AIM 2020: Scene Relighting and Illumination Estimation Challenge

Majed El Helou[1(✉)], Ruofan Zhou[1], Sabine Süsstrunk[1], Radu Timofte[2], Mahmoud Afifi[3], Michael S. Brown[3], Kele Xu[4], Hengxing Cai[4], Yuzhong Liu[4], Li-Wen Wang[5], Zhi-Song Liu[5,6], Chu-Tak Li[5], Sourya Dipta Das[7], Nisarg A. Shah[8], Akashdeep Jassal[9], Tongtong Zhao[10], Shanshan Zhao[11], Sabari Nathan[12], M. Parisa Beham[13], R. Suganya[14], Qing Wang[15], Zhongyun Hu[15], Xin Huang[15], Yaning Li[15], Maitreya Suin[16], Kuldeep Purohit[16], A. N. Rajagopalan[16], Densen Puthussery[17], P. S. Hrishikesh[17], Melvin Kuriakose[17], C. V. Jiji[17], Yu Zhu[18], Liping Dong[18], Zhuolong Jiang[18], Chenghua Li[18], Cong Leng[18], and Jian Cheng[18]

[1] EPFL, Lausanne, Switzerland
{majed.elhelou,ruofan.zhou,sabine.susstrunk}@epfl.ch
[2] ETHZ, Zrich, Switzerland
radu.timofte@vision.ee.ethz.ch
[3] EECS, York University, Toronto, ON, Canada
mafifi@eecs.yorku.ca
[4] National University of Defense Technology, Changsha, China
kelele.xu@gmail.com
[5] Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, Hong Kong, China
liwen.wang@connect.polyu.hk
[6] CS laboratory at the Ecole Polytechnique, Palaiseau, France
[7] Jadavpur University, Kolkata, India
dipta.juetce@gmail.com
[8] Indian Institute of Technology, Jodhpur, India
[9] Punjab Engineering College (PEC), Chandigarh, India
[10] Dalian Maritime University, Dalian, China
daitoutiere@gmail.com
[11] China Everbright Bank, Beijing, China
[12] Couger Inc, Tokyo, Japan
sabarinathantce@gmail.com
[13] Sethu Institute of Technology,, Virudhunagar, India
[14] Thiagarajar College of Engineering, Virudhunagar, India

M. El Helou, R. Zhou, S. Süsstrunk, and R. Timofte are the challenge organizers, and the other authors are challenge participants.
Appendix A lists all the teams and affiliations.
https://github.com/majedelhelou/VIDIT.

Electronic supplementary material The online version of this chapter (https://doi.org/10.1007/978-3-030-67070-2_30) contains supplementary material, which is available to authorized users.

A. Bartoli and A. Fusiello (Eds.): ECCV 2020 Workshops, LNCS 12537, pp. 499–518, 2020.
https://doi.org/10.1007/978-3-030-67070-2_30

[15] Computer Vision and Computational Photography Group, School of Computer Science, Northwestern Polytechnical University, Xi'an, China
zy_h@mail.nwpu.edu.cn
[16] Indian Institute of Technology Madras, Chennai, India
maitreyasuin21@gmail.com
[17] College of Engineering, Trivandrum, India
puthusserydensen@gmail.com
[18] Nanjing Artificial Intelligence Chip Research, Institute of Automation, Chinese Academy of Sciences (AiRiA); MAICRO, Beijing, China
zhuyu.cv@gmail.com

Abstract. We review the AIM 2020 challenge on virtual image relighting and illumination estimation. This paper presents the novel VIDIT dataset used in the challenge and the different proposed solutions and final evaluation results over the 3 challenge tracks. The first track considered one-to-one relighting; the objective was to relight an input photo of a scene with a different color temperature and illuminant orientation (i.e., light source position). The goal of the second track was to estimate illumination settings, namely the color temperature and orientation, from a given image. Lastly, the third track dealt with any-to-any relighting, thus a generalization of the first track. The target color temperature and orientation, rather than being pre-determined, are instead given by a guide image. Participants were allowed to make use of their track 1 and 2 solutions for track 3. The tracks had 94, 52, and 56 registered participants, respectively, leading to 20 confirmed submissions in the final competition stage.

Keywords: Image relighting · Illumination estimation · Style transfer

1 Introduction

Deep image relighting has multiple applications both in research and in practice, and is recently witnessing increased interest. A single-image relighting method would allow aesthetic enhancement applications, such as photo montage of images taken under different illuminations, and illumination retouching without human expert work. Very importantly, in computer vision research image relighting can be leveraged for data augmentation, enabling the trained methods to be robust to changes in light source position or color temperature. It could also serve for domain adaptation, by normalizing input images to a unique set of illumination settings that the down-stream computer vision method was trained on. The relighting task contains multiple sub-tasks, namely, illumination estimation and manipulation, shadow removal or practically inpainting for hardly lit areas, and geometric understanding for shadow recasting. The combination of these tasks makes relighting very challenging.

Recently, datasets limited to interior scenes [33], underexposed images enhanced by professionals [48], and rendered images with randomized light directions [54] have been proposed, but none serve the benchmarking needs for image relighting, namely, having all $M \times N$ combinations of M scenes and N illumination settings. Further datasets are used in the literature on style transfer or intrinsic image decomposition. For instance, IIW [6] and SAW [27] contain human-labeled reflectance and shading annotations, and BigTime [29] contains time-lapse data of scenes illuminated under varying light conditions. Multiple methods are recently being developed for relighting [12,34,42], and the prior literature on intrinsic images, which disentangle surface reflectance from lighting, is rich [5,6,18,39,44,51], notably for applications such as relighting [7] and normalization [32].

The aim of this challenge, and of the novel dataset **V**irtual **I**mage **D**ataset for **I**llumination **T**ransfer (VIDIT), is to gauge the current state-of-the-art for image relighting. The virtual dataset provides a well-controlled setup to provide full-reference evaluation, which is ideal for benchmarking purposes, and is an important step towards real-image relighting. Such virtual datasets have proven useful in multiple applications to augment even the training datasets containing real images, for instance the vKitti data [9]. There could be differences relative to real images such as the distribution of textures that can vary from man-made to natural scenes [8,45], the specifics of the capturing device like chromatic aberrations [15,31,58], or the presence of multiple light sources. VIDIT itself is described in the following section. The goal of the challenge is thus to provide a benchmark on this dataset for future research on image relighting.

This challenge is one of the AIM 2020 associated challenges on: scene relighting and illumination estimation [17], image extreme inpainting [36], learned image signal processing pipeline [24], rendering realistic bokeh [25], real image super-resolution [50], efficient super-resolution [56], video temporal super-resolution [41] and video extreme super-resolution [19].

