The Visual Computer (2022) 38:1665–1675
https://doi.org/10.1007/s00371-021-02095-5

ORIGINAL ARTICLE

Learning to detect soft shadow from limited data

Wen Wu[1] · Shuping Zhang[1] · Mi Tian[1] · Daoqiang Tan[2] · Xiantao Wu[3] · Yi Wan[4]

Accepted: 11 February 2021 / Published online: 9 March 2021

Abstract

Soft shadow is more challenging to detect than hard shadow due to its ambiguous boundary. Existing shadow detection methods pay more attention to hard shadow scene since collecting and annotating hard shadow images is more effortless. Motivated by that soft shadow has similar characteristics with hard shadow, and many traditional hard shadow datasets are publicly available, we propose a novel soft shadow detection method (namely Soft-DA) based on adversarial learning and domain adaptation scheme. Specifically, we create a limited soft shadow dataset, containing 1K soft shadow images with various scenes and shapes. Note that we just only need to annotate 0.4K shadow masks for semi-supervised learning. Besides, to tackle obvious domain discrepancy and potential intention difference between different datasets and similar tasks, we first align data distributions between domains by feature adversarial adaptation. And then, we introduce a novel detector separation strategy to tackle the intention difference issue. In this way, Soft-DA can effectively detect soft shadow with only a small number of soft shadow annotations. Extensive experiments demonstrate that our method can achieve superior performance to state of the arts.

Keywords Deep learning · Image processing · Semi-supervised learning · Domain adaptation · Soft shadow detection

1 Introduction

Shadow is a common illumination phenomenon in our daily life. On the one hand, shadow is the primary source of error and uncertainty. For example, the cast shadow may be wrongly marked as an object in the target tracking [30]. On the other hand, the shadow can also give us instrumental cues about the light source [16], object shapes [27], image depth [31], and geometry [12]. Hence, shadow detection is a fundamental component of scene understanding. In recent years, many researchers have proposed lots of shadow detection methods that mainly aim to shadow images with dark shadows and clear boundaries (hard shadow). According to [6,26], shadow image consists of the sunlit, umbra, and penumbra regions. The weak light source intensity or the longer distances between the ground and the projection object will result in shadows with wide penumbrae and unclear shadow boundaries (soft shadow), leading the detection of soft shadow is very challenging.

In the past couple of years, deep learning has achieved great success in most computer vision and image processing tasks [2,11,41,44]. One of the critical contributions behind it is a large amount of labeled training data. Most works in the field of shadow detection [1,4,7,10,13,25,32,35,36,39,42]

Wen Wu and Shuping Zhang contributed equally to this work.

✉ Yi Wan
19945853@qq.com

Wen Wu
1119764335@qq.com

Shuping Zhang
2714105538@qq.com

Mi Tian
450269537@qq.com

Daoqiang Tan
787342703@qq.com

Xiantao Wu
1185777178@qq.com

1 School of Information Engineering, Xinjiang Institute of Technology, Aksu 843100, China

2 School of Computer and Information Engineering, Hubei University, Wuhan 431400, China

3 College of Information Science and Engineering, Xinjiang University, Ürümqi 830046, China

4 School of Electrical and Electronic Engineering, Wenzhou University, Wenzhou 325035, China

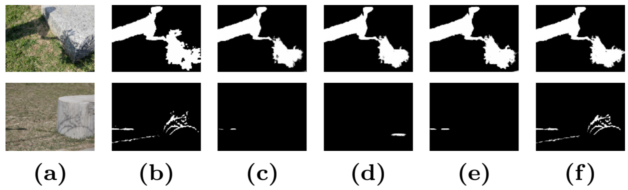


Fig. 1 Comparison with existing shadow detection methods in the hard shadow and soft shadow scene: **a** is the shadow images, **b** is the ground-truth shadow mask, **c** is the results of Wang [36], **d** is the results of Hu [10], **e** is the results of Zheng [42], **f** is our results

pay more attention to detect hard shadows since they can quickly get a rough edge by simple filtering to obtain binary shadow mask for supervised learning. However, such large-scale annotations for soft shadow are unavailable now due to prohibitive annotation costs and time-consuming. In this way, recent deep learning-based shadow detection methods [10,36,42] are good enough to detect hard shadow (first row in Fig. 1), but they cannot further to handle soft shadow (second row in Fig. 1). Considering soft shadow has some similar characteristics with hard shadow and many available benchmark datasets (i.e., SBU [25]), we introduce domain adaptation [37] from hard shadow to soft shadow. To be specific, we utilize the source domain with rich labeled data to help the model generalizing to the target domain.

