Journal of Ambient Intelligence and Humanized Computing (2022) 13:5151–5162
https://doi.org/10.1007/s12652-020-02367-3

ORIGINAL RESEARCH

Sun-sky model estimation from outdoor images

Xin Jin[1] · Pengyue Deng[2] · Xinxin Li[2] · Kejun Zhang[3] · Xiaodong Li[3] · Quan Zhou[4] · Shujiang Xie[5] · Xi Fang[3]

Received: 15 April 2020 / Accepted: 21 July 2020 / Published online: 5 August 2020

Abstract
When a virtual object is inserted into an outdoor image, the recovery of scene illumination has a critical effect on the mix of virtual objects and actual reality. There are two main parts of the object in the outdoor scene: the sun and the sky. In order to represent the illumination conditions of these two natural illumination, this paper uses the Lalonde-Matthew outdoor illumination model to perform the sky and sun in the image. Model use seven parameters represent the illumination of the scene. So the original illumination estimation problem is transformed into a prediction problem of seven illumination parameters. For this problem, this paper proposes a new two-branch network structure, one branch is used to estimate the sun orientation, and the other branch is used to estimate the remaining six parameters. This paper also introduces convolution block attention module (CBAM) based on this structure. The introduction of this module enables the network to select the most important information for the current task target from a large number of information when extracting image features, while suppressing other useless information.

Keywords Outdoor illumination estimation · Convolution block attention module · Two-branch network structure

1 Introduction

Illumination determines the appearance of outdoor scenes, such as the color, brightness of the scene, and the shadows on the ground. Illumination recovery of outdoor scenes has a wide range of applications in many fields ranging from film post-production (Zhou et al. 2015), virtual military exercises (Zhao 2009), graphic design to video game design,vir-tual reality (Lele 2013), augmented reality (Erra and Capece 2019) (Wang et al. 2018a), recover image (Li et al. 2015), image restoration and de-blurring (Lu et al. 2016) (Li et al. 2018). All these applications involve inserting virtual objects into physical scenes. In order to achieve realistic appearance of the mix of virtual objects and actual scenes, the illumination estimation has become a very important task.

However, estimating illumination from a single outdoor image is extremely difficult. In real scenes, the image formation process is affected by many factors, such as material properties, camera parameters, and scene illumination, in extremely complicated ways. In addition, multiple

✉ Xin Jin
jinxinbesti@foxmail.com

Pengyue Deng
dengpy@stu.xidian.edu.cn

Xinxin Li
2948722748@qq.com

Kejun Zhang
zkj@besti.edu.cn

Xiaodong Li
lxdbesti@163.com

Quan Zhou
quan.zhou@njupt.edu.cn

Shujiang Xie
xieshujiang@126.com

Xi Fang
xfang@besti.edu.cn

[1] Beijing Electronic Science and Technology Institute, State Key Laboratory of Cryptology, P.O. Box 5159, Beijing 100878, China

[2] Xidian University, Xi'an 710000, China

[3] Beijing Electronic Science and Technology Institute, Beijing, China

[4] Nanjing University of Posts and Telecommunications, Nanjing, China

[5] School of Economics, Minzu University of China, Beijing, China

combinations of these factors can exert same effect on the image formation, resulting in an uncertainty of the estimated illumination.

A simple and straightforward way to capture scene illumination is to place light probes within the scene, such as mirror spheres (Stumpfel et al. 2006) (Debevec 2008)and integrating spheres. However, this method is not feasible because most of the images we use are not taken in such a scene with light probes, and it is usually impossible for us to place probes within the scene. Another commonly used approach is based on handcrafted features, including sky regions, shading, shadows (Kim and Hong 2005) or combinations of them (Lalonde et al. 2012). These features can convey a lot of information about scene illumination; The sky area can indicate the weather conditions and give us some clues about the sun orientation; The shadows and shading can give us a strong indication of the sun orientation. However, this method has the disadvantage of low prediction accuracy of illumination parameters

Recently, deep learning has developed rapidly and successfully solved many problems in computer vision such as object tracking (Yan et al. 2019) (Li et al. 2019). Applying deep learning to illumination estimation has also become a research hotspot in computer vision. Many works have adopted deep neural networks to estimate illumination from a single outdoor image (Jin et al. 2019) (Hold-Geoffroy et al. 2017) (Ma et al. 2017), or indoor image (Gardner et al. 2017) (Weber et al. 2018), and produced promising results. Nevertheless, how to achieve illumination estimation with high quality and high accuracy from a single image is still worth exploring.

In this paper, we propose a new two-branch network and embed the convolutional block attention modules into this network. Experiments verify the effectiveness of the attention modules and the effect of the number of attention modules on the prediction results. Compared with the state-of-art methods, our method produces results with higher precision, and the recovered illumination is more realistic.

The main contributions of this paper are as follows:

(1) We propose a new end-to-end learning approach based on a new two-branch network and the convolutional block attention module, which achieves higher prediction precision than the state-of-art methods.
(2) This paper proposes a new two-branch network structure, one branch is used to estimate the sun orientation, and the other branch is used to estimate the remaining six parameters.
(3) This paper also introduces the convolution block attention module (CBAM) in this network structure. The introduction of this module enhances the network's feature expression ability and improves the prediction accuracy of illumination parameters. The arrangement of this paper is as follows: Sect. 2 describes related works; Sect. 3 introduces the approach employed in this paper; Sect. 4 shows the experiments; Sect. 5 summaries the conclusion.

2 Related works

A lot of research work have explored this problem and proposed many methods. These methods can be divided into two main categories: traditional methods and the methods based on deep learning.

2.1 Traditional method

Some methods fit a physically-based sky model (Hosek and Wilkie 2012; HošekHošek and Wilkie 2013; Perez et al. 1993; Preetham et al. 1999), which describes the distribution of sky luminance, to the sky area in the image to recover some illumination parameters, such as sun orientation and atmospheric turbidity.

Light probe is a simple and direct way to capture luminance. Some methods capture high dynamic range scene illumination based on mirrored spheres (Stumpfel et al. 2006), and several photographs taken at different exposures.

Lalonde et al. (2012) combined several weak cues, including the shadows, the sky regions, and the shading to generate more reliable and accurate cues. These cues combine some priors computed over a large data set to recover illumination parameters, such as sun orientation and atmospheric turbidity.