2 Scene Relighting and Illumination Estimation Challenge

2.1 Dataset

The challenge, whose 3 tracks are described in the following section, is based on a novel dataset: VIDIT [16]. VIDIT contains 300 virtual scenes used for training, where every scene is captured 40 times in total: from 8 equally-spaced azimuthal angles, each lit with 5 different illuminants. Every image is 1024×1024, but the images are downsampled by a factor of 2, with bicubic interpolation over 4×4 windows, to ease computations for track 3. The dataset is publicly available (https://github.com/majedelhelou/VIDIT).

2.2 Tracks and Competition

Track 1: One-to-one Relighting

Description: the relighting task is pre-determined and fixed for all validation and test samples. In other words, the objective is to manipulate an input image from one pre-defined set of illumination settings (namely, North, 6500K) to another pre-defined set (East, 4500K). The images are in 1024×1024 resolution, both input and output, and nothing other than the input image is provided.

Evaluation Protocol: We evaluate the results using the PSNR and SSIM [49] metrics, and the self-reported run-times and implementation details are also provided. For the final ranking, we define a Mean Perceptual Score (MPS) as the average of the normalized SSIM and LPIPS [57] scores, themselves averaged across the entire test set of each submission

$$0.5 \cdot (S + (1 - L)), \tag{1}$$

where S is the SSIM score, and L is the LPIPS score. We note that normalizing S and $(1 - L)$, by dividing them respectively by their maximum values across all the track's submissions, before averaging the two does not affect the final ranking. We thus do not do this normalization, which also makes it simpler for external comparisons.

Track 2: Illumination Settings Estimation

Description: the goal of this track is to estimate, from a single input image, the illumination settings that were used in rendering it. Given the input image, the output should estimate the color temperature of the illuminant as well as the orientation, i.e. the position of the light source. The input images are also 1024×1024 and no other input is given than the 2D image.

Evaluation Protocol: The evaluation of track 2 is based on the accuracy of predictions following this formula for the loss

$$\sqrt{\sum_{i=0}^{N-1} \left(\frac{|\hat{\phi}_i - \phi_i| mod 180}{180} \right)^2 + (\hat{T}_i - T_i)^2} \tag{2}$$

where $\hat{\phi}_i$ is the predicted angle (0–360) for test sample i and ϕ_i is the ground-truth value for that sample. $\hat{T}_i$ is the temperature prediction for test sample i and T_i is the ground-truth value for that sample. T_i takes values equal to [0, 0.25, 0.5, 0.75, 1], which correspond to the color temperature values [2500K, 3500K, 4500K, 5500K, 6500K].

Track 3: Any-to-any Relighting

Description: this track is a generalization of the first track. The objective is to relight an input image (both color temperature and light source position manipulation) from any arbitrary illumination settings to any arbitrary illumination

settings. The latter settings are dictated by a second input guide image, as in style transfer applications. The participants were allowed to make use of their solutions to the first two tracks to develop a solution for this track. The images are in 512×512 resolution to ease computations, as this track is very challenging.

Evaluation Protocol: We carry out a similar evaluation as for track 1. As the inputs are pairs of possible test images, they cover a larger span of candidate options. For that reason, we double the number of data samples in the validation and test sets for this track.

Challenge Phases for all Tracks. (1) Development: registered teams were given access to the training input and target data, as well as the input validation set data. An online validation server with a leader board provided automated feedback for the submitted image results on the validation set, which was made up of 45 images for tracks 1 and 2, and 90 image pairs for track 3; (2) Testing: registered teams were given access to the input test sets, which are of the same size as the validation ones, and could submit their test results to a private test server. For a submission to be accepted, open-source code and a fact sheet detailing the implemented method needed to be submitted along with the test results. Test results were kept hidden from participating teams, to avoid any chances of test over-fitting, and were only revealed at the end of the challenge.

3 Challenge Results

The results of all three tracks are collected in Tables 1, 2, and 3, respectively. The top solutions are described in the following sections, and the remainder is in the supplementary material.

Table 1. AIM 2020 Image Relighting Challenge Track 1 (One-to-one relighting) results. The MPS, used to determine the final ranking, is computed following Eq. (1). *CET_CVLab and CET_SP are merged into one, due to large similarity between the proposed solutions. We also note that normalizing SSIM and (1-LPIPS) scores by the maximum in the track, for computing the MPS, does not affect the ranking.

Team	Author	**MPS** ↑	SSIM ↑	LPIPS ↓	PSNR ↑	Run-time	Platform	GPU
CET_SP*	hrishikeshps	0.6452 (-)	0.6310 (2)	0.3405 **(1)**	17.0717 (2)	0.03s	Tensorflow	P100
CET_CVLab	Densen	0.6451 **(1)**	0.6362 **(1)**	0.3460 (3)	16.8927 (6)	0.03s	Tensorflow	P100
lyl	tongtong	0.6436 (2)	0.6301 (3)	0.3430 (2)	16.6801 (8)	13s	PyTorch	V100
YorkU	mafifi	0.6216 (3)	0.6091 (4)	0.3659 (5)	16.8196 (7)	6s	PyTorch	1080TI
IPCV_IITM	ms_icpv	0.5897 (4)	0.5298 (7)	0.3505 (4)	17.0594 (3)	0.04s	PyTorch	Titan X
DeepRelight	leven	0.5892 (5)	0.5928 (6)	0.4144 (7)	17.4252 **(1)**	0.5s	PyTorch	2080TI
Withdrawn	tomanut	0.5603 (6)	0.5236 (8)	0.4029 (6)	16.5136 (9)	0.01s	PyTorch	2080TI
Hertz	souryadipta	0.5339 (7)	0.5666 (6)	0.4989 (8)	16.9234 (4)	0.006s	PyTorch	1080TI
Image Lab	sabarinathan	0.3746 (8)	0.3769 (9)	0.6278 (9)	16.8949 (5)	0.12s	Tensorflow	1080TI
input image	–	0.6438	0.6288	0.3412	16.2796			

Table 2. AIM 2020 Image Relighting Challenge Track 2 (Illumination settings estimation) results. The loss is computed based on the angle and color temperature predictions, following Eq. (2), and is used to determine the final ranking.