There are two significant challenges to conduct domain adaptation for soft shadow detection. (1) *Domain discrepancy* Different capture devices, shooting environments, and exposure parameters will lead to large domain discrepancies in data collection procedures. In this work, the domain discrepancy [9,14,37] of data distributions mainly originates from different lighting conditions, different shooting countries, and the distances between the ground and the projection object. This way, directly employing hard shadow detection models to soft shadow scenes will not work; (2) *Intention difference* Hard shadow detection task mainly aims to distinguish the shadow regions, shadow-like non-shadow regions, and non-shadow regions, while the soft shadow detection task need to pay more attention at shadow regions with non-shadow patterns and then generate a structural edge. These two tasks are similar but not entirely the same, leading to a weak generation from the source domain (hard shadow) to the target domain (soft shadow). However, most existing domain adaptation methods [37,40] overlook the intention difference issue and adopt only one domain-shared detector to detect both hard shadow and soft shadow.

To tackle these two above issues, we propose a novel deep domain adaptation framework Soft-DA for soft shadow detection, including feature adversarial adaptation and detector separations strategy. Specifically, we reduce the domain discrepancy by aligning the feature distributions between two domains based on an adversarial learning manner. Owing to this, our Soft-DA can learn domain-invariant features for detection. Moreover, we introduce a novel detector separation strategy to overcome the intention difference, disentangling the detector into a domain-shared detector and two domain-specific detectors. The domain-shared detector tends to learn the task-shared detection knowledge among hard shadow and soft shadow, while the domain-specific detectors aim to conduct task-specific detection for only hard or soft shadow detection. For this purpose, domain-shared detector learns task-shared semantic information by aligning the two domains' joint distributions, while domain-specific detectors focus on task-specific detection information by maximizing the domain-shared detector's diversity and domain-specific detectors. In this way, we can detect soft shadow effectively from limited data.

The main contributions of this work are three aspects:

1. We create a Limited Soft Shadow Dataset (LSSD), which contains 1K shadow images and 0.4K corresponding shadow masks for semi-supervised learning.
2. To our best knowledge, we are the first to introduce the domain adaptation method for soft shadow detection. The extensive experiments demonstrate that our Soft-DA can detect hard shadow and soft shadow well.
3. We propose a novel detector separation strategy to further overcome the domain discrepancy and potential intention difference issue.

2 Related works

Shadow detection Tradition methods locate the shadow regions using shadow detection [7] or user annotation [1,13,39]. However, statistical learning-based methods rely too much on handcrafted shadow features. Although user annotation-based methods can accurately locate shadows, processing many images is time- consuming. For example, Khan et al. [13] and Shen et al. [32] take advantage of the representation learning ability of convolutional neural networks (CNNs) to learn the features of shadow, but these two networks are relatively shallow due to the limitation of dataset size. Vicente et al. [35] present a stacked CNN for large-scale shadow detection. Moreover, these methods process images in a patch-wise manner, so global optimization is necessary to maintain the consistency of the detection results. More recently, Nguyen et al. [25] first propose adversarial training for detection and introduce a novel conditional GAN with a sensitivity parameter. Mohajerani et al. [23] exploit a novel CNN architecture to identify and extract shadow features in an end-to-end manner based on U-Net. Le et al. [17] introduce a GAN-based framework in which a shadow detection network (D-Net) is trained together with a shadow attenuation network (A-Net) that generates adversarial training examples. Very recently, Zhu et al. [43] also develop the recurrent attention residual module to combine and process spatial contexts in two adjacent CNN layers. Zheng et al. [42] design attention modules by considering the distraction. More recently, Wang et al. [38] design an end-to-end framework named after Light-guided Instance Shadow-object Association to automatically predict the shadow and object instances, together with the shadow object associations and light direction.