Some methods rely on priors on scene geometry, illumination, and reflectance (Barron and Malik 2014; Lombardi and Nishino 2015), to recover illumination from an image. However, these priors are only applicable to specific scenes and cannot be used in other scenes, which limits the application of this method.

2.2 Method based on deep learning

Ma et al. (2017) proposed to directly predict the sun orientation from an input image based on deep learning. The network used was adapted from AlexNet, and different loss functions were used.

Jin et al. (2019) proposed to add short-cuts to deep neural network to achieve the contact of low-level features and high-level features, which can enhance the extracted image features from an image.

A method based on CNN to recover high dynamic range illumination from low dynamic range panoramas was proposed by Hold-Geoffroy et al. (2017), A physically-based sky model (Hosek and Wilkie 2012; HošekHošek and Wilkie 2013) was fitted to the sky regions of panoramas to generate the illumination parameters such as sun orientation, atmospheric turbidity, and camera parameters. Then these parameters were used to train the CNN in accompany with the images extracted from panoramas.

Cheng et al. (2018) proposed to estimate scene illumination from paired images, which were captured by front and rear cameras. Illumination was represented as spherical harmonic lighting (SHL) (Green 2003) rather than illumination parameters. The method first needs to calculate the spherical harmonic coefficients corresponding to each picture and use them as label data. Then use pairs of images and their corresponding label data to train the deep neural network. This method is not only suitable for outdoor images but also indoor images.

3 Approach

The outdoor illumination estimation method proposed in this paper is shown in Fig 1. Next, this paper will introduce the content involved in this method in detail. Section 3.1 describes in detail how to use the Lalonde–Matthews outdoor illumination model (Lalonde and Matthews 2014) to represent illumination; Sect. 3.2 mainly introduces how to use the SUN360 dataset to generate the dataset required for this paper; Sect.3.3 details the two-branch network structure proposed in this paper; Sect. 3.4 mainly introduces the use of the attention module force of the convolution block; Sect. 3.5 is the loss function.

3.1 Lalonde–Matthews outdoor illumination model

The Lalonde–Matthews outdoor illumination model (Lalonde and Matthews 2014) is a parametric model, which assumes that the only light source is the sun and the sky in an outdoor scene, and ignores the effects of local illumination, such as light reflected from the ground or the surface of nearby objects. Thus, it express outdoor illumination as the sum of the sun and sky light sources, where the sun is a bright and collimated light source and the sky is a low-frequency hemispherical light source. The expression is as follows:

$$f_h^c(l, q_h) = f_{sun}^c(l, q_{sun}, l_{sun}) + f_{sky}^c(l, q_{sky}, l_{sky}) \quad (1)$$

where, $q_h = [q_{sun}\ q_{sky}\ l_{sun}], q_{sky} = [w_{sky}^c\ t], q_{sun} = [w_{sun}^c\ \beta\ k]$, $l_{sun} = [\theta_{sun}\ \phi_{sun}]$, θ_{sun} represents the azimuth angle of the sun, and ϕ_{sun} represents the zenith angle of the sun.

Finally, the Lalonde-Matthews outdoor illumination model (Lalonde and Matthews 2014) can be represented by a 6-dimensional parameterized vector with the following expression:

$$q_{LM} = \{w_{sky}, t, w_{sun}, \beta, k, l_{sun}\} \quad (2)$$

In addition, this paper also adds the camera vertical field of view parameter to this parameter vector. This parameter is generated during the dataset generation process and can be used to control the camera during rendering.

As a result, the problem of illumination estimation from a single outdoor image is transformed into the estimation of seven illumination parameters.

3.2 Dataset generation

This paper is mainly based on the SUN360 dataset (Xiao et al. 2012) to generate the data needed for this part. The SUN360 dataset (Xiao et al. 2012) mainly includes 360-degree panoramic images of outdoor and indoor scenes. This paper mainly uses outdoor images, and some of them are shown in Fig 2.

For some of the label information in this dataset, namely the 6 illumination parameters in the Lalonde-Matthews outdoor illumination model (Lalonde and Matthews 2014), this paper mainly uses the data generated by Zhang et al.

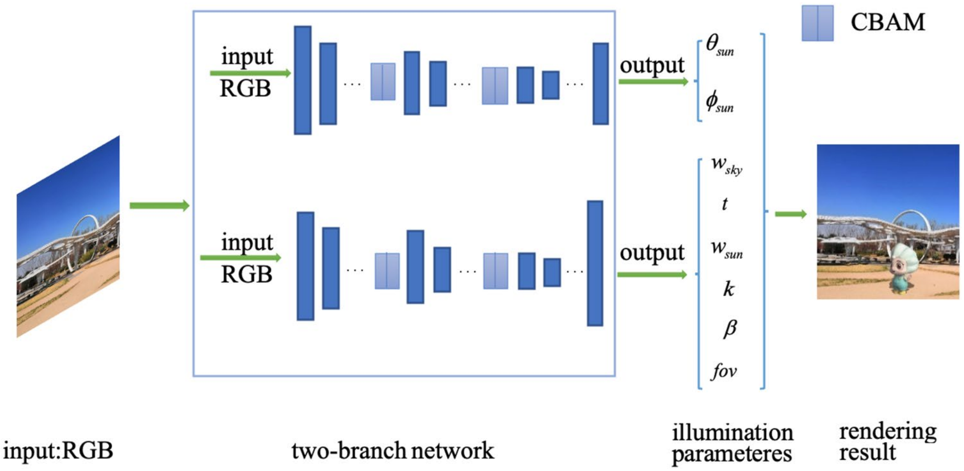


Fig. 1 The outdoor illumination estimation method proposed in this paper

Fig. 2 Outdoor panorama images in part of the SUN360 dataset

(2019) in their paper. However, the information of label data is not comprehensive, and only contains the five parameters and lacks parameters of the sun orientation. Aiming at this problem, this paper adopts the method proposed by Hold-Geoffroy et al. (2017) to calculate sun orientation in the panoramic image. This method uses the centroid of the region of maximum saturation in the image sky region as the sun orientation.