Team	Author	**Loss** ↓	AngLoss ↓	TempLoss ↓	Run-time	Platform	GPU
AiRiA_CG	Airia_CG	0.0875 **(1)**	0.0722 (3)	0.0153 **(1)**	0.03s	PyTorch	Titan Xp
YorkU	mafifi	0.0887 (2)	0.0639 (2)	0.0248 (2)	0.95s	MATLAB	1080TI
Image Lab	sabarinathan	0.0984 (3)	0.0513 **(1)**	0.0471 (5)	0.02s	Tensorflow	1080TI
debut_kele	debut_kele	0.1431 (4)	0.1125 (4)	0.0306 (3)			
RGETH	Georgechogovadze	0.1708 (5)	0.1347 (5)	0.0361 (4)	0.026s	PyTorch	
random guess	–	0.5987	0.3729	0.2257			

Table 3. AIM 2020 Image Relighting Challenge Track 3 (Any-to-any relighting) results. The MPS, used to determine the final ranking, is computed following Eq. (1). We also note that normalizing SSIM and (1-LPIPS) scores by the maximum in the track, for computing the MPS, does not affect the ranking.

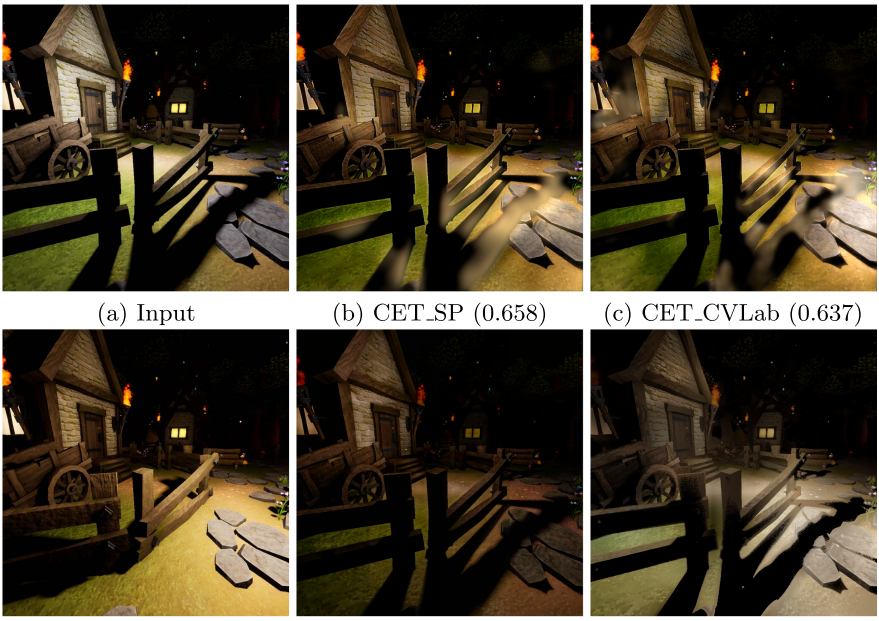
Team	Author	**MPS** ↑	SSIM ↑	LPIPS ↓	PSNR ↑	Run-time	Platform	GPU
NPU-CVPG	walden	0.6484 **(1)**	0.6353 **(1)**	0.3386 (3)	18.5436 (2)	0.15s	PyTorch	1080TI
YorkU	mafifi	0.6428 (2)	0.6195 (2)	0.3338 (2)	18.2384 (4)	6s	PyTorch	1080TI
IPCV_IITM	ms_icpv	0.6424 (3)	0.6042 (3)	0.3194 **(1)**	19.3559 **(1)**	0.3s	PyTorch	Titan X
lyl	tongtong	0.6213 (4)	0.5881 (4)	0.3455 (4)	17.6314 (5)	13s	PyTorch	V100
AiRiA_CG	Airia_CG	0.5258 (5)	0.4451 (5)	0.3936 (5)	18.3493 (3)		PyTorch	Titan Xp
RGETH	Georgechogovadze	0.3465 (6)	0.4123 (6)	0.7192 (6)	10.4483 (6)	0.0289s	PyTorch	
Input image	–	0.6750	0.6603	0.3103	17.9391			

Visual results of some top submissions along with input and ground-truth images for track 1 are shown in Fig. 1. We notice that most of the outputs generate the relit image with the correct color temperature, however, the shadows are harder to estimate. For instance, lyl and YorkU suffer from shadow removal. Both CET_SP and CET_CVLab tend to remove the unnecessary shadows, although not perfectly, which underlines the difficulty of the shadow-relighting sub-task. We show visual results of some submissions to track 3 in Fig. 2. Among the top 3 submissions, only NPU-CVPG is able to successfully relight the bottom-right part and produce the closest color temperature to the ground-truth.

4 Track 1 Methods

4.1 CET_CVLab: Wavelet Decomposed RelightNet (WDRN)

The architecture of the proposed Wavelet Decomposed RelightNet (WDRN) [37] is shown in Fig. 3. The network structure used is similar to that of an encoder-decoder U-Net. The downsampling operation used in the contraction path is a discrete wavelet transform (DWT) based decomposition instead of a downsampling convolution or pooling. Similarly, in the expansion path, the inverse discrete wavelet transform (IDWT) is used instead of an upsampling convolution.



(a) Input (b) CET_SP (0.658) (c) CET_CVLab (0.637)

(d) Ground-truth (e) lyl (0.610) (f) YorkU (0.580)

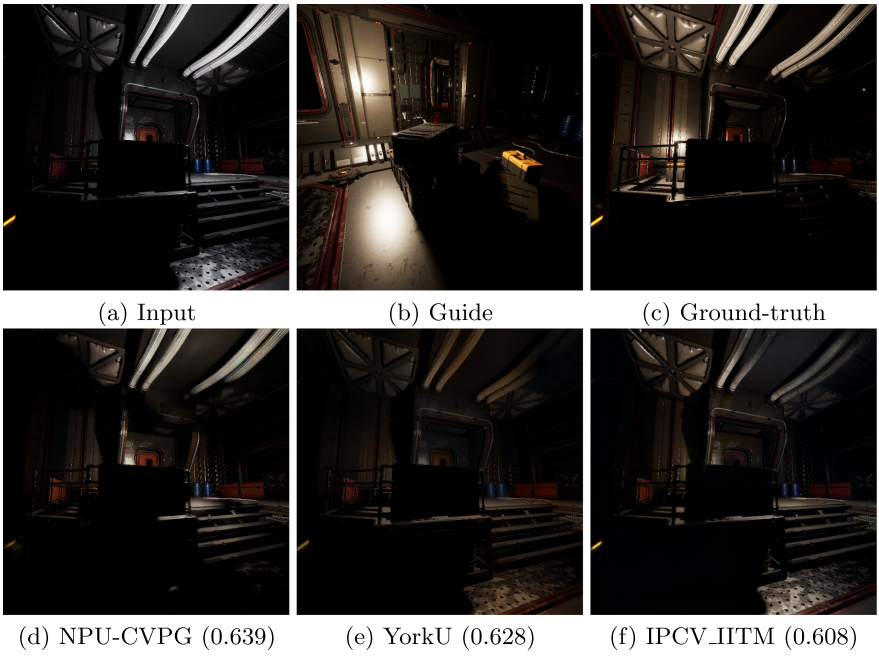
Fig. 1. Sample visual results from top submissions in track 1, with MPS scores. We observe that relighting previous shadows is the most difficult sub-task.