Domain adaption Most existing domain adaptive methods try to alleviate the discrepancy between domains by adding an adaptation layer to match high-order moments of distributions [34] or designing a domain discriminator [3,29] to learn domain invariant features in an adversarial manner. Similarly, Lou et al. [20] propose category-aware domain adaptation instead of only conducting global feature alignment. In the task of object detection [9,14,37], they successfully propose a series of object detection methods to improve the model's generalization ability. However, these above methods ignore the intention difference between domains. By merely transferring these methods to soft shadow detection task, we cannot obtain a satisfactory result. So we propose a novel soft shadow domain adaptive method to solve this problem, which aims to alleviate domain discrepancy and overcome intension difference simultaneously.

3 A new dataset of limited soft shadow-LSSD

In recent years, large-scale datasets (i.e., SBU [35], ISTD [36]) have been created successively for shadow detection. Thanks to the massive training data, the accuracy of shadow detection is improving by a large margin. The SBU contains 4727 pairs of shadow/shadow mask for shadow detection only, while ISTD is consists of 1870 triplets of shadow/shadow mask/shadow-free for shadow detection and shadow removal.

However, these two above large-scale datasets have two significant drawbacks: (1) more than 98% percent of images in these datasets are hard shadow since the main goal of these training models is to distinguish the shadow regions, shadow-like non-shadow regions, and non-shadow regions. These models, however, cannot work well in soft shadow scenes. As we all know, the soft shadow is also widespread in daily life; (2) besides, especially in the ISTD dataset, it has only 0.1K+ unique scenes since most samples share the same background with different shadow positions. Moreover, there are only 20+ occluders in the dataset, limiting the diversity of cast shadow. Such limited scenes will hugely influence the understanding of scenes.

In this work, we create a limited dataset for soft shadow detection according to the following observations: (1) large-scale soft shadow annotation work will be complicated since the soft shadow needs to observe even through human eyes; (2) lacking of soft shadow training data, shadow detection model will not be suitable for soft shadow scenes.

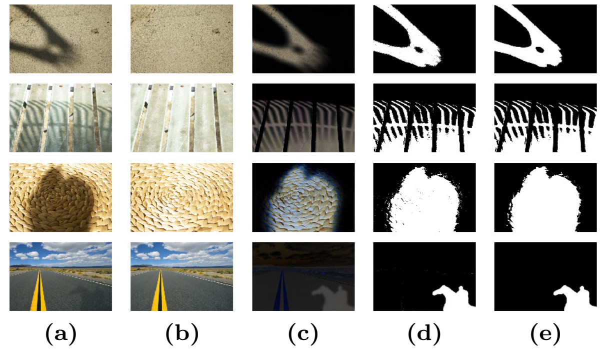


Fig. 2 Pipeline for annotating shadow masks of LSSD: **a** is the shadow images, **b** is the shadow-free images, **c** is the difference between **a** and **b**, **d** is the thresholding results of **c**, **e** is the shadow masks of manually adjusting

To this end, we propose a novel soft shadow detection model, which only needs a small amount of limited soft shadow training data. To be specific, we collect 1K soft shadow images and then annotate only 0.4K binary shadow mask for training and testing. We feed 0.3K labeled image pairs and 0.6K unlabeled shadow images into our semi-supervised learning model. Finally, we use the rest of the 0.1K labeled image pairs to validate our model.

Note that even soft shadow is hard to identify. As illustrated in Fig. 2a and b, we first take shadow and shadow-free image pairs by a fixed camera. Next, we threshold the difference (Fig. 2c) between Fig. 2a and b to obtain Fig. 2d. Then, we conduct morphological filtering and manually label mask adjusting to obtain the final shadow mask (Fig. 2e).

4 Our approach

In this section, we will introduce the overall strategy and details of Soft-DA. This work aims to train a model with traditional labeled shadow data (source domain) and a small amount of limited data (target domain) for soft shadow detection.

The main challenges in this work are: (1) we only feed a small amount of labeled soft shadow data into our network; (2) there is an apparent domain discrepancy [37] and a potential intention difference in soft shadow detection task. However, existing domain adaptation methods for object detection [9,14,37] only focus on reducing the domain discrepancy, while ignoring the intention difference causes perform poorly in target task. To solve this, we propose a novel deep domain adaptation framework for soft shadow detection, namely Soft-DA.

