Research Article | Vol. 32, No. 15 / 15 Jul 2024 / Optics Express 26708
Optics EXPRESS

# Robust pixel-wise illuminant estimation algorithm for images with a low bit-depth

SHUWEI YUE[1] AND MINCHEN WEI[1,2,*]

[1]Color, Imaging, and Illumination Laboratory, The Hong Kong Polytechnic University, Kowloon, Hong Kong SAR, China
[2]Color, Imaging, and Metaverse Research Center, The Hong Kong Polytechnic University, Kowloon, Hong Kong SAR, China
*minchen.wei@polyu.edu.hk

**Abstract:** Conventional illuminant estimation methods were developed for scenes with a uniform illumination, while recently developed methods, such as pixel-wise methods, estimate the illuminants at the pixel level, making them applicable to a wider range of scenes. It was found that the same pixel-wise algorithm had very different performance when applied to images with different bit-depths, with up to a 30% decrease in accuracy for images having a lower bit-depth. Image signal processing (ISP) pipelines, however, prefer to deal with images with a lower bit-depth. In this paper, the analyses show that such a reduction was due to the loss of details and increase of noises, which were never identified in the past. We propose a method combining the L1 loss optimization and physical-constrained post-processing. The proposed method was found to result in around 40% higher estimation accuracy, in comparison to the state-of-the-art DNN-based methods.



## 1. Introduction

Image signal processing (ISP) pipeline contains multiple stages, such as demosaicing [1], denoising [2], and white balance [3,4], to process the RAW images captured by camera sensors. White balance adjusts the RGB values of all the pixels based on the estimated illuminant, which helps to remove the color cast caused by the illumination and makes the image color appear similar to what humans perceive. Many algorithms have been developed to automatically estimate the illuminant in a scene and perform the adjustment, which is known as auto white balance (AWB).

Most AWB algorithms, such as Hu et al. [5] and Yue et al. [6], assume that there is only one illuminant in a scene. It is commonly violated in real scenes, resulting in bad performance. Thus, new algorithms have been proposed in recent years to deal with scenes containing multiple illuminants, such as [7–11]. In particular, patch-wise and pixel-wise algorithms are popular. Patch-wise algorithms (e.g., [8]) divide an image into multiple patches and estimate the illuminant for each patch, with the boundaries between patches found to cause issues. In contrast, pixel-wise algorithms (e.g., [7,9,10,12]) estimate the illuminant for each pixel, which is considered promising.

Gray pixel algorithm [13] is an early example of pixel-wise algorithms. It identifies gray pixels in an image and uses these pixels, together with a filtering process, to estimate the illuminant for each pixel. Its performance significantly depends on how accurately the gray pixels are predicted. Since the adoption of DNN in AWB algorithms in 2015 [5,6,14,15], DNN-based pixel-wise algorithms have also been developed. The GAN-based algorithm introduced in [12] becomes a milestone for pixel-wise algorithms. It uses a generative adversarial network to address the challenges of non-uniform and multiple illuminates, but sometimes introduces artifacts [10]. The collection of the Large Scale Multi-Illuminant (LSMI) dataset [7] allowed the development of more DNN-based methods, such as a U-Net method [7]. TranCC [10] method adopts a

#528359 https://doi.org/10.1364/OE.528359
Journal © 2024 Received 24 Apr 2024; revised 1 Jul 2024; accepted 1 Jul 2024; published 9 Jul 2024

Optics EXPRESS

transformer-based multi-task learning framework, which outperforms the past methods; SMM [9] uses a self-supervised technique and employs Transformer-based encoders for multi-illuminant scenes, setting a new benchmark for the LSMI dataset.

When performing analyses on pixel-wise methods, such as U-Net, we found that the performance decreases for cameras or images with lower bit-depths and for images having a single-illuminant, as illustrated in Fig. 1 and described in detail in Section 2. This is especially critical to the ISP pipeline, as a lower bit-depth is always preferred due to the higher efficiency and lower memory. In this article, we propose a new method to mitigate such a weakness, which has been found effective.