In the wavelet based decomposition, the information from all channels is combined in the downsampling process such that there is minimal information loss when compared to that of a convolutional subsampling. For the given task, it can be deduced that the network must learn to re-calibrate the illumination gradient within the image. To this end, the network should be able to establish the relation between distant pixels. The proposed WDRN can achieve a high receptive field and hence establish this relation with the multi-scale wavelet decomposition. Also, this methodology is computationally efficient and is inspired by the multi-level wavelet-CNN (MWCNN) proposed by Liu *et al.* [30]. The training loss used in this work is a weighted sum of the SSIM loss, MAE loss and a *gray loss* (the gray loss term is used in the CET_SP submission, and omitted in that of CET_CVLab). Gray loss is the $\ell 1$ distance between the grayscale version of the restored image and that of the ground-truth image.

4.2 lyl: Coarse-to-Fine Relighting Net (CFRN)

The proposed Coarse-to-Fine Relighting Net (CFRN) is illustrated in Fig. 4. The solution consists of two networks: (1) progressive coarse network and (2) a network merging the output of the coarse network, with channel attention, to correct the input in each level. Such a progressive process helps to achieve the principle for image relighting: high-level information is a good guide to obtain a

better relit image. In the proposed method, there are three indispensable parts; (1) tying the loss at each level (2) using the FineNet structure and (3) providing a lower-level extracted feature input to ensure the availability of low-level information. To make full use of the training data, the team augments data in three ways; (1) scaling: randomly downscaling between [0.5,1.0], (2) rotation:



(a) Input (b) Guide (c) Ground-truth

(d) NPU-CVPG (0.639) (e) YorkU (0.628) (f) IPCV_IITM (0.608)

Fig. 2. Sample visual results from top submissions in track 3, with MPS scores.

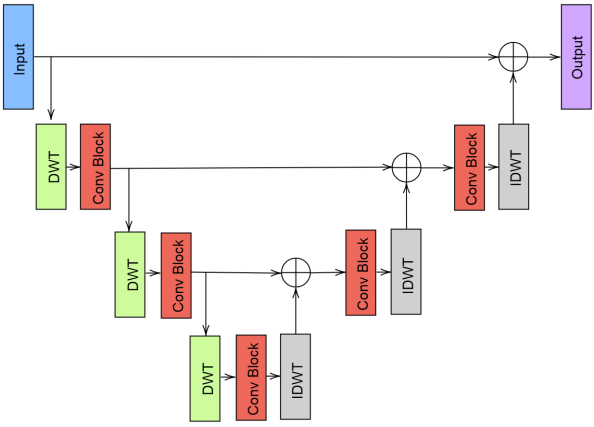


Fig. 3. Architecture of the Wavelet Decomposed RelightNet (WDRN).

randomly rotating the image by 90, 180, and 270 degrees, and (3) flipping: randomly flipping images horizontally or vertically with equal probability.

4.3 YorkU: Norm-Relighting-U-Net (NRUNet)

The method adopts a U-Net architecture [38] as the main backbone of the proposed framework. The solution consists of two networks: (1) the normalization network, which is responsible for producing uniformly-lit white-balanced images, and (2) the relighting network, which performs the one-to-one image relighting. An instance normalization [46] is applied after each stage in the encoder of the normalization network, while batch normalization is used for the encoder of the relighting network. The relighting network is fed the input image and the latent representations of the uniformly-lit image produced by the normalization network. The team uses the white-balance augmenter in [2] to augment the training data. To produce the ground-truth of the normalization network, the team uses the training data provided for tracks 2 and 3, which include a set

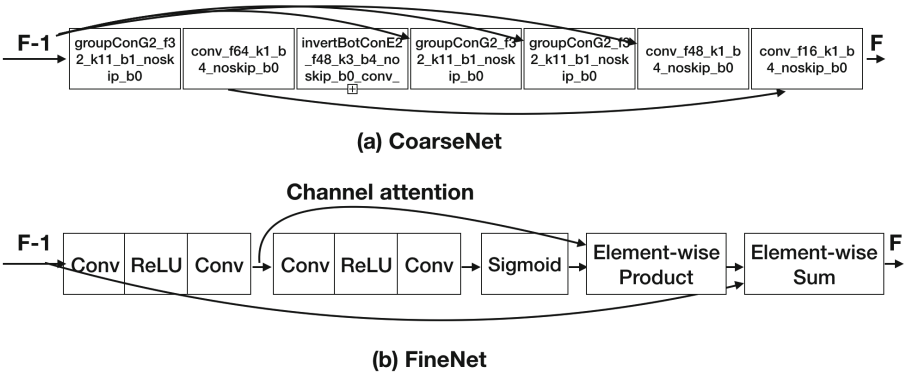


Fig. 4. Architecture diagram of the Coarse-to-Fine Relighting Net (CFRN).

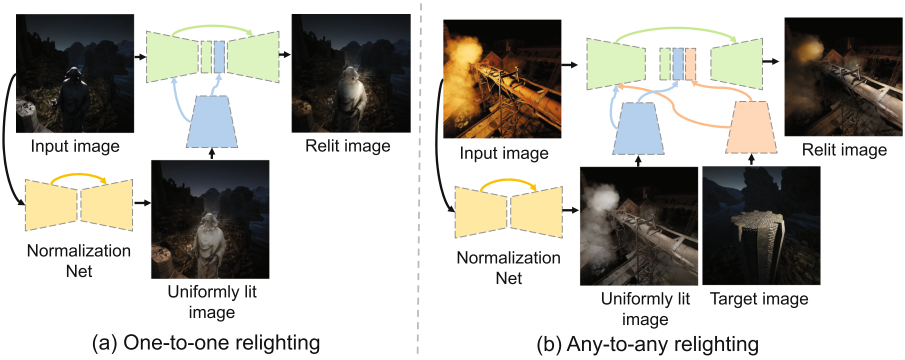


Fig. 5. Overview of YorkU team's NRUNet framework.

of images taken from each scene under different lighting directions. The team exploits their solution for the illumination settings estimation task (see Sect. 5.2) to predict the target scene settings for the one-to-one mapping. Hence, the team increases the number of training images by including the training images provided for tracks 2 and 3. The team pre-trains the normalization network then fixes its weights and the entire framework is jointly trained. The training uses the Adam optimizer [26] with $\ell 1$ loss. At inference, the team processes a resized version of the input image, then a guided up-sampling [10] is applied to obtain the full-resolution image. The team ensembles the final results by utilizing their one-to-any framework (more details on the one-to-any framework in Sect. 6.2). To relight the image using the one-to-any framework, the team randomly selects six images with the predicted illumination settings of the current track to use them as targets. This procedure generates six relit images that are used along with the result image produced by the one-to-one framework to generate the final result. Figure 5-(a) shows an overview of the proposed one-to-one mapping framework. The source code for the three tracks is available at https://github.com/mahmoudnafifi/image_relighting.

