125-1134.
[36] Buzzelli M, Riva R, Bianco S, Schettini R. Consensus-driven illuminant estimation with GANs. *Thirteenth International Conference on Machine Vision, International Society for Optics and Photonics, SPIE*. Vol 11605; 2021:578-584. https://doi.org/10.1117/12.2587589.
[37] Aghaei H, Funt B. A flying gray ball multi-illuminant image dataset for color research. *Journal of Imaging Science and Technology*. 2020;64(5):50411-1-50411-8(8).

AUTHOR BIOGRAPHIES

Egor Ershov (ORCID = https://orcid.org/0000-0001-6797-6284) received a PhD degree from the Faculty of Radio Engineering and Computer Technology in 2019. He is a senior researcher in Vision System Laboratory in Institute of Information Transmission Problems (Kharkevic Institute) of the Russian Academy of Science (IITP RAS), Moscow. He is also senior lecturer at Computer Science Faculty in Higher School of Economy. His main areas of research are image processing and analysis, particularly color computer vision, color reproduction technologies, colorimetry, human vision system, Hough transform.

Alex (Alexey) Savchik, Vision Systems Laboratory, Institute for Information Transmission Problems (Kharkevich Institute) of the Russian Academy of Sciences (IITP RAS), Moscow, Russia. Alex Savchik in 2014 graduated from Department of Mechanics and Mathematics of Moscow State University, in 2012 from Yandex School of Data Analysis. His research interests include deep learning and computer vision.

Ilya (Illya) Semenkov, Vision Systems Laboratory, Institute for Information Transmission Problems

(Kharkevich Institute) of the Russian Academy of Sciences (IITP RAS), Moscow, Russia, received bachelor's degree at National Research University Higher School of Economics (NRU HSE) in 2019 and is currently master's student at Faculty of Computer Science of NRU HSE. His research interests include deep learning, computer vision, statistical learning and optimal transport.

Nikola Banić received BS, MS, and PhD degrees in computer science in 2011, 2013, and 2016, respectively. He is currently working as a senior computer vision engineer at Gideon Brothers, Zagreb, Croatia. He has worked in real-time image enhancement for embedded systems, digital signature recognition, people tracking and counting, and image processing for stereo vision. His research interests include image enhancement, color constancy, image processing for stereo vision, and tone mapping.

Karlo Koščević received BS and MS degrees in computer science in 2016 and 2018, respectively. He is currently in his second year of the technical sciences in the scientific field of computing PhD program at the Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia. His research interests include image processing, image analysis, and deep learning. His current research is in the area of color constancy with a focus on learning-based methods for illumination estimation.

Marko Subašić received a PhD degree from the Faculty of Electrical Engineering and Computing at the University of Zagreb in 2007. Since 1999, he has been working in the Department for Electronic Systems and Information Processing at the Faculty of Electrical Engineering and Computing at the University of Zagreb, currently as an associate professor. He teaches several courses at the graduate and undergraduate levels. His research interests lie in image processing and analysis and neural networks, with a particular interest in image segmentation, detection techniques, and deep learning. He is a member of the IEEE—Computer Society, the Croatian Center for Computer Vision, the Croatian Society for Biomedical Engineering and Medical Physics, and the Centre of Research Excellence for Data Science and Advanced Cooperative Systems.

Alexander Belokpytov is a well-known author in the field of the journal scope. His/Her research interests include interest 1, interest 2.

Arseniy Terekhin works in the Vision Systems Laboratory, Institute for Information Transmission Problems (Kharkevich Institute) of the Russian Academy of Sciences (IITP RAS), Moscow, Russia. His research interests include software architecture and computer vision.

Daria Senshina, Vision Systems Laboratory, Institute for Information Transmission Problems (Kharkevich Institute) of the Russian Academy of Sciences (IITP RAS), Moscow, Russia, received a bachelor's degree at Moscow Institute of Physics and Technology (MIPT) in 2020. Daria Senshina is currently master's student at Phystech School of Applied Mathematics and Informatics of MIPT and simultaneously obtaining MS degree in industrial and applied maths at the University Grenoble Alpes, Grenoble, France. Her research interests include computer vision and image processing.

Artem Nikonorov received a Candidate and Doctor of Sciences degree in 2005 and 2016. Now he works as a Head of AI Center of the Samara University and Head of Video Intelligence Lab of the Image Processing Systems Institute of RAS. He has more than 80 publications, two monographs. Scope of scientific interest: pattern recognition and image analysis, machine learning, computational imaging, parallel and distributed programming.

Zhihao Li is a graduate student in Vision Lab of Nanjing University, supervised by Prof Zhan Ma. His research interests are in RAW image processing and biomedical science.