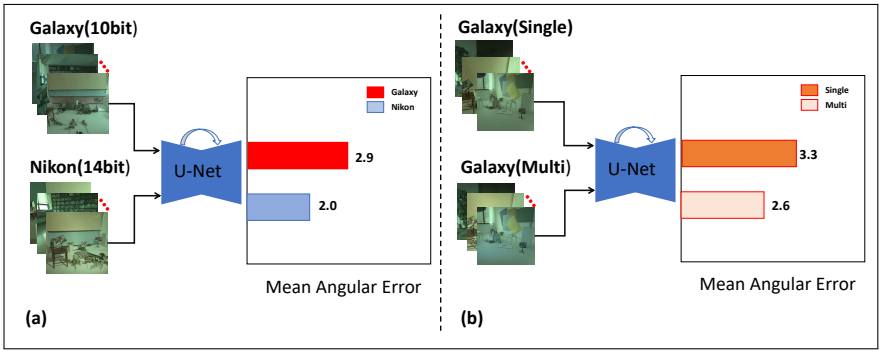


**Fig. 1.** Illustration of the lower performance of U-Net [7], when applying on cameras/images with different bit-depths and images with single and multiple illuminants. (a) Comparison between 14-bit images captured by a Nikon camera and 10-bit images captured by a Galaxy camera; (b) Comparison between 10-bit images with a single illuminant and multiple illuminants captured by a Galaxy camera.

## 2. Problem and preliminary analysis

### 2.1. Observations

Based on our initial observations, we performed an analysis. U-Net [7] and FC4 [5] methods, the state-of-the-art DNN-based pixel-wise methods, were applied on images captured by a Nikon camera in the LSMI dataset [7]. Here, FC4 was revised to consider multi-illuminant scenes with weighted output layers. To evaluate the impact of image bit-depth, the images were originally 14-bit, but were converted to 12-, 10-, and 8-bit using:

$$I_{\text{bd}} = \mathbf{R}\left(\frac{I}{2^{14}-1} \times (2^{bd}-1)\right), \tag{1}$$

where $\mathbf{R}(\cdot)$ is the rounding operation and $bd$ is the bit-depth.

It can be clearly observed that the performance of both methods, as characterized using the mean angular error between the ground truth and estimated illuminants, was worse when the image bit-depth was lower, as shown in Fig. 2.

### 2.2. Preliminary analyses and problem statement

Such an effect was speculated to be due to the loss of image details and increase of image noises, since image details play an important role in estimating the scene illuminants and are considered in various AWB methods. Thus, the gray-edge method [16], which is sensitive to image details, the white-patch [17], gray-world [18], and Cheng-PCA methods [19], which are less reliant on image details, were used to estimate the illuminants of the images in the LSMI dataset [7], with

Optics EXPRESS

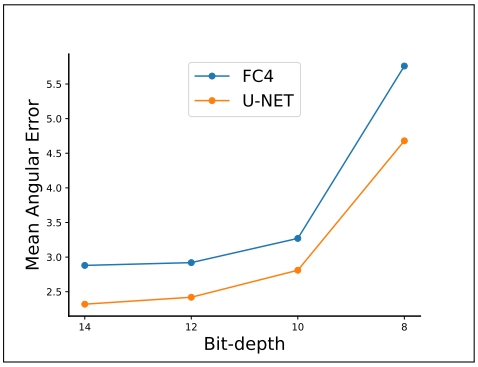


**Fig. 2.** Performance of U-Net [7] and FC4 [5] methods, the state-of-the-art DNN-based pixel-wise methods, in terms of the mean angular error on the images captured by a Nikon camera in the LSMI dataset [7] with different bit-depths.

the results shown in Fig. 3. It can be clearly observed that the performance of the gray-edge method is very sensitive to the change of image bit-depth, while the other three methods have very similar performance regardless of the bit-depths. This clearly suggests that the images having a lower bit-depth lose image details. Meanwhile, it was found that images with a lower bit-depth also have higher noises, especially in the regions with smooth changes in color or brightness, since a lower bit-depth makes it difficult to capture smooth changes. Figure 4 illustrates that an image with a lower bit-depth contains greater noises.

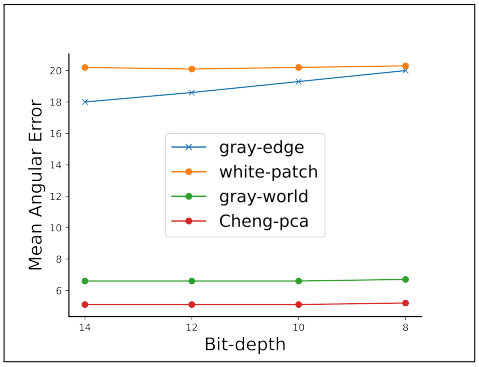


**Fig. 3.** Comparison of the performance of the methods that are sensitive to details (i.e., gray-edge) and those that are not sensitive to details on the images with different bit-depths.

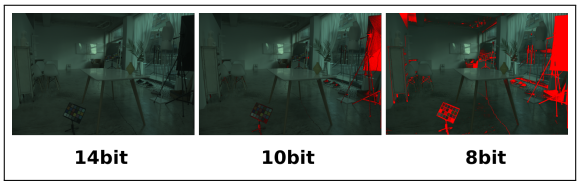


**Fig. 4.** Illustration of the increase of image noises due to the lower image bit-depth, with the noises shown in red. (Note: the noise was the difference between the 14-bit image and the images having lower bit-depths.)