4.4 IPCV_IITM: Deep Residual Network for Image Relighting (DRNIR)

Figure 6 shows the structure of the proposed residual network with skip connections, based on the hourglass network [59]. The network has an encoder-decoder structure with skip connections [23]. Residual blocks are used in the skip connections, and Batch-Norm and ReLU non-linearity in each of the blocks. The encoder features are concatenated with the decoder features of same level. The network takes the input image and directly produces the target image. The team converts the input RGB images to LAB for better processing. To reduce the memory consumption without harming the performance, the team uses a pixel-shuffle block [40] to downsample the image. The network is first trained using the $\ell 1$ loss, then fine-tuned with the MSE loss. Note that experiments with adversarial loss did not lead to stable training. The learning rate of the Adam

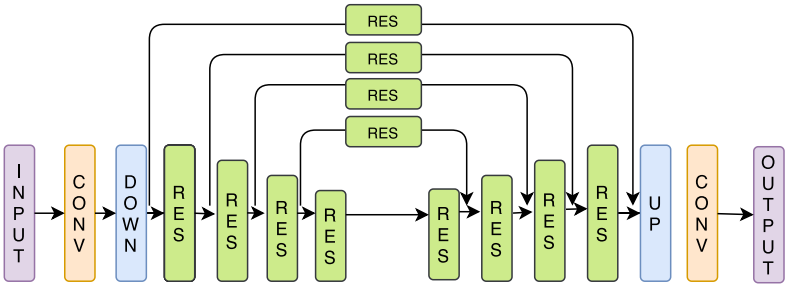


Fig. 6. Diagram illustration of the DRNIR network architecture.

optimizer is 0.0001 with a decay cycle of 200 epochs, and a 512×512 patch size for training. Data augmentation is used to make the network more robust.

4.5 Other Submitted Solutions

The DeepRelight team addresses the one-to-one relighting task by recovering the structure information of the scene, target illumination information, and renders the output with a GAN strategy [47]. Another solution makes use of two pairs of encoder-decoder networks, such that the encoding and decoding are illumination specific, and the learning is also supervised with discriminators. Transforming an image becomes equivalent to encoding it with the first encoder and decoding it with the second. Hertz tackle the problem using a multi-scale hierarchical network, the image is encoded at multiple resolutions and feature information is transferred from lower to higher levels to obtain the final transformation. Lastly, Image Lab [35] build on the multilevel hyper vision net [14], adding convolution block attention [52] in their skip connections. Further details of each of these submitted solutions can be found in the supplementary material.

5 Track 2 Methods

5.1 AiRiA_CG: Dual Path Ensemble Network (DPENet)

The proposed DPENet has two sub-networks, one for angle prediction and one for temperature classification [13]. The full DPENet is shown in Fig. 7. ResNeXt-101_32×4d [53] is adopted for the angle prediction sub-network. The temperature classification sub-network is based on ResNet-50 [20]. The two sub-networks are pre-trained on ImageNet [11]. The solution adopts random flipping and random rotation for data augmentation.

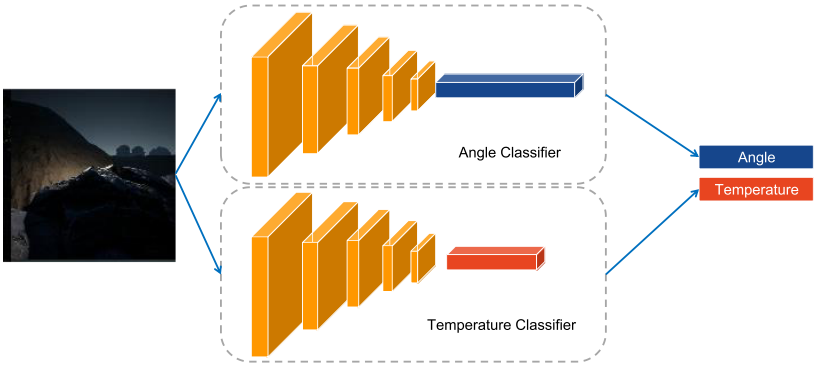


Fig. 7. The structure of Dual Path Ensemble Network (DPENet).

5.2 YorkU: Illuminant-ResNet (I-ResNet)

The team treats the task as two independent classification tasks; (1) illuminant temperature classification and (2) illuminant angle classification. The team adopts the ResNet-18 model [20] trained on ImageNet [11]. The last fully-connected layer is replaced with a new layer with n neurons, where n is the number of output classes for each task. The Adam optimizer [26] is used with cross entropy loss. For angle classification, the team applies the white-balance augmenter proposed in [2] to augment the training data. For temperature classification, the team follows previous work [1,3,4] that uses image histogram features instead of the 2D input image. Specifically, the team feeds the network with 2D RGB-uv projected histogram features [1,3], instead of the original training images. This histogram-based training, rather than image-based, improves the model's generalization. Figure 8 shows an overview of the team's solution, including the white-balance augmentation process.

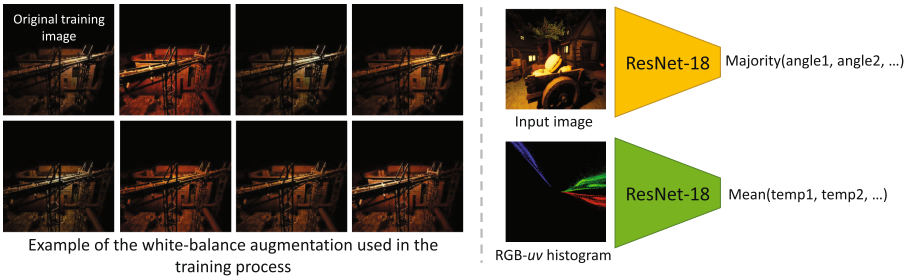


Fig. 8. Overview of the YorkU solution, with the white-balance augmentation [2].