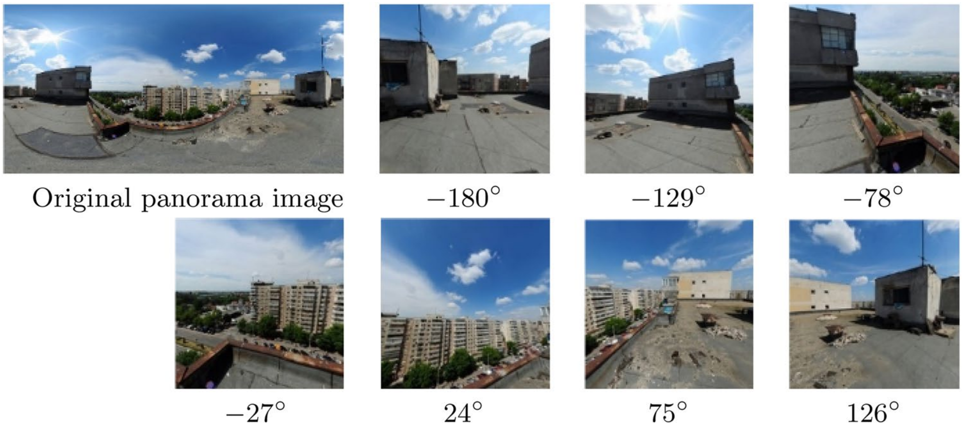
Through the above calculations, a total of 22, 126 outdoor panoramas and their corresponding 7 illumination parameter label data are obtained. Next, this paper took 7 images from each panorama, as shown in Fig 3. 7 images corresponding to the azimuth of the camera are:$-180°, -129°, -78°, -27°, 24°, 75°, 126°$, The camera zenith angle and camera vertical field of view of each image are randomly selected between the interval $[-20°, 20°]$ and $[20°, 70°]$, and the size of the image is 256×256.

After the above operations, a total of 154,882 images and corresponding seven illumination parameters were obtained. In order to facilitate the experiment, this paper divides it into training set, validation set and test set according to the ratio of 8:1:1. Among them, the scene is randomly selected for these three sets, but 7 images of a scene must appear in a set at the same time. A total of 123,906 training set images, 15,488 verification set images, and 15,488 test set images were obtained. At the same time, the label data is also split using the same method.

3.3 Two-branch network

Most of illumination estimation methods based on deep learning predict illumination parameters, or use a fully connected layer to output all the parameters in the last layer of the network, or use multiple fully connected layers to output multiple parameters. A similar method is used in this paper, but the prediction results obtained are not good. After carefully analysis and experimental verification of these 7 illumination parameters, this paper proposes the reasons for this phenomenon. The sun orientation and the remaining 6 parameters used different loss functions. The values of the two loss functions are quite different. Although they are assigned different weight values, they cannot produce good

Fig. 3 Seven images taken from the panorama



results. Based on the above reason, this paper proposes a new two-branch network structure, as shown in Fig 4.The detailed description of the structure is shown in Table 1.

The first branch in the network structure is mainly used to predict sun orientation. Its input is the three channels (*RGB*) of the original image. There are 11 convolutional layers and one fully connected layer but no pooling layer in this branch. Convolutional layers are used instead of the pooling layer to complete the downsampling operation. The last fully connected layer is used to output the sun zenith angle and sun azimuth. The second branch is mainly used to predict the remaining six parameters. Its input is the three channels (*RGB*) of the original image. The structure of this branch is the same as the first branch, except that the last layer is fully connected and the output is 6 parameters. Except for the last layer in both branches, all convolutional layers are followed by BatchNormation (Zhang et al. 2018) and Relu activation function. During the training process, these two branches are independent of each other and alternately perform weight update operations. They do not share weight parameters and

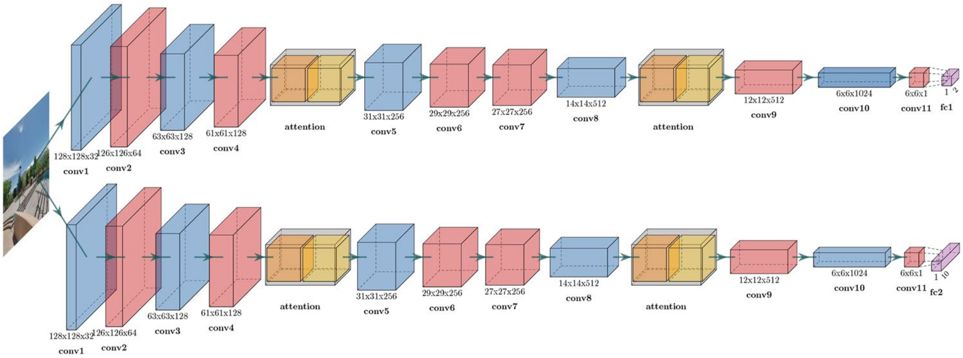


Fig. 4 Network structure

Table 1 Detailed description of the two-branch network structure

Input	256 × 256 × 3	
	branch-1: 256 × 256 × 3	branch-2: 256 × 256 × 3
1	conv: kernel = (3, 3) stride = 2 filter = 32 output: 128 × 128 × 32	conv: kernel = (3, 3) stride = 2 filter = 32 output: 128 × 128 × 32
1	conv: kernel = (3, 3) stride = 1 filter = 64 output: 126 × 126 × 64	conv: kernel = (3, 3) stride = 1 filter = 64 output: 126 × 126 × 64
1	conv: kernel = (3, 3) stride = 2 filter = 128 output: 63 × 63 × 128	conv: kernel = (3, 3) stride = 2 filter = 128 output: 63 × 63 × 128
1	conv: kernel = (3, 3) stride = 1 filter = 128 output: 61 × 61 × 128	conv: kernel = (3, 3) stride = 1 filter = 128 output: 61 × 61 × 128
	CBAM	CBAM
1	conv: kernel = (3, 3) stride = 2 filter = 256 output: 31 × 31 × 256	conv: kernel = (3, 3) stride = 2 filter = 256 output: 31 × 31 × 256
1	conv: kernel = (3, 3) stride = 1 filter = 256 output: 29 × 29 × 256	conv: kernel = (3, 3) stride = 1 filter = 256 output: 29 × 29 × 256
1	conv: kernel = (3, 3) stride = 1 filter = 256 output: 27 × 27 × 256	conv: kernel = (3, 3) stride = 1 filter = 256 output: 27 × 27 × 256
1	conv: kernel = (3, 3) stride = 2 filter = 512 output: 14 × 14 × 512	conv: kernel = (3, 3) stride = 2 filter = 512 output: 14 × 14 × 512
	CBAM	CBAM
1	conv: kernel = (3, 3) stride = 1 filter = 512 output: 12 × 12 × 512	conv: kernel = (3, 3) stride = 1 filter = 512 output: 12 × 12 × 512
1	conv: kernel = (3, 3) stride = 2 filter = 1024 output: 6 × 6 × 1024	conv: kernel = (3, 3) stride = 2 filter = 1024 output: 6 × 6 × 1024
1	conv: kernel = (1, 1) stride = 1 filter = 1 output: 6 × 6 × 1	conv: kernel = (1, 1) stride = 1 filter = 1 output: 6 × 6 × 1
	fc: output: 2	fc: output: 10

have different hyperparameters, such as initial learning rate, number of iterations, and so on.