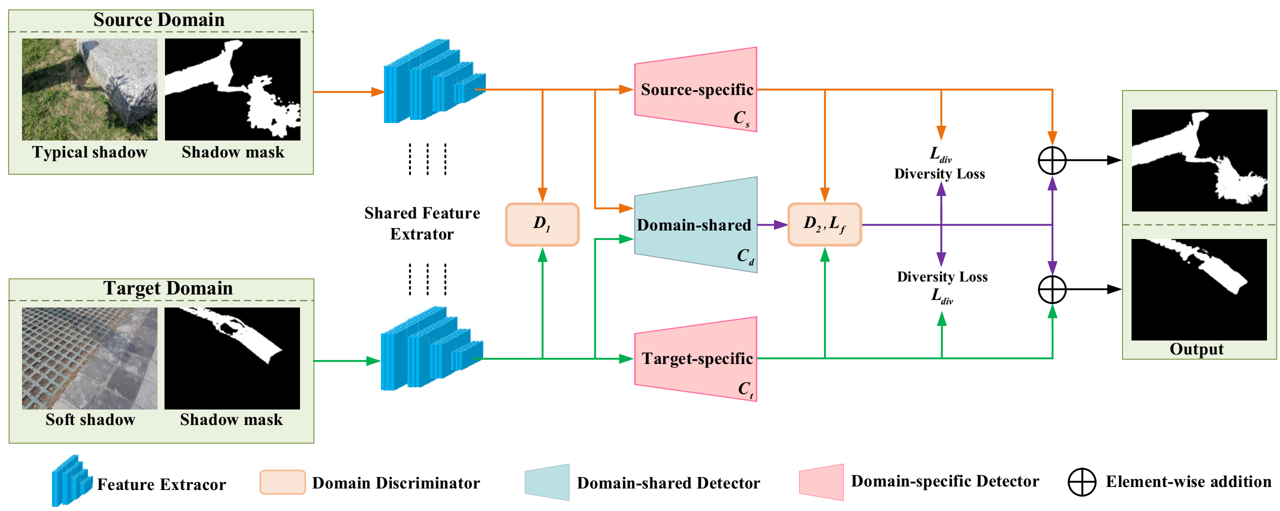


Fig. 3 Overall architecture and scheme of our Soft-DA

4.1 Overall scheme of Soft-DA

To enforce effective domain knowledge adaptation, we seek to alleviate the domain discrepancy with adversarial domain adaptation and handle the intention difference via a novel detector separation scheme. To this end, as shown in Fig. 3, Soft-DA consists of three main parts: (1) a domain-shared feature extractor G_f for extracting domain-invariant feature; (2) two domain discriminators $\{D_1, D_2\}$ for feature adaptation and detector adaptation, respectively; and (3) a domain-shared detector C_d and two domain-specific detectors $\{C_t, C_s\}$ for the shadow detecting. The separation of domain-shared and domain-specific detectors helps disentangle task-shared and task-specific illumination information regarding hard shadow and soft shadow.

Overall, Soft-DA has three main strategies as follows. (a) *Feature distribution alignment* We impose a domain loss L_{d_1} to align the feature distributions of two domains to minimize the domain discrepancy in an adversarial learning manner; (b) *Joint feature distribution adaptation* We exploit a domain loss L_{d_2} to conduct joint distribution alignment for the domain-shared detector C_d, making it able to learn domain-shared illumination information in an adversarial learning manner; (c) *Detector diversity maximization* We maximize the diversity between the domain-shared and domain-specific detectors via a diversity loss L_{div}, so that the domain-specific detectors can learn task-specific information for two different domains. Note that the strategies (b) and (c) help Soft-DA effectively tackle the intention difference. Finally, we train the discriminator D_2 via a focal loss L_f [18], leading the model class imbalance-aware and discriminative. This way, Soft-DA can adapt the source domain knowledge to the target domain and detect soft shadow accurately.

The overall training procedure of Soft-DA is to solve the following mini-max problem [5]:

$$\min\max\{-\alpha[L_{d_1} + L_{d_2}] + \beta L_{\text{div}} - L_f\}, \tag{1}$$

where α, β denote the weight of each component.

We next detail two domain loss L_{d_1}, L_{d_2}, and L_f in Sects. 4.2 and 4.3, respectively. Following that, we detail the detector diversity loss L_{div} in Sect. 4.4.