Optics EXPRESS

When developing AWB algorithms, angular error is commonly selected as a loss function [5,6,8], and L2 loss is also sometimes selected [7,8]. The selection of the loss function seems to be a minor issue for developing and comparing conventional AWB algorithms for single-illuminant scenes. Our investigations, however, suggest that it seem to be critical for developing the pixel-wise algorithms that are noise-sensitive, and L1 loss seems to be less sensitive to outliers and can better handle the noises in low bit-depth images.

On the other hand, the development of AWB algorithms for practical implementations in ISP pipelines needs to consider the memory usage and computational power. Therefore, the algorithms that can be deployed on images with a lower bit-depth are always preferred, though a greater bit-depth allows to have a greater number of tones, helping to capture more details and subtle variations in a scene.

The above analyses clearly show a trade-off between image bit-depth and algorithm performance for pixel-wise AWB algorithms. Deploying these algorithms on images with a lower bit-depth in ISP pipelines can help to reduce the computational power but at the expense of accuracy.

## 3. Proposed method

With the above in mind, we propose a pixel-wise AWB algorithm to achieve a balance between accuracy and computational power for images with a lower bit-depth. The algorithm is developed based on the U-Net architecture proposed by Kim et al. [7]. L1 loss, instead of L2 loss, is used to reduce the influence of image noise. A physical-constrained post-processing technique is employed to significantly improve the estimation accuracy. Figure 5 shows the overall strategy of our proposed method. It should be noted that a large number of the images included in the LSMI dataset are indoor scenes with single or multiple LED sources, which commonly have a purplish tint as shown in the predicted illuminant maps in Fig. 5.

### 3.1. Baseline method: LSMI-U

The LSMI-U model [7], the basis of our proposed method, employs a U-Net architecture optimized for $256 \times 256$ input images. It pre-processes the images by converting RGB values into one luminance ($l$) and two chrominance values ($u, v$), as defined:

$$\begin{aligned} l &= \log(I_G + \epsilon), \\ u &= \log(I_R + \epsilon) - \log(I_G + \epsilon), \\ v &= \log(I_B + \epsilon) - \log(I_G + \epsilon). \end{aligned} \tag{2}$$

Where $I_R$, $I_G$, and $I_B$ represent the red, green, and blue channel values respectively and $\epsilon$ is a small constant for avoiding a zero value.

### 3.2. Using L1 loss for training

In our proposed method, L1 loss is used to train a U-Net model for predicting the pixel-wise illuminants. L1 loss is less sensitive to outliers and more robust to noises in low bit-depth images, as it reduces the impacts of extreme deviations that are likely to be noise artifacts, while L2 loss uses the square of the error term. The L1 loss is calculated for all the pixels in an image as follows:

$$L_{\text{model}} = \frac{1}{N} \sum_{i=1}^{N} |l_i - \hat{l}_i|, \tag{3}$$

where $N$ is the total number of pixels, $l_i$ and $\hat{l}_i$ are the ground-truth and estimated illuminants of a pixel $i$.