3.4 Convolution block attention module

Attention mechanism has been widely used in various fields of deep learning in recent years. Attention mechanism can be seen in various types of tasks such as natural language processing, speech recognition or image processing. This mechanism was first proposed in the field of computer vision. It mimics the general process of humans observing the behavior of things. When humans observe something at a certain time, their attention will generally focus on a certain focus on the thing. Area and ignore the rest. Similar to this, the attention mechanism in the field of deep learning is essentially a weight allocation model. The key information for the current task will be assigned a larger weight, and the useless information will be assigned a smaller weight. It depends on the current application scenario.

At present, in the field of image illumination estimation based on deep learning, there are few methods to use the attention mechanism, and this paper has tried to introduce this mechanism and achieved some results. As shown in Fig 4, this paper adds the attention module after the 4th and 8th convolutional layers in the first network branch and the second network branch, respectively. Because there are many attention mechanisms in the field of computer vision, such as self-attention mechanism (Zhang et al. 2018; Wang et al. 2018), spatial domain attention mechanism (Jaderberg et al. 2015), channel domain attention mechanism (Hu et al. 2018) and mixed domain attention mechanism (Woo et al. 2018) and many more. Different types of attention mechanisms are suitable for different visual tasks. In order to find the attention mechanisms that are more suitable for image illumination estimation tasks, this paper conducted a comparison experiment between various attentions, and finally selected Woo et al. (2018). The introduction of this module has improved the feature expression ability of two-branch networks.

3.5 Loss function

During the training of the two-branch network, two loss functions are used: one is the cosine distance loss function for the estimation of the sun orientation; the other is the mean square error (MSE) loss function for the remaining six parameters.

For the second loss function, this paper first performs some preprocessing operations on the label data before its definition. Due to the uneven distribution of the values of some parameters, the values vary widely and contain a small number of extreme data values. For example, the range of the smaller data is between [0, 1] and the value range of big data is between [30, 50] in the parameter β. If such raw data is used to train the network directly, the loss value will cause large oscillations, the network will be difficult to converge, and eventually a poor result will be generated. To solve this problem, this paper uses the average and variance of each parameter to normalize all data to a distribution with a mean of 0 and a variance of 1. The process can be expressed by the following formula.

$$x^{'} = \frac{x - x_{mean}}{x_{std}} \tag{3}$$

where, x indicates the original data, x' indicates the normalized value, x_{mean} indicates the average value, x_{std} indicates the variance. Next, calculate the mean squared error (MSE) loss for all sky, sun, and camera parameters:

$$\begin{aligned} loss_{wsky} &= \frac{1}{3}\left\|\widehat{w}_{sky} - \widetilde{w}_{sky}\right\|_2^2 & loss_t &= \left\|\hat{t} - \tilde{t}\right\|_2^2 \\ loss_{wsun} &= \frac{1}{3}\left\|\widehat{w}_{sun} - \widetilde{w}_{sun}\right\|_2^2 & loss_\beta &= \left\|\hat{\beta} - \tilde{\beta}\right\|_2^2 \\ loss_k &= \left\|\hat{k} - \tilde{k}\right\|_2^2 & loss_{fov} &= \left\|\widehat{fov} - \widetilde{fov}\right\|_2^2 \end{aligned} \tag{4}$$

where, $w_{sky} \in R^3$, $w_{sun} \in R^3$, fov represent the vertical field of view of the camera, the superscript $\widehat{(\ )}$ represents the true value, and $\widetilde{(\ )}$ represents the predicted value.

The resulting optimization goals are as follows:

$$loss = \frac{1}{6}\left(loss_{wsky} + loss_t + loss_{wsun} + loss_\beta + loss_k + loss_{fov}\right) \tag{5}$$

4 Experiments and analysis

First, in order to find the one that is most suitable for outdoor illumination estimation from many attention mechanisms and can verify the effectiveness of the convolution block attention module introduced in this paper, this paper conducted a comparison experiment between different attention mechanisms; In this paper, the effects of the number and position of convolution block attention module (CBAM) are verified experimentally. Finally, the effectiveness of the method proposed in this paper is verified by comparison with existing methods.

4.1 Comparison between different attention mechanisms

This paper mainly compares three attention mechanisms proposed by Wang et al. (2018), Hu et al. (2018), and Woo et al. (2018). The three can be expressed respectively by self-attention, SE, and CBAM. In addition, in order to prove the effectiveness of attention mechanism, this paper also carried out experiments named base on the case without adding any attention module. The final results are shown in Table 2 and 3 . Among them, Table 2 is the prediction result of the sun orientation. The evaluation index used is the angle difference between the predicted sun orientation and the true sun orientation. The unit of the data is percentage; It is the prediction result of the remaining six illumination parameters. The evaluation index used is the root mean square error (RMSE) between the predicted value and the true value.