4.2 Feature distribution alignment

As previous mentioned, there are large domain discrepancies in data collection procedures. So we align the feature space distribution by domain adversarial learning manner [5,33]. Let $\mathbf{S} = \{\mathbf{x}_s, \mathbf{y}_s\}$ as labeled source data, $\mathbf{T}_l = \{\mathbf{x}_t^l, \mathbf{y}_t\}$ is a small amount of labeled target data, and $\mathbf{T}_u = \{\mathbf{x}_t^u, \phi\}$ is an unlabeled target data, where n_s, n_t^l, n_t^u represent the number of images in the source data, labeled target data, and unlabeled target data, respectively, where $n_t^l \ll n_t^u \ll n_s$, $\mathbf{T} = \mathbf{T}_l \cup \mathbf{T}_u$, and $n_t = n_t^l + n_t^u$.

We introduce a domain discrimination loss L_{d_1} to align the feature distribution between the two domains for eliminating the domain discrepancy. On the one hand, we minimize L_{d_1} to train the domain discriminator D_1 to identify the input image domain. On the other hand, we maximize L_{d_1} to train the feature extractor G_f for better confusing discriminator D_1. Owing to this, the G_f can extract domain-invariant features, which can confuse the discriminator well. Based on the least square distance [22], the domain loss L_{d_1} can be defined as follows:

$$L_{d_1} = \frac{1}{n_s}\sum_{\mathbf{x}_s \in \mathbf{S}} d_1(\mathbf{x}_s)^2 + \frac{1}{n_t}\sum_{\mathbf{x}_t \in \mathbf{T}} (1 - d_1(\mathbf{x}_t))^2, \tag{2}$$

where $d_1(\mathbf{x}) = D_1(G_f(\mathbf{x}))$ is the prediction of d_1. We set the 1 and 0 for the label of the target domain and the source domain, respectively.

4.3 Joint feature distribution adaptation

Existing adversarial domain adaptation methods [3,15,37,40] argue that the source and target domain deal with same task. Therefore, they often only focus on aligning the feature distribution between the two domains and then use the detector trained on source data to process the data in the target domain. These strategy, however, are not entirely suitable for soft shadow detection. In this work, we call this phenomenon as intention difference.

Let $P(\mathrm{f}^s)$ and $P(\mathrm{f}^t)$ represent the feature distributions in the source, and the target domains, $P(\mathbf{y}^s|f^s)$ and $P(\mathbf{y}^t|f^t)$ is the prediction conditional distributions of these two domains. In Sect. 4.2, we have aligned the feature distributions (i.e., $P(f^s) = P(f^t)$), but the potential intention difference will lead to different prediction conditional distributions(i.e., $P(\mathbf{y}^s|f^s) \neq P(\mathbf{y}^t|f^t)$) [21,24]. Therefore, only training a detector by hard shadow data will not detect soft shadow well. To reduce the intention difference, we propose a novel detector separation strategy, which separates a domain-shared detector C_d and two domain-specific detectors $\{C_s, C_t\}$.

Specifically, after aligning the feature distribution ($P(f^s) = P(f^t)$), we further train the detector to align the joint feature distributions ($P(\mathbf{y}^s|f^s)=P(\mathbf{y}^t|f^t)$). The domain shared detector will learn task-shared detection knowledge [19]. Inspired by it, we combine the domain adversarial learning manner with training a domain-shared detector by aligning the joint distributions. In detail, we first minimize the domain loss L_{d_2} to train discriminator D_2, which can distinguish the joint distribution between domains adequately. And then, we maximum L_{d_2} to train a domain-shared detector C_d to confuse D_2 better. Based on the least square distance, the loss function can be defined as follows:

$$L_{d_2} = \frac{1}{n_s}\sum_{x_s \in \mathbf{S}} d_2(\mathbf{x})^2 + \frac{1}{n_t}\sum_{x_t \in \mathbf{T}} (1 - d_2(\mathbf{x}))^2, \tag{3}$$

where $d_2(x) = D_2(C_d(x))$ denotes the detection result of domain discriminator.

In essence, the domain discriminator D_2 aims to solve a binary classification problem. Considering the imbalance of sample categories (i.e., hard shadow and soft shadow), we use the focal loss [18] as follows:

$$L_f = \frac{1}{n_t^l + n_s}\sum_{\mathbf{T}_l \cup \mathbf{S}} d^{\mathrm{T}}((1-\hat{d})^{\gamma} \odot \log(\hat{d})), \tag{4}$$

where d is the label of the domains, $\hat{d} = d_2(\mathbf{x})$, and $\odot$ represents the element-wise production, and γ is a hyper-parameter. Note that the focal loss is widely used for class-imbalance issues.