Optics EXPRESS

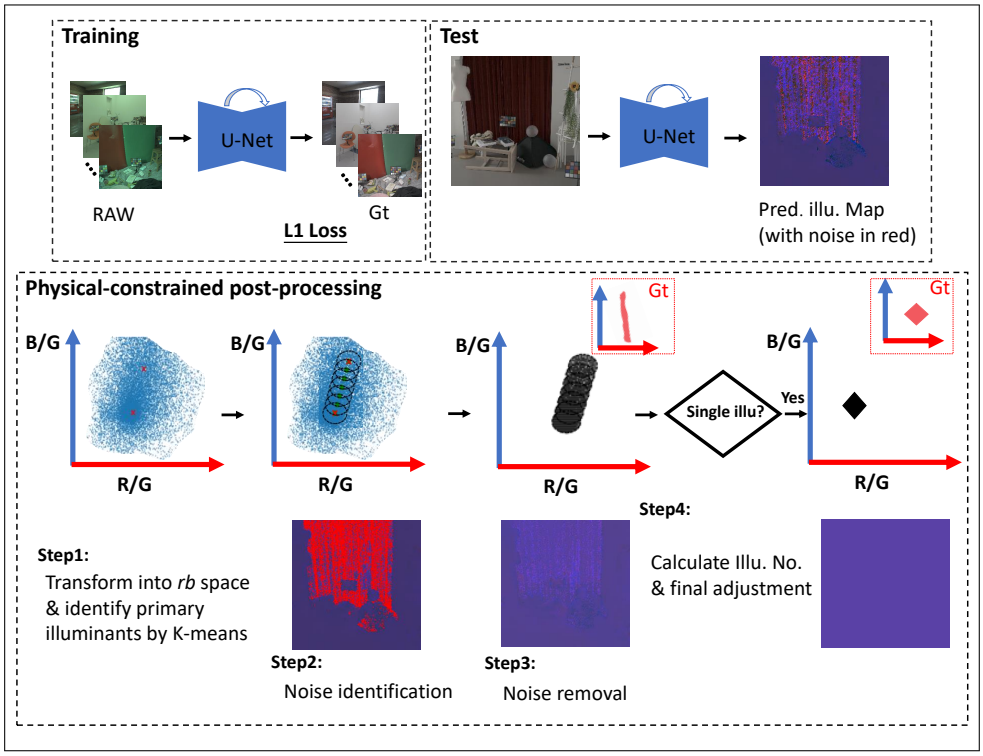


**Fig. 5.** Overview of our proposed method.

### 3.3. *Physical-constrained post-processing*

Physical-constrained processing is commonly used in AWB algorithms for illuminant estimation [20–22] and correction [23] in single-illuminant scenes. For multi-illuminant scenes, the illuminant at a pixel may be considered as a mixture of two major illuminants the scene [7] for simplification [24], expressed as $l_i = \alpha \cdot l_1 + (1-\alpha) \cdot l_2$ with $l_1$ and $l_2$ representing the two major illuminants. Such an assumption is believed to be reasonable for most common cases in practice, but not valid for all the cases.

Based on such a concept, our proposed method clusters the predicted pixel-wise illumination map to identify the two major illuminants. This helps to differentiate the pixels with and without high noises, facilitating noise removal and illuminant estimation. The proposed physical-constrained post-processing on the illumination map $\hat{l}_{\text{orig}}$, as illustrated in Fig. 5, includes four steps.

1. **Identification of two major illuminants**: $\hat{l}_{\text{orig}}$ is transformed into the *rb* space using Eq. (2), with the two major illuminants $l_1$ and $l_2$ identified using the K-means clustering method.

2. **Identification of noise**: To classify the pixels from the predicted illumination map $\hat{l}_{\text{orig}}$ into those with and without noise, we first discretize the segment between the two estimated major illuminants $l_1$ and $l_2$ into equidistant points $\{P_k\}$. Each point can be considered as the center of a tolerance range, with a circle representing the possible range of the actual illuminants of the mixture. Pixels falling outside these circles are considered noises. Such a method allows certain levels of deviations from an ideal mixture of the two major illuminants along a linear line, providing some levels of flexibility and tolerance.

Optics EXPRESS

The identification of pixels with and without noise can be expressed using the following equation:

$$\text{noisy}(p) = \begin{cases} 1 & \text{if } \forall k, \|\hat{l}_p - P_k\| > \sigma \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

In our implementation, it is found that the number of equidistant of 10 and the radius of the tolerance circle of $\sigma = 0.2$ can result in a good performance, in terms of noise reduction.

3. **Noise removal**: For removing the noise, the image is firstly divided into four quadrants (i.e., upper left, upper right, bottom left, and bottom right). The estimated illuminant of a noise pixel is replaced with the average estimated illuminant in the corresponding quadrant with all the noise pixels in the quadrant removed. Such a method is based on the premise that light propagation tends to be linear, making the localized average a better representation of the actual illuminant than the global mean:

$$\hat{l}'_p = \frac{1}{|Q|} \sum_{q \in Q} \hat{l}_q, \tag{5}$$

where $Q$ is the quadrant containing pixel $p$, and $|Q|$ is the number of the pixels in $Q$ with all the noise pixels removed.

4. **Adjustment of illuminant**: After noise removal, the angular difference between the identified major illuminants $l_1$ and $l_2$ is evaluated to judge whether the scene contains a single illuminant or not. If the angular difference is smaller than $\theta_{\text{threshold}}$, the scene is considered to have a single illuminant, and the illumination map $\hat{l}_{\text{proc}}$ is adjusted to be the average values after the noise removal, which is considered to produce more reliable estimation.

$$\hat{l}_{\text{proc}} = \begin{cases} \frac{1}{N} \sum_{p \in \text{non-noisy}} \hat{l}'_p & \text{if } \angle(l_1, l_2) < \theta_{\text{threshold}} \\ \hat{l}'_p & \text{otherwise} \end{cases} \tag{6}$$

In our implementation, $\theta_{\text{threshold}}$ is set to 3 degrees.