5.3 Image Lab: Virtual Image Illumination Estimation (LightNet)

As shown in Fig. 9, the team adopts a Densenet [22] architecture for the task. The team trains ten different pre-trained networks and also creates a custom network with selective blocks [28]. From these networks, the Densnet121 network achieves the best performance. DenseNet121 consists of fifty-eight dense blocks, followed by three transition blocks and three fully-connected layers. The global average pooling and fully connected layers are removed from the pre-trained network, and replaced with a new global average pooling and fully connected layers with a degree and temperature output layer. From the training dataset, the team creates a random splitting, with 67% of samples taken for training and the rest for validation. The training images are normalized to [0,1]. The Adam optimizer with a learning rate decaying from 0.001 to 0.00001 over 500 epochs is used for training the model with the categorical loss. Attention layers [52] were tested in the development phase but did not yield any improvement.

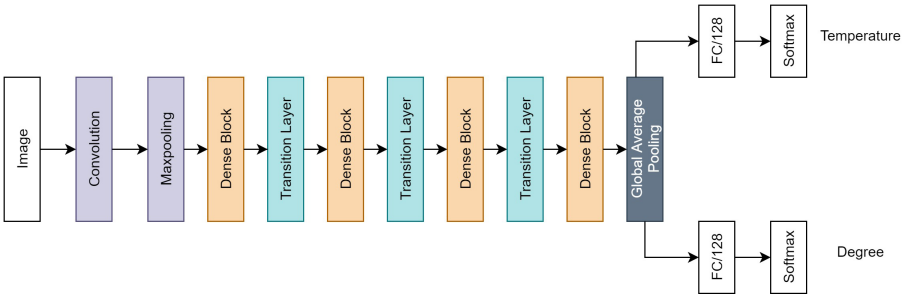


Fig. 9. Overview of the LightNet model's architecture.

5.4 Other Submitted Solution

The debut_kele team proposes to use a single EfficientNet [43] backbone, pre-trained on ImageNet. Further details of this submitted solution can be found in the supplementary material.

6 Track 3 Methods

6.1 NPU-CVPG: Self-Attention AutoEncoder (SA-AE)

As shown in Fig. 10, the team presents the novel Self-Attention AutoEncoder (SA-AE) [21] model for generating a relit image from a source image to match the illumination settings of a guide image. In order to reduce the learning difficulty, the team adopts an implicit scene representation [59] learned by the encoder to render the relit images using the decoder. Based on the learned scene representation, an illumination estimation network is designed as a classier to predict the illumination settings of the guide image. A lighting-to-feature network is also designed to recover the corresponding implicit scene representation from the illumination settings, similar to the inverse of the illumination estimation process. In addition, a self-attention [55] mechanism is introduced in the decoder to focus on the rendering of the regions requiring relighting in the source images.

6.2 YorkU: Norm-Relighting-U-Net (NRUNet)

As for the one-to-one mapping proposed (Sect. 4.3), the U-Net architecture [38] is used as the main backbone of the any-to-any relighting framework, and two networks are used for normalization and relighting, as shown in Fig. 5-(b). The relighting network is fed the input image, the latent representation of the guide image and the uniformly lit image produced by the normalization network. The team uses the white-balance augmentation [2] on the training data for the normalization network. The team trains two frameworks; one framework on 256×256 random patches and one on 256×256 resized images. The final result is generated by taking the mean of the two relit images and applying a guided upsampling [10].

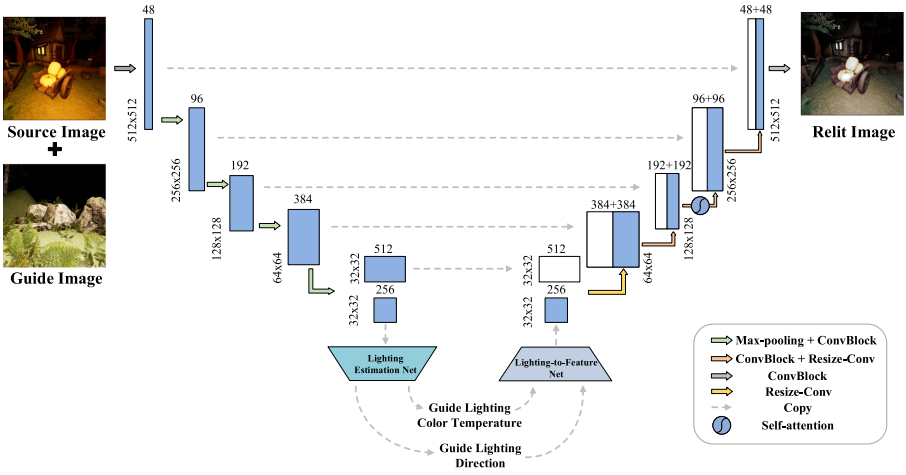


Fig. 10. Overview of the proposed SA-AE network.

6.3 IPCV_IITM: Deep Residual Network for Image Relighting (DRNIR)

Figure 11 shows the structure of the proposed residual network with skip connections, based on the hourglass network [59]. The network has an encoder-decoder structure similar to [23]. The team also uses residual blocks in the skip connections. The encoder features are concatenated with the decoder features of the same level. Along with the input image, the network is given a guide image that is used in two places. First, both the input and the guide image are concatenated. Second, the team adds a separate loss to match the illumination properties between the guide image and the predicted image. A separate network predicts the illumination settings of an image, and is trained with the provided ground-truth labels. The team passes both the guide image and the predicted image through the network and minimizes the distance between intermediate feature representations. The feature representation of the guide image is further concatenated with the encoder output and fed to the decoder. The team converts the input RGB images to LAB for better processing. To reduce memory consumption, pixel-shuffle blocks [40] are used as in track 1.

6.4 lyl: Coarse-to-Fine Relighting Net (CFRN)

The proposed Coarse-to-Fine Relighting Net (CFRN) is shown in Fig. 4, as in track 1. Training is divided in two stages: incomplete training and full training. During an incomplete training, the fine network is trained with a batch size of 16 for 200 epochs. The Adam optimizer ($\beta_1 = 0.9$,$\beta_2 = 0.999$) is used to minimize the $\ell 1$ loss between the generated relit images and the ground-truth. The learning rate is initialized to 10^4 and kept unchanged. After the incomplete training with

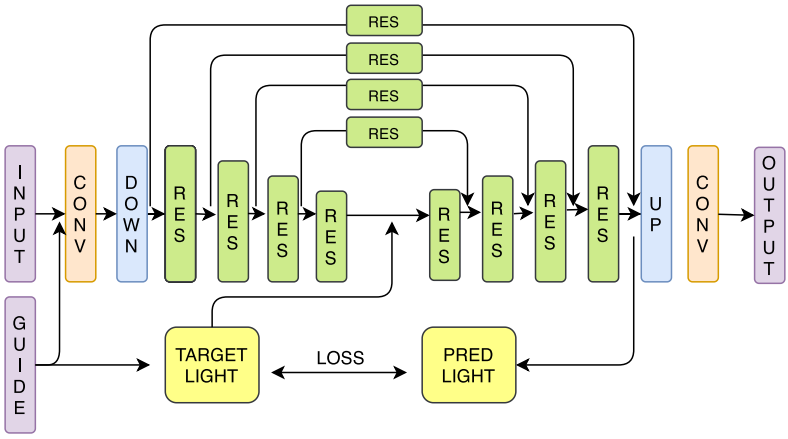


Fig. 11. Network architecture of the DRNIR method.

the fine network, the whole CFRN is fully trained. In each full training batch, the team randomly samples 64 patches for 20k epochs.