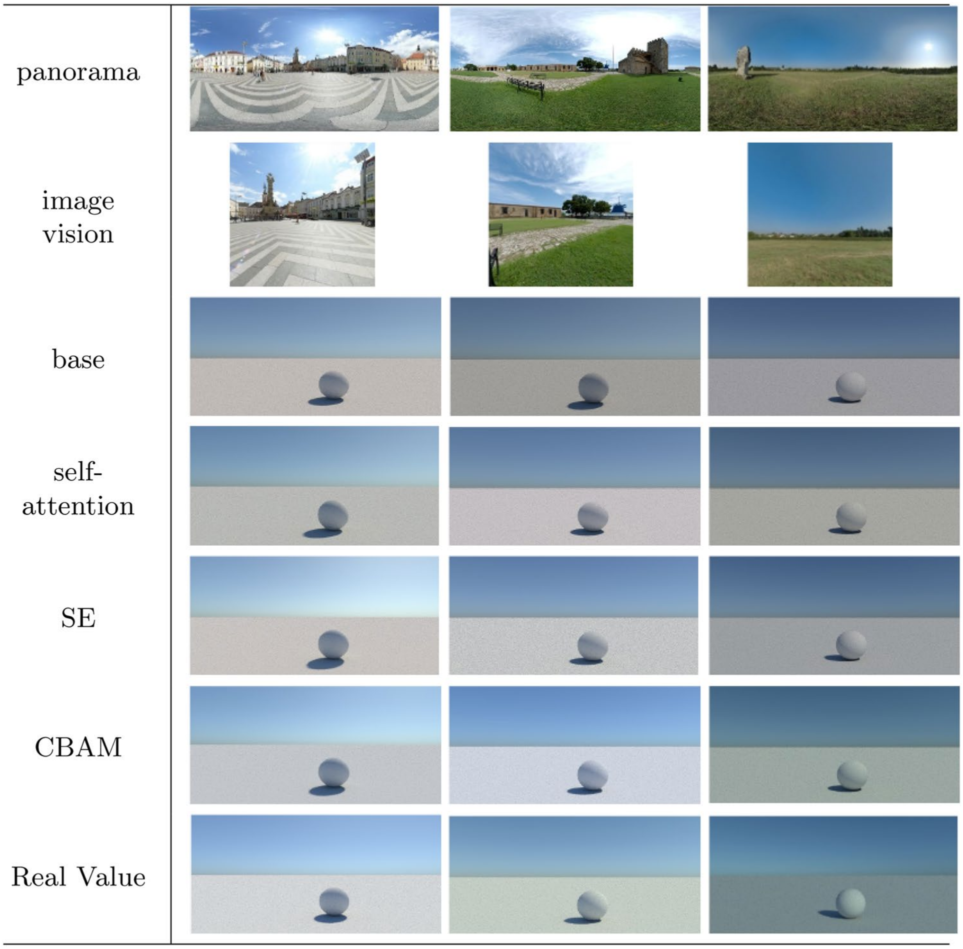


Fig. 5 Rendering results of different attention mechanisms

Table 2 Error of sun orientation estimation with different attention mechanisms

Attention	Degree							
	0	10	20	30	40	50	60	70
Base	0.0	13.32	34.41	51.49	64.55	74.34	82.23	88.01
Self-attention	0.0	16.38	39.07	55.92	67.71	76.43	83.23	88.32
SE	0.0	15.99	37.31	53.63	66.10	74.86	81.99	87.98
CBAM	0.0	20.06	44.79	61.55	72.09	79.60	85.34	89.53

Table 3 Error of remaining 6 parameters with different attention mechanisms (RMSE)

Attention	Parameters					
	W_{sun}	W_{sky}	k	β	t	fov
Base	0.4005	0.0954	0.1662	9.7823	1.2029	0.0529
Self-attention	0.3753	0.0939	0.1552	9.1151	1.1439	0.0643
SE	0.3776	0.0914	0.1523	8.9298	1.1161	0.0475
CBAM	0.3597	0.0854	0.1503	8.4783	1.0602	0.0449

In addition to the quantitative comparison of results, this paper also carried out a qualitative comparison by visual methods, as shown in Fig 5. Among them, the first row is a 360-degree panorama, which is randomly selected from the test set; The second row is a image of a certain field of view taken from the panorama; The third, fourth, fifth, and sixth rows are respectively the renderings of base, self-attention, SE, and CBAM rendering; The 7th line is rendering of real value.

From Table 2 and Table 3, it can be known that no matter what kind of attention mechanism is used, the prediction of the sun orientation and the remaining six parameters are favorable, and the prediction accuracy can be improved, but the effect is different. Among them, the effect of CBAM is the most significant.

From Fig 5, it can be known that the rendering effect generated by the three attention mechanisms is closer to the real value than the base. And the result rendered by CBAM is more realistic. Therefore, the convolution block attention module (CBAM) was finally selected in this paper.

4.2 Comparison between different numbers of attention modules

In the previous section, although this paper demonstrates the effect of the convolution block attention module. The influence of the module embedded in the network will have on the question is worth exploring. Via experiments, this paper verified this problem. The network structures used for comparison are Network-1, Network-2, Network-3, and Network-4. Among them, there are only one attention module on the two branches of network-1; There are two attention modules on the two branches of network-2, which are located behind the convolutional layers of the 4th and 8th layers respectively; There are 3 attention modules on the two branches in network-3, which are located behind the convolution layers on the 2nd, 5th, and 8th layers; Threr are 4 attention modules on the 2 branches in the network-4, which located behind the second, fourth, sixth, and eighth convolution layers. The prediction of the sun orientation for the four network structures are shown in Table 4, and the prediction of the remaining six parameters are shown in Table 5.

For Tables 4 and 5, the comparison of the four networks show that whether it is the prediction of the sun orientation or the remaining six parameters, the more attention modules are introduced the better the result is. But when the number of modules reaches 2, the results change slow or even little change with the attention modules increase. Considering the balance between the amount of computation and performance, this paper sets 2 attention modules.

Then, this paper designs an experiment for the location of the module.The network structures used for comparison are Network-1, Network-2 and Network-3. Among them, the attention modules in Network-1 are in the front position in the network; The attention modules in Network-2 are evenly distributed in the network; The attention modules in Network-3 are in the back position in the network. The prediction results of the three network structures for the sun orientation are shown in Table 6, and the prediction results of the remaining six parameters are shown in Table 7.

It can be seen from Tables 6 and 7 that the prediction of the sun orientation and the remaining six parameters have a good performence when the attention modules are evenly distributed in the network. so this paper evenly distributes the two attention modules in the two network.

4.3 Comparison with existing methods

This paper mainly compares with the method of Hold-Geoffroy et al. (2017). This method is also for the illumination

Table 4 Error of cumulative sun orientation estimation with four network structures

Net-Structure	Degree							
	0	10	20	30	40	50	60	70
Network-1	0.0	16.75	39.78	57.45	68.73	77.14	83.69	88.56
Network-2	0.0	20.09	44.99	61.55	72.09	79.60	85.34	89.73
Network-3	0.0	20.10	45.05	61.39	72.08	79.49	85.15	89.72
Network-4	0.0	20.11	45.07	61.58	71.87	79.50	84.83	89.94

Table 5 Error of the remaining six parameters with four networks (RMSE)

Net-Structure	Parameters					
	W_{sun}	W_{sky}	k	β	t	fov
Network-1	0.3689	0.0887	0.1505	8.7224	1.0664	0.0564
Network-2	0.3567	0.0854	0.1501	8.4783	1.0602	0.0449
Network-3	0.3580	0.0857	0.1497	8.5360	1.0502	0.0516
Network-4	0.3552	0.0856	0.1484	8.5059	1.0582	0.0448