By following these steps, the original predicted illumination map $\hat{l}_{\text{orig}}$ is revised to $\hat{l}_{\text{proc}}$, which achieves a balance between performance and memory size.

## 4. Experiment and results

### 4.1. Benchmark dataset: LSMI dataset

The LSMI dataset [7] was used to evaluate the performance of the proposed method. It includes 7,486 images from around 3,000 real scenes, captured by three distinct cameras under various lighting conditions. The images also have different bit-depths, such as 14-bit for those captured by a Nikon DSLR, 12-bit for those captured by a Sony camera, and 10-bit for those captured by a Samsung Galaxy Note 20 Ultra. Such different bit-depths allow us to test the performance of the proposed method. For the pixel-wise ground truth illuminants in the LSMI dataset, interpolation between two illuminants was used, which is a common approach due to the difficulty in collecting accurate pixel-wise illuminant data. Such an interpolation method is based on the physical principles and considered accurate for most scenes with gradual changes between illuminants.

### 4.2. Experiments settings

Two strategies (i.e., training from scratch and fine-tuning from LSMI-U [7]), were explored, with both using the L1 loss. For the from-scratch approach, the model underwent 1000 epochs

Optics EXPRESS

of training with an initial learning rate of $5 \times 10^{-4}$. For the fine-tuning approach, the model underwent 500 epochs with a learning rate of $5 \times 10^{-5}$. The experiment was carried out using an NVIDIA GeForce RTX 4090 GPU.

### 4.3. Error function

The mean angular error (MAE) across all the pixels in an image was employed as the error function. The angular error characterizes the discrepancy between the predicted and ground-truth illuminants for each pixel, with the mean of the angular errors characterizing the entire image:

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}\frac{180}{\pi}\arccos\left(\frac{l_i \cdot \hat{l}_i}{\|l_i\| \cdot \|\hat{l}_i\|}\right), \tag{7}$$

where $N$ is the total number of pixels in an image, $l_i$ is the ground-truth illuminant for a pixel $i$, and $\hat{l}_i$ is the estimated illuminant for the same pixel. Vectors $l_i$ and $\hat{l}_i$ are normalized to a unit length, and the error is presented in degrees.

### 4.4. Results

Table 1 summarizes the angular errors derived from various AWB methods, including the traditional methods (i.e., gray-world [18], white-patch [17], gray-edge [16], and Cheng-PCA [19]), DNN-based methods (i.e., patch-CNN [8], angular-GAN [12], SMM [9], TRCC [10], and LSMI-U [7]), and our proposed methods. It can be observed that our proposed methods had better performance, especially when the physical-constrained post-processing was included. In addition, the fine-tuning strategy resulted in slightly better results than the training from scratch strategy, which was likely to due to the initial L2 loss training for bringing a more stable global optimum.

**Table 1. Summary of the performance of various AWB methods, including the traditional methods (i.e., gray-world [18], white-patch [17], gray-edge [16], and Cheng-PCA [19]), DNN-based methods (i.e., patch-CNN [8], angular-GAN [12], SMM [9], TRCC [10], and U-Net [7]), and our proposed method, with the best results highlighted in bold.**

| Method | Mean | Median | Worst%25 | Best%25 |
|---|---|---|---|---|
| gray-world | 11.3 | 8.8 | 20.7 | 4.9 |
| white-patch | 12.8 | 14.3 | 23.5 | 5.6 |
| gray-edge (1st) | 12.1 | 10.8 | 22.6 | 5.3 |
| Cheng-PCA | 10.9 | 10.7 | 20.8 | 4.8 |
| patch-CNN | 4.82 | 4.24 | - | - |
| angular-GAN | 4.69 | 3.88 | - | - |
| SMM | 2.73 | 2.10 | - | - |
| TRCC | 2.78 | 2.15 | - | - |
| LSMI-U | 2.63 | 1.91 | 5.01 | 1.12 |
| U-L1 (from scratch) | 2.50 | 2.10 | 4.91 | 1.13 |
| +Post-processing | 2.00 | 1.67 | 4.0 | 0.43 |
| U-L1 (fine-tuned) | 2.23 | 1.76 | 4.23 | 0.89 |
| +Post-processing | **1.80** | **1.47** | **3.70** | **0.40** |

The proposed method, together with the state-of-the-art DNN methods (i.e., SMM [9] and LSMI-U [7])), was evaluated on images with different bit-depths from 10-bit to 14-bit, with the