Yanlin Qian is a senior camera engineer at Honor Std. in Beijing, China, before that he was a camera engineer in Huawei Camera Team. He received a PhD in computer vision from Tampere University in Finland in 2020, where he was sponsored by the Tampere University PhD scholarship and supervised by Prof Joni Kamarainen and Prof Jiri Matas. He received Nokia scholarship in his PhD study due to for his novel research on camera color. His research interests cover auto-white balance, color correction matrix and intrinsic image decomposition.

Marco Buzzelli (ORCID = https://orcid.org/0000-0003-1138-3345) obtained a bachelor degree and masters degree in computer science respectively in 2012 and 2014, focusing on image processing and computer vision tasks. He received a PhD in computer science in 2019 at the University of Milano, Bicocca, Italy, where he is currently employed as a post-doctoral researcher. His main topics of research include characterization of digital imaging devices, and object recognition in complex scenes.

Riccardo Riva obtained a bachelor degree and subsequently a masters degree in computer science at the University of Milano, Bicocca, Italy, in 2018 and 2020 respectively: both studies were completed with thesis work related to the world of image processing and neural networks. He is currently tackling the challenges of systems integration, but maintaining a strong passion for neural networks and image processing.

Simone Bianco (ORCID = https://orcid.org/0000-0002-7070-1545) obtained a PhD in computer science at DISCo (Dipartimento di Informatica, Sistemistica e Comunicazione) of the University of Milano-Bicocca, Italy, in 2010. He obtained the BS and the MS degrees in mathematics from the same university, respectively in 2003 and 2006. He is currently an associate professor and his research interests include computer vision, machine learning, optimization algorithms, and color imaging.

Raimondo Schettini (ORCID = https://orcid.org/0000-0001-7461-1451) is a Professor at the University of Milano Bicocca. He is Head of the Imaging and Vision Lab. He has been associated with the Italian National Research Council since 1987, where he led the color imaging lab from 1990 to 2002. He has been a team leader in several research projects and published more than 300 refereed papers and six patents about color reproduction, and image processing, analysis, and classification. He is a Fellow of the International Association of Pattern Recognition for his contributions to pattern recognition research and color image analysis.

Jonathan T. Barron is a staff research scientist at Google in San Francisco, California, USA, where he works on computer vision and machine learning. He received a PhD in computer science from the University of California, Berkeley in 2013, where he was advised by Jitendra Malik, and he received an Honors BS in computer science from the University of Toronto, Canada, in 2007. He received a National Science Foundation Graduate Research Fellowship in 2009, the C.V. Ramamoorthy Distinguished Research Award in 2013, the PAMI Young Researcher Award in 2020, and the ECCV Best Paper Honorable Mention in both 2016 and 2020.

Sven Lončarić received B.S., M.S., and Ph.D. degrees in 1985, 1989, and 1994, respectively. After earning his doctoral degree, he continued his academic career in the Faculty of Electrical Engineering and Computing, University of Zagreb, where he is currently a full Professor. He was an assistant professor at the Department of Electrical and Computer Engineering, New Jersey Institute of Technology, NJ, United States, from 2001 to 2003. His main areas of research are image processing and analysis. Together with his students and collaborators, he has published more than 200 publications in scientific peer-reviewed journals and has presented his work at international conferences. He is a senior member of the IEEE, director of the Center for Artificial Intelligence, and co-director of the National Center of Research Excellence for Data Science and Advanced Cooperative Systems. He is a recipient of several awards for his scientific and professional work.

Dmitry Nikolaev (ORCID = https://orcid.org/0000-0001-5560-7668) received a Master's degree in physics and Ph.D. degree in computer science from Moscow State University, Moscow, Russia, in 2000 and 2004, respectively. Since 2007, he has been the Head of the Vision Systems Laboratory, Institute for Information Transmission Problems, Russian Academy of Sciences (Kharkevich Institute), Moscow, and he has been the CTO of SmartEngines Service LLC, Moscow, since 2016. Since 2016, he has been an Associate Professor with the Moscow Institute of Physics and Technology (State University), Moscow, teaching the Image Processing and Analysis Course. He has authored over 250 scientific publications and 10 patents. His research activities are in the area of computer vision with primary application to color image understanding. Dr Nikolaev led a team of authors to win the Document Image Binarization Competition (DIBCO), in 2017.

How to cite this article: Ershov E, Savchik A, Semenkov I, et al. Illumination estimation challenge: The experience of the first 2 years. *Color Res Appl*. 2021;1–14. https://doi.org/10.1002/col.22675