Table 6 Error of sun orientation estimation with three network structures

Net-Structure	Degree							
3	0	10	20	30	40	50	60	70
Network-1	0.0	18.98	42.54	58.51	69.88	78.06	84.40	89.16
Network-2	0.0	20.09	44.99	61.55	72.09	79.60	85.34	89.73
Network-3	0.0	18.35	42.46	59.01	70.67	78.69	84.62	89.31

Table 7 Error of the remaining six parameters with three networks (RMSE)

Net-Structure	Parameters					
	W_{sun}	W_{sky}	k	β	t	fov
Network-1	0.3624	0.0871	0.1531	8.6997	1.0653	0.0511
Network-2	0.3567	0.0854	0.1501	8.4783	1.0602	0.0449
Network-3	0.3580	0.0864	0.1500	8.5615	1.0567	0.0495

Table 8 Error of two methods for illumination parameters (RMSE)

Approach	Parameters	
	t	fov
Our method	1.0602	0.0449
Hold-Geoffroy et al. (2017)	1.1715	0.1473

estimation of outdoor images, and it uses the Hošek-Wilkie (HošekHošek and Wilkie 2013) sky model. This model mainly uses the parameters of the sun orientation, atmospheric turbidity, exposure, camera zenith angle, and vertical field of view to represent the scene illumination. Because this paper is different from the sky model used by Hold-Geoffroy et al. (2017), the illumination parameters obtained by the sky model are also different. The comparison between the two can only be based on parameters that are present in both sky models, such as the sun orientation, atmospheric turbidity, the vertical field of view of the camera, and so on. The comparison curve of the sun orientation are shown in Fig 6, and the comparison results for the atmospheric turbidity and the vertical field of view of the camera are shown in Table 8.

As can be seen from Fig 6, the method proposed in this paper has higher accuracy of sun orientation prediction than

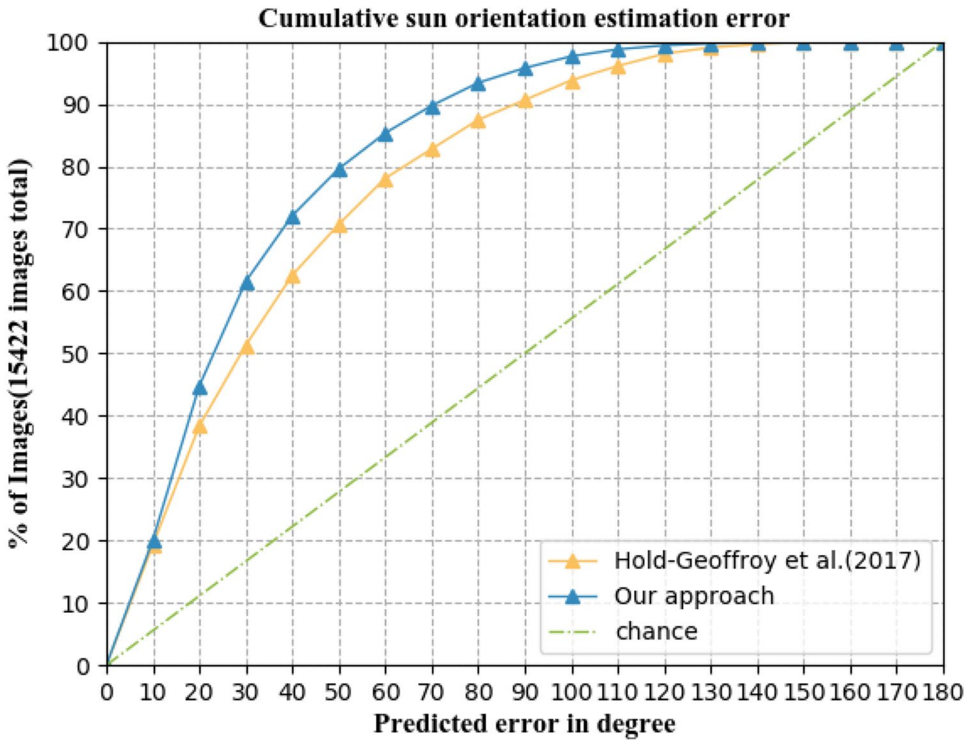


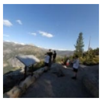
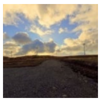
Fig. 6 Curve of sun orientation estimation error between the two methods

the method of Hold-Geoffroy et al. (2017). For example, for the method of Hold-Geoffroy et al. (2017) and the method proposed in this paper, the number of images with a sun orientation estimation error of less than 20 degrees accounted for 38%, 45% of the total test images; The number of images less than 30 degrees accounted for 51%, 62%. It can be seen from Table 8 that no matter whether it is atmospheric turbidity or the vertical field of view of the camera, the proposed method has lower prediction error than the method of Hold-Geoffroy et al. (2017).

In addition, this paper also randomly selected some images from the test set and compared the errors between the three illumination parameters predicted by the two methods and the true values, as shown in Table 9. These images cover a variety of more common situations such as no sky in the image, a small sky area, a large sky area and so on. At the same time, they also cover a variety of weather conditions, such as sunny, cloudy, foggy and so on. It can be seen from Table 9, the method proposed in this paper is more accurate for most cases than the method of Hold-Geoffroy et al. (2017) in predicting the sun orientation, atmospheric turbidity, and camera vertical field of view.