Optics EXPRESS

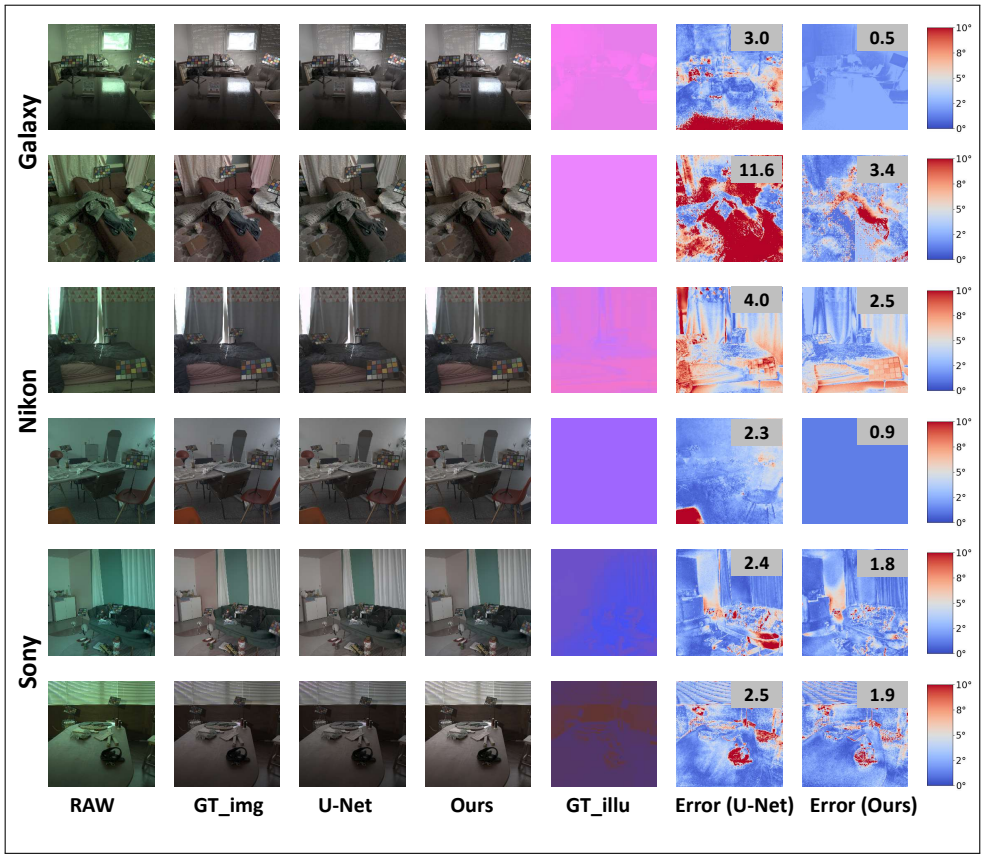


**Fig. 6.** Examples of the images processed with U-Net and our proposed method. From left to right: input RAW image (RAW), ground-truth image (GT_img), LSMI-U corrected image (U-Net), our proposed method corrected image (Ours), ground-truth illumination map (GT_illu), angular error map between U-Net estimation and ground-truth (Error U-Net), and angular error map between our proposed method estimation and ground-truth (Error Ours). The mean angular error value is shown at the top right corner of each error map. (Note: the images are shown in a nonlinear gamma space for better visualization).

**Table 2. Summary of the performance, in terms of the mean and median angular errors, of the SMM, U-Net, and our proposed method on images with the different bit-depths in the LSMI dataset, with the best performance highlighted in bold.**

| Method | 10 bit (Galaxy) | | | | | | 14 bit (Nikon) | | | | | | 12 bit (Sony) | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Single | | Multi | | Mixed | | Single | | Multi | | Mixed | | Single | | Multi | | Mixed | |
| | Mean | Median | Mean | Median | Mean | Median | Mean | Median | Mean | Median | Mean | Median | Mean | Median | Mean | Median | Mean | Median |
| SMM [9] | 3.20 | 2.36 | 2.63 | 2.33 | 2.88 | 2.33 | 1.60 | 1.20 | 2.48 | 1.99 | 2.05 | 1.49 | 2.60 | 2.29 | 2.98 | 2.69 | 2.82 | 2.49 |
| LSMI-U [7] | 3.31 | 1.90 | 2.56 | 2.09 | 2.90 | 1.91 | 1.52 | 1.19 | 2.36 | 1.76 | 1.96 | 1.44 | 2.76 | 2.36 | 3.00 | 2.77 | 2.90 | 2.62 |
| **Ours** | **1.38** | **0.77** | **2.28** | **1.91** | **1.87** | **1.57** | **1.01** | **0.65** | **2.24** | **1.68** | **1.65** | **1.25** | **1.11** | **0.74** | **2.36** | **2.01** | **1.81** | **1.57** |

results summarized in Table 2 and examples of images shown in Fig. 6. It can be clearly observed that the proposed method outperformed the two state-of-the-art methods, and it was much less sensitive to the change of the image bit-depths. We also performed such tests and analyses on images captured by a same camera (i.e., Nikon) with the image bit-depth changed using Eq. (1), and found similar results.