4.4 Maximum the diversity of detector

For better solve the intension difference, we will further train the two domain-specific detectors to learn the domain-private detection knowledge. To this end, we maximize the diversity of the domain-specific detectors $\{C_s, C_t\}$ and the domain-shared detector C_d.

Given a shadow image $\mathbf{x}$, let $\hat{\mathbf{y}}_d$ denote the domain-shared detectors prediction result and let $\hat{\mathbf{y}}_{\mathbf{t}}$ and $\hat{\mathbf{y}}_{\mathbf{s}}$ represent the prediction results of the domain-specific detector C_t and C_s. We maximum the diversity of these detectors are as follows:

$$L_{\mathrm{div}} = \frac{1}{n_s}\sum_{x_s \in \mathbf{S}} L_{\mathrm{ce}}(\hat{\mathbf{y}}_d, \hat{\mathbf{y}}_s) + \frac{1}{n_t}\sum_{x_t \in \mathbf{T}} L_{\mathrm{ce}}(\hat{\mathbf{y}}_d, \hat{\mathbf{y}}_t), \tag{5}$$

where L_{ce} is cross-entropy and we define it as $L_{\mathrm{ce}}(y_1, y_2) = -(y_2 \log y_1 + (1-y_2)\log(1-y_1))$. To this end, two domain-specific detectors will be as different as possible from the domain-shared detector, which can better learn domain-specific detection knowledge. Besides, the final prediction result is the average ensemble of the two detectors. We maximize the diversity between detectors can enhance the performance of ensemble learning.

5 Experiments

In this section, we first introduce the implementation details (Sect. 5.1), evaluation datasets and evaluation metrics (Sect. 5.2). We then compare our results both quantitatively and qualitatively to the existing shadow detection methods from the following two aspects: (1) the performance of soft shadow detection and (2) the properties of our algorithm (Sect. 5.3). Finally, we conduct thorough ablation studies to analyze the proposed model (Sect. 5.4).

5.1 Implemented details

In our experiments, we create our network in TensorFlow on a computer with Intel i9-10900K and RTX 2080Ti. Specifically, we first implement our feature extractor based on ResNet-18 [8] and then construct two domain discriminators based on a two-layer fully connected network and implement all detectors decoder architecture based on U-net [28]. In the training process, we use the SGD optimizer with a learning rate of 0.001 for training the overall network. We set the batch size of each domain to 16. In our experiments, we set α= 0.1, β = 0.1, γ = 2.0 for better results. We show the training and inference details of Soft-DA in Algorithms 5.1 and 5.1.

Algorithm 1 Training stage

Input: Labeled source data $\mathbf{S} = \{\mathbf{x}_s, \mathbf{y}_s\}$, labeled target data $\mathbf{T}_l = \{\mathbf{x}_t^l, \mathbf{y}_t\}$, and unlabeled target data $\mathbf{T}_u = \{\mathbf{x}_t^u, \phi\}$. Training epoch M; hyper-parameter α, β and γ.

1: **for** $m = 1 \to M$ **do**
2: Extract image feature **f** using G_f.
3: Obtain detection results $\{\hat{\mathbf{y}}_d, \hat{\mathbf{y}}_t, \hat{\mathbf{y}}_s\}$ of the detector $\{C_d, C_t, C_s\}$, respectively.
4: Calculate domain adversarial losses L_{d_1}, L_{d_2}(Eq. 2 and Eq. 3)via $\{C_d, D_1, D_2\}$.
5: **if x** is labeled data **then**
6: Compute the domain classification loss L_f (Eq. 4) based on $\{\hat{\mathbf{y}}_d, \hat{\mathbf{y}}_t\}$ and $\{\hat{\mathbf{y}}_d, \hat{\mathbf{y}}_s\}$.
7: **end if**
8: We calculate detection diversity loss L_{div}(Eq. 5) based on $\{\hat{\mathbf{y}}_d, \hat{\mathbf{y}}_t, \hat{\mathbf{y}}_s\}$.
9: Compute the total loss (Eq. 1).
10: Back propagation.
11: **end for**

Output: G_f, C_d, and C_t.

Algorithm 2 Inference stage

Input: Soft shadow image **x**, the detector $\{C_d, C_t\}$ for the target domain.