In order to simplify the representation in Table 9, this paper uses *sp* represents the sun orientation, *t* represents atmospheric turbidity, *fov* represents camera vertical field of view. The sun orientation error uses the angle error between the predicted value and the true value, and the other parameters use the root mean square error(RMSE). The unit of the

Table 9 Comparison of the error of the three illumination parameters with two methods

					
Real value	*sp*	(0.440, −4.121)	(0.783, 2.160)	(1.327, −1.577)	(0.474, −0.059)
	t	21.6190	20.2154	17.0702	23.1224
	fov	1.3908	1.5285	1.6428	1.3304
Hold-Geoffroy et al. (2017)	*sp*	(−0.082, −0.034)	(−0.379, −0.903)	(0.465, −0.246)	(−0.482, 0.376)
	t	22.0848	20.5387	19.3405	22.4158
	fov	1.5561	1.6685	1.7024	1.6021
Error	*sp*	22.78°	23.27°	71.44°	33.42°
	t	0.4658	0.3233	2.2703	0.7066
	fov	0.1653	0.1400	0.0596	0.2717
This paper method	*sp*	(−0.287, −0.409)	(−0.547, −0.972)	(0.583, −1.194)	(0.032, −0.328)
	t	20.4535	20.0046	18.7234	23.7563
	fov	1.3758	1.5428	1.7007	1.3505
Error	*sp*	14.27°	13.53°	45.85°	25.36°
	t	1.1655	0.2108	1.6532	0.6339
	fov	0.0150	0.0143	0.0579	0.0201
Real value	*sp*	(−0.076, 3.117)	(0.228, 4.203)	(0.320, 1.390)	(0.980, 2.078)
	t	23.0485	19.2110	17.7632	19.9791
	fov	1.3454	1.6858	1.3454	1.7046
Hold-Geoffroy et al. (2017)	*sp*	(−0.112, 0.321)	(−0.100, 0.341)	(−0.027, 0.135)	(−0.737, 0.308)
	t	20.5621	19.0737	18.5621	20.4590
	fov	1.5976	1.7162	1.5976	1.6668
Error	*sp*	10.70°	9.51°	18.89°	28.47°
	t	2.4864	0.1373	0.7989	0.4799
	fov	0.2522	0.0304	0.2522	0.0378
This paper method	*sp*	(0.007, 0.043)	(−0.066, 1.119)	(−0.143, 0.549)	(−0.680, −1.153)
	t	21.2422	19.4676	17.1376	19.5563
	fov	1.3275	1.6411	1.3661	1.6716
Error	*sp*	3.98°	9.28°	24.59°	17.55°
	t	1.8063	0.2566	0.6256	0.4228
	fov	0.0179	0.0447	0.0207	0.0330

sun orientation is radians, and the unit of the angle error is the angle.

Furthermore, this paper also proposes a method for sun orientation estimation problem, as shown in the Fig 7. The loss function used in this method is Kullback-Leibler divergence loss function proposed by Zhang et al. (2019)which also different from the previous one. This paper mainly compares with the method of Zhang et al. (2019). The experimental results are shown in Fig 8. It can be seen from the figure that the method has a higher accuracy of sun orientation prediction than the method of Zhang et al. (2019).

5 Conclusion

For the problem of outdoor illumination estimation, this paper uses seven illumination parameters in the Lalonde-Matthews illumination model to represent the illumination. For the prediction of these seven parameters, a new two-branch network structure is proposed in this paper. One branch is used to predict the sun's orientation, and the other branch is used to predict the remaining parameters. The two branches are independent and do not share weight parameters. At the same time, this paper also analyzes and explains why the network is designed. This paper also introduces the Convolution Block Attention Module (CBAM) into this network structure. The introduction of this module enhances the network's feature expression ability and improves the prediction accuracy of the illumination parameters. Comparison

Fig. 7 Sun orientation estimation network

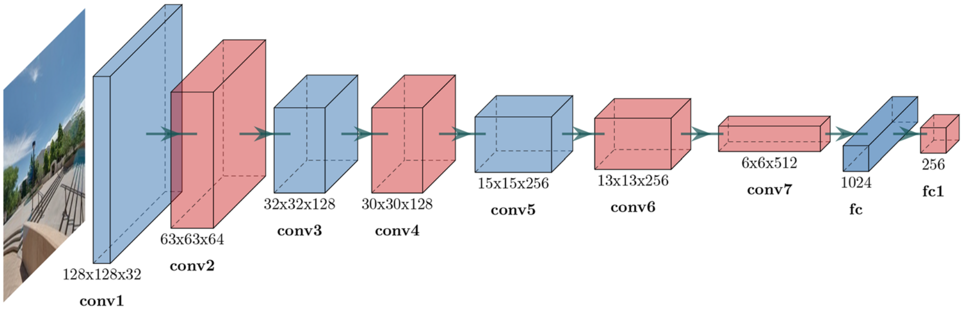
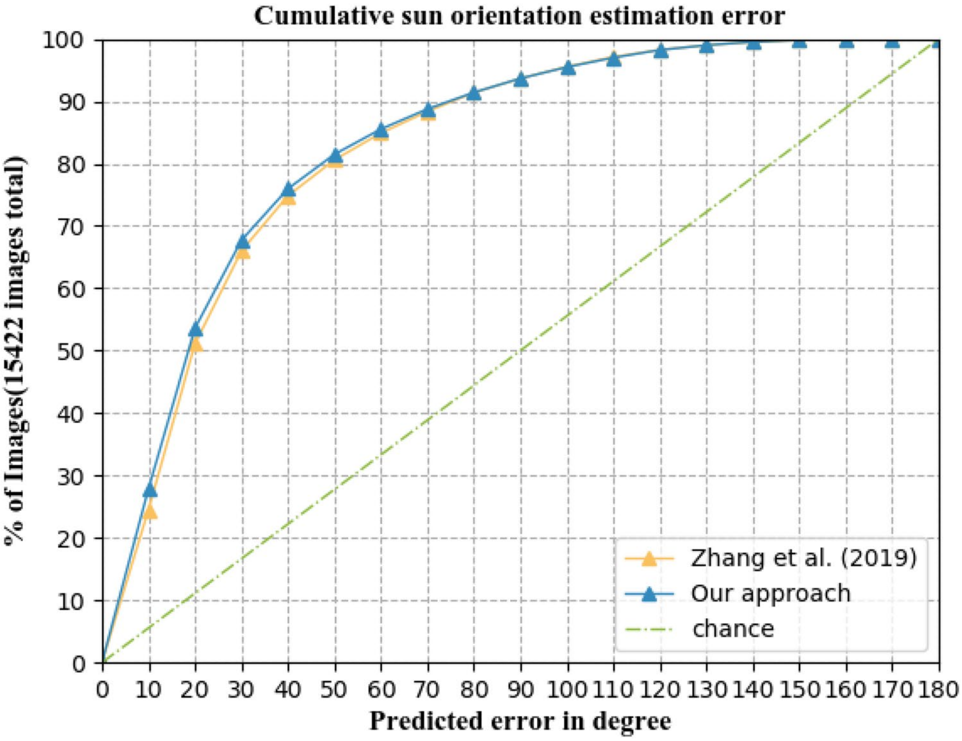


Fig. 8 Curve of sun orientation estimation error between the two methods



with existing methods validates the effectiveness of the proposed method.