Optics EXPRESS

## 5. Discussions

### 5.1. Performance on single-illuminant dataset

To further validate the robustness of our proposed method, an additional experiment was carreid out using the Cube+ dataset [25]. The dataset includes a large number of high-quality images with a single illuminant, which was specifically collected for developing DNN-based methods and provides a reliable benchmark for evaluating different methods. We used three-fold cross-validation and standard angular error metrics as described in [6,19]. The predicted illumination map was used following [10], with the mean of the map used for the LSMI-U data. For our proposed method, we used the L1 loss for training to derive the prediction, applied the noise-free strategy to obtain the noise-free map, and then calculated the mean as the final predicted illuminant. The results, as summarized in Table 3, show that our proposed method results in smaller mean angular error for single-illuminant scenes. The smaller median and worst 25% results of the TRCC method were likely due to the use of contrastive learning for data augmentation. The improvements introduced by our proposed method on the Cube+ dataset seem to be less significant, in comparison to those on the LSMI dataset. This is believed due to the fact that the images included in the Cube+ dataset were mainly captured in daytime with good lighting conditions, making the impact of lower bit-depths less pronounced.

**Table 3. Summary of the performance on the Cube+ dataset for single-illuminant scenes, with the best performance highlighted in bold.**

| Method | Mean | Median | Worst 25% | Best 25% |
|---|---|---|---|---|
| gray-world | 7.71 | 4.29 | 20.19 | **1.01** |
| white-patch | 9.69 | 7.48 | 20.49 | 1.72 |
| TRCC | 2.71 | **1.59** | **3.21** | 1.05 |
| LSMI-U | 2.68 | 2.14 | 3.47 | 1.19 |
| L1+Post-processing | **2.54** | 1.94 | 3.33 | 1.10 |

### 5.2. Comparison to noise removal using filters

Filters are commonly used in image processing to remove noises, so a comparison was made between our proposed method and the implementation of two filters (i.e., Gaussian and median filters). The results, as summarized in Table 4, clearly suggest the better performance of our proposed method. In particular, a median filter was able to reduce the mean angular error by 5% and a Gaussian filter did not introduce any improvement at all, which was likely due to the prevalence of salt-and-pepper noise. In contrast, our proposed method was able to introduce a significant reduction of the mean angular error, with the adoption of the physical-constrained post-processing effectively reducing the mean angular error by 20%. When combined with the L1 loss optimization, the mean angular error was reduced by 30%, which clearly illustrates the effectiveness of the proposed method.

### 5.3. Limitations

Though the results above clearly show the effectiveness of our proposed method in improving the accuracy of illuminant estimation, it relies heavily on the results of the initial illuminant estimation $\hat{l}_{\text{orig}}$. When the initial estimation has poor performance, the proposed method adopting the L1 loss optimization and physical-constrained post-processing may have little effect. This highlights the importance of the initial estimation.

Additionally, the proposed method is based on the assumption that any pixel in a multi-illuminant scene can be considered as a mixture of two major illuminants, which is believed to be

Optics EXPRESS

**Table 4. Comparison of the performance, in terms of the mean and median angular errors, of using median filters, Gaussian filters, and our proposed method on the images in the LSMI dataset. (*s* represents the filter size and $\sigma$ is the standard deviation of the Gaussian filter.)**

| Method | Mean | Median |
|---|---|---|
| Baseline (LSMI-U) | 2.63 | 1.91 |
| Median filter (s=5) | 2.53 | 1.80 |
| Median filter (s=9) | 3.54 | 1.78 |
| Gaussian filter (s=10, $\sigma$ = 0.05) | 3.20 | 1.92 |
| Gaussian filter (s=50, $\sigma$ = 0.25) | 6.07 | 2.07 |
| LSMI-U+Post-processing | 2.09 | 1.78 |
| L1+Post-processing | **1.80** | **1.47** |

a reasonable assumption for most common scenes in real life. This can also be supported by the good performance when applied to the scenes containing three illuminants in the LSMI dataset. Future work, however, is still needed to explore scenes containing more illuminants to allow a good performance for a wider range of conditions.

## 6. Conclusion

The performance of pixel-wise AWB algorithms was found to vary with image bit-depths. We performed an analysis and found that they tend to have worse performance on images with a lower bit-depth. This was found due to the increase of noises and loss of details in the images with a lower bit-depth. Given the fact that a lower bit-depth is always preferred in ISP pipelines, we proposed a method using an L1 loss optimization and a physical-constrained post-processing. The proposed method was found to outperform the traditional and the state-of-the-art DNN-based methods, especially on images with a lower bit-depth.