6.5 Other Submitted Solution

The AiRiA_CG team proposes a creative solution consisting of a dual encoder and single decoder [13]. The input image is encoded, and so is the target image. However, the encoder of the target image is mirrored to match the decoder of the input image latent representation, and the feature layers of the former are thus transferred, layer by layer, to the decoder of the latter. This allows the illumination information to be transferred from the guide image to the input image during the decoding process. Further details of this submitted solution can be found in the supplementary material.

Acknowledgements. We thank all AIM 2020 sponsors: Huawei, MediaTek, NVIDIA, Qualcomm, Google and CVL, ETH Zurich (https://data.vision.ee.ethz.ch/cvl/aim20/). We also note that all tracks were supported by the CodaLab infrastructure (https://competitions.codalab.org).

A Teams and Affiliations

AIM challenge organizers

Members: Majed El Helou, Ruofan Zhou, Sabine Süsstrunk ({*majed.elhelou, ruofan.zhou,sabine.susstrunk*}*@epfl.ch*, EPFL, Switzerland), and Radu Timofte (*radu.timofte@vision.ee.ethz.ch*, ETH Zürich, Switzerland).

– AiRiA_CG –

Members: Yu Zhu(*zhuyu.cv@gmail.com*), Liping Dong, Zhuolong Jiang,

Chenghua Li, Cong Leng, Jian Cheng
Affiliation: Nanjing Artificial Intelligence Chip Research, Institute of Automation, Chinese Academy of Sciences (AiRiA); MAICRO.
– **CET_CVLab** –
Members: Densen Puthussery (*puthusserydensen@gmail.com*), Hrishikesh P S, Melvin Kuriakose, Jiji C V
Affiliation: College of Engineering, Trivandrum, India.
– **debut_kele** –
Members: Kele Xu (*kelele.xu@gmail.com*), Hengxing Cai, Yuzhong Liu
Affiliation: National University of Defense Technology, China.
– **DeepRelight** –
Members: Li-Wen Wang[1] (*liwen.wang@connect.polyu.hk*), Zhi-Song Liu[1,2], Chu-Tak Li[1], Wan-Chi Siu[1], Daniel P. K. Lun[1]
Affiliation: [1]Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, [2]CS laboratory at the Ecole Polytechnique (Palaiseau).
– **Hertz** –
Members: Sourya Dipta Das[1] (*dipta.juetce@gmail.com*), Nisarg A. Shah[2], Akashdeep Jassal[3]
Affiliation: [1]Jadavpur University, Kolkata, India, [2]Indian Institute of Technology Jodhpur, India, [3]Punjab Engineering College (PEC), Chandigarh, India.
– **Image Lab** –
Members: Sabari Nathan[1] (*sabarinathantce@gmail.com*), M.Parisa Beham[2], R.Suganya[3]
Affiliation: [1]Couger Inc, Tokyo, Japan, [2]Sethu Institute of Technology, India, [3]Thiagarajar College of Engineering, India.
– **IPCV_IITM** –
Members: Maitreya Suin (*maitreyasuin21@gmail.com*), Kuldeep Purohit, A. N. Rajagopalan
Affiliation: Indian Institute of Technology Madras, India.
– **lyl** –
Members: Tongtong Zhao[1] (*daitoutiere@gmail.com*), Shanshan Zhao[2]
Affiliation: [1]Dalian Maritime University,[2] China Everbright Bank.
– **NPU-CVPG** –
Members: Zhongyun Hu (*zy_h@mail.nwpu.edu.cn*), Xin Huang, Yaning Li, Qing Wang
Affiliation: Computer Vision and Computational Photography Group, School of Computer Science, Northwestern Polytechnical University.
– **RGETH** –
Members: George Chogovadze (*chogeorg@student.ethz.ch*), Rémi Pautrat
Affiliation: ETH Zurich, Switzerland.
– **YorkU** –
Members: Mahmoud Afifi (*mafifi@eecs.yorku.ca*), Michael S. Brown
Affiliation: EECS, York University, Toronto, ON, Canada.

References

1. Afifi, M., Brown, M.S.: Sensor-independent illumination estimation for DNN models. In: British Machine Vision Conference (BMVC), p. 11 (2019)
2. Afifi, M., Brown, M.S.: What else can fool deep learning? addressing color constancy errors on deep neural network performance. In: IEEE International Conference on Computer Vision (ICCV), pp. 243–252 (2019)
3. Afifi, M., Price, B., Cohen, S., Brown, M.S.: When color constancy goes wrong: correcting improperly white-balanced images. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1535–1544 (2019)
4. Barron, J.T.: Convolutional color constancy. In: IEEE International Conference on Computer Vision (ICCV), pp. 379–387 (2015)
5. Barron, J.T., Malik, J.: Color constancy, intrinsic images, and shape estimation. In: European Conference on Computer Vision (ECCV), pp. 57–70 (2012)
6. Bell, S., Bala, K., Snavely, N.: Intrinsic images in the wild. ACM Trans. Graph. (TOG) **33**(4), 159 (2014)
7. Bousseau, A., Paris, S., Durand, F.: User-assisted intrinsic images. In: ACM SIGGRAPH Asia, pp. 1–10 (2009)
8. Burton, G.J., Moorhead, I.R.: Color and spatial structure in natural scenes. Appl. Opt. **26**(1), 157–170 (1987)
9. Cabon, Y., Murray, N., Humenberger, M.: Virtual kitti 2. arXiv preprint arXiv:2001.10773 (2020)
10. Chen, J., Adams, A., Wadhwa, N., Hasinoff, S.W.: Bilateral guided upsampling. ACM Trans. Graph. (TOG) **35**(6), 1–8 (2016)
11. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: ImageNet: a large-scale hierarchical image database. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 248–255 (2009)
12. Dherse, A.P., Everaert, M.N., Gwizdała, J.J.: Scene relighting with illumination estimation in the latent space on an encoder-decoder scheme. arXiv preprint arXiv:2006.02333 (2020)
13. Dong, L., Jiang, Z., Li, C.: An ensemble neural network for scene relighting with light classification. In: Proceedings of the European Conference on Computer Vision Workshops (ECCVW) (2020)
14. D. Sabarinathan, Beham, M., Roomi, S.: Moire image restoration using multi level hyper vision net. Image and Video Processing arXiv:2004.08541 (2020)
15. El Helou, M., Dümbgen, F., Süsstrunk, S.: AAM: an assessment metric of axial chromatic aberration. In: IEEE International Conference on Image Processing (ICIP), pp. 2486–2490 (2018)
16. El Helou, M., Zhou, R., Barthas, J., Süsstrunk, S.: VIDIT: virtual image dataset for illumination transfer. arXiv preprint arXiv:2005.05460 (2020)
17. El Helou, M., et al.: AIM 2020: scene relighting and illumination estimation challenge. In: European Conference on Computer Vision Workshops (2020)
18. Finlayson, G.D., Drew, M.S., Lu, C.: Intrinsic images by entropy minimization. In: European Conference on Computer Vision (ECCV), pp. 582–595 (2004)
19. Fuoli, D., et al.: AIM 2020 challenge on video extreme super-resolution: methods and results. In: European Conference on Computer Vision Workshops (2020)
20. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR).,pp. 770–778 (2016)