Acknowledgements This work is partially supported by the National Natural Science Foundation of China (Grant Numbers 61701008, 61772047), the Beijing Natural Science Foundation (Grant Number 19L2040), the Open Project Program of State Key Laboratory of Cryptology (Grant Number MMKFKT201804), the Open Project Program of State Key Laboratory of Virtual Reality Technology and Systems, Beihang University (Grant Number VRLAB2019C03) and the Fundamental Research Funds for the Central Universities (Grant Number 328201907).

References

Barron JT, Malik J (2014) Shape, illumination, and reflectance from shading. IEEE Trans Pattern Anal Mach Intell 37(8):1670–1687

Cheng D, Shi J, Chen Y, Deng X, Zhang X (2018) Learning scene illumination by pairwise photos from rear and front mobile cameras. Comput Graph Forum, Wiley Online Library 37:213–221

Debevec P (2008) Rendering synthetic objects into real scenes: bridging traditional and image-based graphics with global illumination and high dynamic range photography. In: ACM SIGGRAPH 2008 classes, pp 1–10

Erra U, Capece N (2019) Engineering an advanced geo-location augmented reality framework for smart mobile devices. J Ambient Intell Hum Comput 10(1):255–265

Gardner MA, Sunkavalli K, Yumer E, Shen X, Gambaretto E, Gagné C, Lalonde JF (2017) Learning to predict indoor illumination from a single image. arXiv preprint arXiv:170400090

Green R (2003) Spherical harmonic lighting: the gritty details. In: Archives of the game developers conference, vol 56, p 4

Hold-Geoffroy Y, Sunkavalli K, Hadap S, Gambaretto E, Lalonde JF (2017) Deep outdoor illumination estimation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 7312–7321

Hosek L, Wilkie A (2012) An analytic model for full spectral sky-dome radiance. ACM Trans Graph (TOG) 31(4):1–9

HošekHošek L, Wilkie A (2013) Adding a solar-radiance function to the hošek-wilkie skylight model. IEEE Comput Graph Appl 33(3):44–52

Hu J, Shen L, Sun G (2018) Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 7132–7141

Jaderberg M, Simonyan K, Zisserman A, et al. (2015) Spatial transformer networks. In: Advances in neural information processing systems, pp 2017–2025

Jin X, Sun X, Zhang X, Sun H, Xu R, Zhou X, Li X, Liu R (2019) Sun orientation estimation from a single image using short-cuts in dcnn. Opt Laser Technol 110:191–195

Kim T, Hong KS (2005) A practical single image based approach for estimating illumination distribution from shadows. In: Tenth IEEE international conference on computer vision (ICCV'05) Volume 1, IEEE, vol 1, pp 266–271

Lalonde JF, Matthews I (2014) Lighting estimation in outdoor image collections. In: 2014 2nd International conference on 3D vision, IEEE, vol 1, pp 131–138

Lalonde JF, Efros AA, Narasimhan SG (2012) Estimating the natural illumination conditions from a single outdoor image. Int J Comput Vis 98(2):123–145

Lele A (2013) Virtual reality and its military utility. J Ambient Intell Hum Comput 4(1):17–26

Li Y, Lu H, Serikawa S (2015) Underwater image devignetting and colour correction. International conference on image and graphics, pp 510–521

Li Y, Lu H, Li K, Kim H, Serikawa S (2018) Non-uniform de-scattering and de-blurring of underwater images. Mob Netw Appl 23(2):352–362

Li P, Chen B, Ouyang W, Wang D, Yang X, Lu H (2019) Gradnet: Gradient-guided network for visual object tracking. In: Proceedings of the IEEE international conference on computer vision, pp 6162–6171

Lombardi S, Nishino K (2015) Reflectance and illumination recovery in the wild. IEEE Trans Pattern Anal Mach Intell 38(1):129–141

Lu H, Li Y, Nakashima S, Serikawa S (2016) Turbidity underwater image restoration using spectral properties and light compensation. IEICE Trans Inf Syst 99(1):219–227

Ma WC, Wang S, Brubaker MA, Fidler S, Urtasun R (2017) Find your way by observing the sun and other semantic cues. In: 2017 IEEE international conference on robotics and automation (ICRA), IEEE, pp 6292–6299

Perez R, Seals R, Michalsky J (1993) All-weather model for sky luminance distribution-preliminary configuration and validation. Solar Energy 50(3):235–245

Preetham AJ, Shirley P, Smits B (1999) A practical analytic model for daylight. In: Proceedings of the 26th annual conference on computer graphics and interactive techniques, pp 91–100

Stumpfel J, Jones A, Wenger A, Tchou C, Hawkins T, Debevec P (2006) Direct hdr capture of the sun and sky. In: ACM SIGGRAPH 2006 Courses, pp 5–es

Wang M, Callaghan V, Bernhardt J, White K, Peña-Rios A (2018a) Augmented reality in education and training: pedagogical approaches and illustrative case studies. J Ambient Intell Hum Comput 9(5):1391–1402

Wang X, Girshick R, Gupta A, He K (2018) Non-local neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 7794–7803

Weber H, Prévost D, Lalonde JF (2018) Learning to estimate indoor lighting from 3d objects. In: 2018 International conference on 3D vision (3DV), IEEE, pp 199–207

Woo S, Park J, Lee JY, So Kweon I (2018) Cbam: Convolutional block attention module. In: Proceedings of the European conference on computer vision (ECCV), pp 3–19

Xiao J, Ehinger KA, Oliva A, Torralba A (2012) Recognizing scene viewpoint using panoramic place representation. In: 2012 IEEE conference on computer vision and pattern recognition, IEEE, pp 2695–2702

Yan B, Zhao H, Wang D, Lu H, Yang X (2019) 'Skimming-perusal' tracking: a framework for real-time and robust long-term tracking. In: Proceedings of the IEEE international conference on computer vision, pp 2385–2393

Zhang H, Goodfellow I, Metaxas D, Odena A (2018) Self-attention generative adversarial networks. arXiv preprint arXiv:180508318

Zhang J, Sunkavalli K, Hold-Geoffroy Y, Hadap S, Eisenman J, Lalonde JF (2019) All-weather deep outdoor lighting estimation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 10158–10166

Zhao Q (2009) A survey on virtual reality. Sci China Ser F Inf Sci 52(3):348–400

Zhou Z, Zhou Y, Xiao J (2015) Survey on augmented virtual environment and augmented reality. SCIENTIA SINICA Inf 45(2):157–180