**Funding.** National Natural Science Foundation of China (62222516).

**Disclosures.** The authors declare no conflicts of interest.

**Data availability.** No data were generated or analyzed in the presented research.

## References

1. W. Choi, H. S. Park, and C. M. Kyung, "Color reproduction pipeline for an RGBW color filter array sensor," Opt. Express **28**(10), 15678–15690 (2020).
2. X. Yao, W. Yu, X. Liu, *et al.*, "Iterative denoising of ghost imaging," Opt. Express **22**(20), 24268–24275 (2014).
3. S. Gao, M. Zhang, and Y. Li, "Improving color constancy by selecting suitable set of training images," Opt. Express **27**(18), 25611–25633 (2019).
4. R. Ma, Q. Gao, Y. Qiang, *et al.*, "Robust categorical color constancy along daylight locus in red-green color deficiency," Opt. Express **30**(11), 18571–18588 (2022).
5. Y. Hu, B. Wang, and S. Lin, "Fc4: Fully convolutional color constancy with confidence-weighted pooling," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2017), pp. 4085–4094.
6. S. Yue and M. Wei, "Color constancy from a pure color view," J. Opt. Soc. Am. A **40**(3), 602–610 (2023).
7. D. Kim, J. Kim, S. Nam, *et al.*, "Large scale multi-illuminant (LSMI) dataset for developing white balance algorithm under mixed illumination," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2021), pp. 2410–2419.
8. S. Bianco, C. Cusano, and R. Schettini, "Single and multiple illuminant estimation using convolutional neural networks," IEEE Trans. on Image Process. **26**(9), 4347–4362 (2017).
9. Z. Feng, Z. Xu, H. Qin, *et al.*, "Smm: Self-supervised multi-illumination color constancy model with multiple pretext tasks," in *Proceedings of the 31st ACM International Conference on Multimedia*, (2023), pp. 8653–8661.
10. S. Li, J. Wang, M. S. Brown, *et al.*, "Transcc: Transformer-based multiple illuminant color constancy using multitask learning," arXiv, (2022).

Optics EXPRESS


11. I. Domislović, D. Vršnjak, M. Subašić, *et al.*, "Color constancy for non-uniform illumination estimation with variable number of illuminants," Neural Comput. & Applic. **35**(20), 14825–14835 (2023).
12. O. Sidorov, "Conditional gans for multi-illuminant color constancy: Revolution or yet another approach?" in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, (2019), pp. 0–0.
13. K. Yang, S. Gao, and Y. Li, "Efficient illuminant estimation for color constancy using grey pixels," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2015), pp. 2254–2263.
14. S. Bianco, C. Cusano, and R. Schettini, "Color constancy using CNNS," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, (2015), pp. 81–89.
15. S. Yue and M. Wei, "Effective cross-sensor color constancy using a dual-mapping strategy," J. Opt. Soc. Am. A **41**(2), 329–337 (2024).
16. J. V. D. Weijer, T. Gevers, and A. Gijsenij, "Edge-based color constancy," IEEE Trans. on Image Process. **16**(9), 2207–2214 (2007).
17. E. H. Land, "The retinex theory of color vision," Sci. Am. **237**(6), 108–128 (1977).
18. G. Buchsbaum, "A spatial processor model for object colour perception," J. Franklin Inst. **310**(1), 1–26 (1980).
19. D. Cheng, D. K. Prasad, and M. S. Brown, "Illuminant estimation for color constancy: why spatial-domain methods work and the role of the color distribution," J. Opt. Soc. Am. A **31**(5), 1049–1058 (2014).
20. P. V. Gehler, C. Rother, A. Blake, *et al.*, "Bayesian color constancy revisited," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2008), pp. 1–8.
21. G. D. Finlayson, S. D. Hordley, and P. M. Hubel, "Color by correlation: A simple, unifying framework for color constancy," IEEE Trans. Pattern Anal. Machine Intell. **23**(11), 1209–1221 (2001).
22. X. Zhang, S. Yue, R. Gong, *et al.*, "Illumination estimation based on a weighted color distribution," Optik **185**, 965–971 (2019).
23. G. D. Finlayson, "Corrected-moment illuminant estimation," in *Proceedings of the IEEE International Conference on Computer Vision*, (2013), pp. 1904–1911.
24. D. Cheng, A. Abdelhamed, B. Price, *et al.*, "Two illuminant estimation and user correction preference," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2016), pp. 469–477.
25. N. Banić, K. Koščević, and S. Lončarić, "Unsupervised learning for color constancy," arXiv, (2017).