21. Hu, Z., Huang, X., Li, Y., Wang, Q.: SA-AE for any-to-any relighting. In: Proceedings of the European Conference on Computer Vision Workshops (ECCVW) (2020)
22. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2261–2269 (2017)
23. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4700–4708 (2017)
24. Ignatov, A., et al.: AIM 2020 challenge on learned image signal processing pipeline. In: European Conference on Computer Vision Workshops (2020)
25. Ignatov, A., et al.: AIM 2020 challenge on rendering realistic bokeh. In: European Conference on Computer Vision Workshops (2020)
26. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
27. Kovacs, B., Bell, S., Snavely, N., Bala, K.: Shading annotations in the wild. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6998–7007 (2017)
28. Li, X., Wang, W., Hu, X., Yang, J.: Selective kernel networks. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 510–519 (2019)
29. Li, Z., Snavely, N.: Learning intrinsic image decomposition from watching the world. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 9039–9048 (2018)
30. Liu, P., Zhang, H., Zhang, K., Lin, L., Zuo, W.: Multi-level wavelet-CNN for image restoration. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, pp. 773–782 (2018)
31. Llanos, B., Yang, Y.H.: Simultaneous demosaicing and chromatic aberration correction through spectral reconstruction. In: IEEE Conference on Computer and Robot Vision (CRV), pp. 17–24 (2020)
32. Matsushita, Y., Nishino, K., Ikeuchi, K., Sakauchi, M.: Illumination normalization with time-dependent intrinsic images for video surveillance. Trans. Pattern Anal. Mach. Intell. **26**(10), 1336–1347 (2004)
33. Murmann, L., Gharbi, M., Aittala, M., Durand, F.: A dataset of multi-illumination images in the wild. In: IEEE International Conference on Computer Vision (ICCV), pp. 4080–4089 (2019)
34. Nagano, K., et al.: Deep face normalization. ACM Trans. Graph. (TOG) **38**(6), 183 (2019)
35. Nathan, D.S., Beham, M.P.: LightNet: deep learning based illumination estimation from virtual images. In: European Conference on Computer Vision Workshops (2020)
36. Ntavelis, E., et al.: AIM 2020 challenge on image extreme inpainting. In: European Conference on Computer Vision Workshops (2020)
37. Puthussery, D., P S, H., Kuriakose, M., C V., J.: WDRN: a wavelet decomposed relightnet for image relighting. In: European Conference on Computer Vision Workshops (2020)
38. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28

39. Shen, J., Yang, X., Jia, Y., Li, X.: Intrinsic images using optimization. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3481–3487 (2011)
40. Shi, W., et al.: Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1874–1883 (2016)
41. Son, S., et al.: AIM 2020 challenge on video temporal super-resolution. In: European Conference on Computer Vision Workshops (2020)
42. Sun, T., et al.: Single image portrait relighting. ACM Trans. Graph. (TOG) **38**(4), 79 (2019)
43. Tan, M., Le, Q.V.: Efficientnet: rethinking model scaling for convolutional neural networks. arXiv preprint arXiv:1905.11946 (2019)
44. Tappen, M.F., Freeman, W.T., Adelson, E.H.: Recovering intrinsic images from a single image. In: Advances in Neural Information Processing Systems, pp. 1367–1374 (2003)
45. Torralba, A., Oliva, A.: Statistics of natural image categories. Netw. Comput. Neural Syst. **14**(3), 391–412 (2003)
46. Ulyanov, D., Vedaldi, A., Lempitsky, V.: Instance normalization: the missing ingredient for fast stylization. arXiv preprint arXiv:1607.08022 (2016)
47. Wang, L.W., Siu, W.C., Liu, Z.S., Li, C.T., Lun, D.P.: Deep relighting networks for image light source manipulation. In: Proceedings of the European Conference on Computer Vision Workshops (ECCVW) (2020)
48. Wang, R., Zhang, Q., Fu, C.W., Shen, X., Zheng, W.S., Jia, J.: Underexposed photo enhancement using deep illumination estimation. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6849–6857 (2019)
49. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004)
50. Wei, P., et al.: AIM 2020 challenge on real image super-resolution. In: European Conference on Computer Vision Workshops (2020)
51. Weiss, Y.: Deriving intrinsic images from image sequences. In: IEEE International Conference on Computer Vision (ICCV), vol. 2, pp. 68–75 (2001)
52. Woo, S., Park, J., Lee, J.Y., Kweon, I.S.: CBAM convolutional block attention module. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 1–17 (2018)
53. Xie, S., Girshick, R., Dollár, P., Tu, Z., He, K.: Aggregated residual transformations for deep neural networks. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1492–1500 (2017)
54. Xu, Z., Sunkavalli, K., Hadap, S., Ramamoorthi, R.: Deep image-based relighting from optimal sparse samples. ACM Trans. Graph. (TOG) **37**(4), 126 (2018)
55. Zhang, H., Goodfellow, I., Metaxas, D., Odena, A.: Self-attention generative adversarial networks. In: International Conference on Machine Learning (ICML), pp. 7354–7363 (2019)
56. Zhang, K., et al.: AIM 2020 challenge on efficient super-resolution: methods and results. In: European Conference on Computer Vision Workshops (2020)

57. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 586–595 (2018)
58. Zhao, J., Hou, Y., Liu, Z., Xie, H., Liu, S.: Modified color CCD moiré method and its application in optical distortion correction. Precis. Eng. **65**, 279–286 (2020)
59. Zhou, H., Hadap, S., Sunkavalli, K., Jacobs, D.W.: Deep single-image portrait relighting. In: IEEE International Conference on Computer Vision (ICCV), pp. 7194–7202 (2019)