1: Extract the feature of **x** via G_f.
2: Compute the detection result $\{\hat{\mathbf{y}}_d, \hat{\mathbf{y}}_t\}$ by $\{C_d, C_t\}$.
3: Compute the ensemble detection result $\hat{\mathbf{y}} = (\hat{\mathbf{y}}_d + \hat{\mathbf{y}}_t)/2$.

Output: $\hat{\mathbf{y}}$.

Table 1 Detail of the two datasets, where SBU servers as the source domain and LSSD servers as the target domain

Set	Domain	Number	
		Labeled	Unlabeled
Training	SBU	4727	0
	LSSD	300	600
Test	LSSD	100	0

5.2 Dataset and evaluation metrics

Datasets We set SBU [35] and LSSD dataset are source domain and target domain in our experiments, respectively. The number of training and testing samples is shown in Table 1.

Evaluation metrics The detection task needs to consider the unbalanced distribution of different types of pixels. To measure the accuracy of shadow detection in this paper, we use the Balance Error Rate (BER) as the following:

$$\text{BER} = 1 - \frac{1}{2}\left(\frac{\text{TP}}{\text{TP} + \text{FN}} + \frac{\text{TN}}{\text{TN} + \text{FP}}\right), \tag{6}$$

where TP, TN, FP, and FN are the number of true positive, true negative, false positive, and false negative pixels.

Table 2 Quantitative soft shadow detection results on LSSD dataset

Methods	BER	Shadow	Nonshadow
Zheng [42]-only source	14.1	13.3	14.6
Zheng [42]-only target	**14.5**	**14.1**	**15.5**
Zheng [42]-fine-tuning	13.8	13.2	14.1
Wang [38]-fine-tuning	13.2	12.5	13.4
BEAL for shadow detection	10.6	9.2	11.1
Soft-DA	*8.8*	*7.9*	*9.3*

5.3 Comparison with state-of-the-arts

In fact, soft shadow detection is more challenging than hard shadow detection, and the ability of soft shadow detection is also an essential indicator of the generalization ability. We select various shadow images with different complex scenes and then perform extensive experiments.

5.3.1 Soft shadow detection

For soft shadow detection (1) Source-only: only use labeled source data to train Zheng [42]; (2) Target-only: use only a small number of labeled target data to train Zheng; (3) Fine-tuning: after training on the source data for Zheng and Wang [38], use a small amount of label data for fine-tuning on the target domain; (4) BEAL [37]: a semi-supervised learning method via domain adaptation, which use the labeled source data and partially labeled target dataset for training.

As shown in Table 2 and Fig. 4, we measure and visualize the detection result of above all methods on LSSD dataset. We can observe that the soft shadow detection ability of Soft-DA is the best. We can also get the following observation: (1) the target-only model and fine-tuning model do not both certainly outperform the source-only model. We argue that the deep learning model may have overfitted in the target domain, so its performance is limited; (2) the semi-supervised domain adaptation method (BEAL) further improves the models performance on the target domain. This result proves the concept of domain adaptation and the contribution of a small amount of labeled target domain data; (3) our Soft-DA outperforms all the above methods, which shows that the superiority of Soft-DA to leverage both well-labeled source data and partially labeled target data.

5.3.2 Hard shadow detection

For hard shadow detection we compare our shadow detection results with some state of the arts including one traditional method, i.e., Guo [7], and five recent deep learning-based methods, i.e., ScGAN [25], CPNet [23], BDRAR [43], Zheng [42], and Wang [38]. For the fair comparison, we train all the

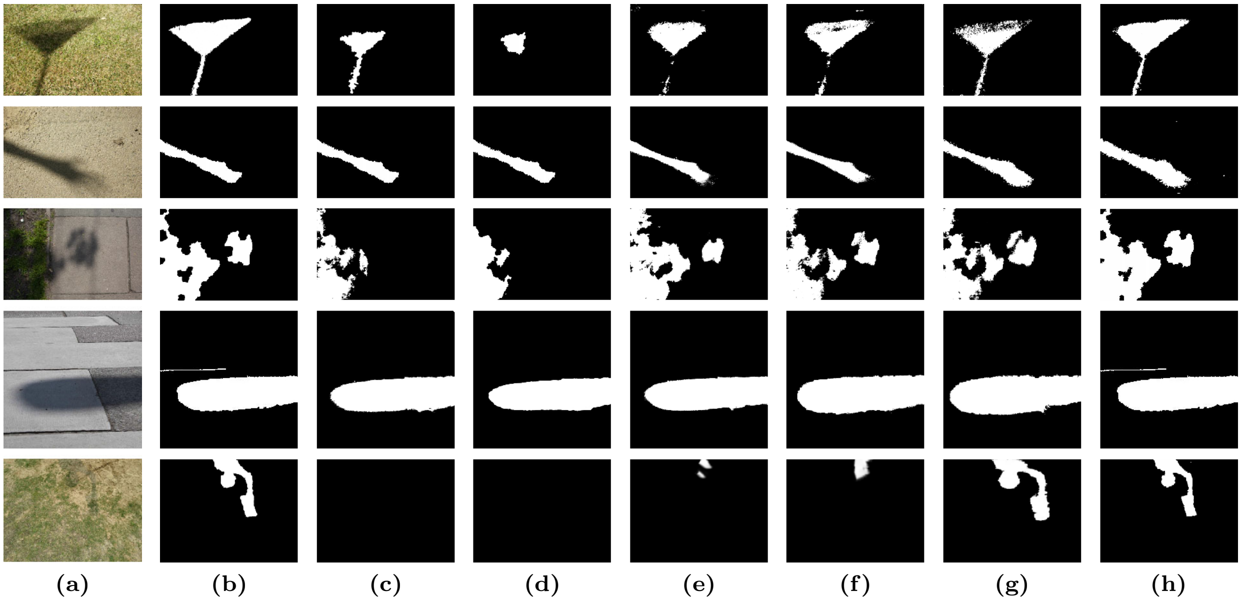


Fig. 4 Visual comparison of soft shadow detection results in LSSD dataset: **a** is the input images, **b** is the ground-truth shadow masks, **c** is the results of source-only based on Zheng [42], **d** is the results of target-only based on Zheng [42], **e** is the results of fine-tuning based on Zheng [42], **f** is the results of fine-tuning based on Wang [38], **g** is the results of BEAL [37], **h** is our results

other models and our proposed Soft-DA on the same training data and evaluate the shadow detection performance on the SBU, ISTD dataset.

We summarize the hard shadow detection results in Table 3. As we can see, our Soft-DA works the best BER on all datasets among all the competing methods, which can demonstrate that our strategy can also help our Soft-DA detect accurate hard shadow regions, although we have no particular attention-aware components designed in our shadow detection generator.

To further explain the performance of our proposed Soft-DA, we provide some visualization results in Figs. 5 and 6, including the traditional methods and the learning-based methods. As we can observe, (1) traditional method, i.e., Guo [7] is not able to usefully detect shadows in the image(the second, fourth and, fifth row); (2) Among all deep learning methods, comparing ScGAN [25], CPNet [23], BDRAR [43], Zheng [42], and Wang [38], our proposed Soft-DA can detect more fine-grained shadow details(e.g., the boundary of shadow) and even more close to our human observation.

Table 3 Quantitative shadow detection results on SBU and ISTD datasets

Methods	Venue/year	SBU	ISTD
Guo	CVPR/2011	25.03	27.16
ScGAN	ICCV/2017	13.23	5.72
CPNet	MMSP/2018	7.64	4.95
BDRAR	ECCV/2018	7.34	4.51
Zheng	CVPR/2019	6.52	4.19
Wang	CVPR/2020	**6.23**	**4.32**
Soft-DA	2021	*6.17*	*4.01*

5.4 Ablation study

We conduct an ablation study to verify the effectiveness of each component in Soft-DA. In this work, both cross-entropy loss L_{ce} and diversity loss L_{div} are designed for shadow detection; domain losses L_{d_1} and L_{d_2} are designed for domain adaptation; L_f is designed for domain classification. We remain the network architecture and training procedure of Soft-DA and create different loss functions to train our proposed network.

As shown in Table 4 and Fig. 7, we find that all components make empirical contributions and play essential roles in our method, where domain confrontation losses L_{d_1} and L_{d_2} are relatively more critical than other components. Besides, the detector diversity loss L_{div} can further improve the performance of soft shadow detection.

6 Conclusion

In this work, we propose a deep domain adaptation method for detecting soft shadow, which aims to transfer domain knowledge from a fully labeled source domain (i.e., hard shadow data) to a partially labeled target domain (i.e., soft

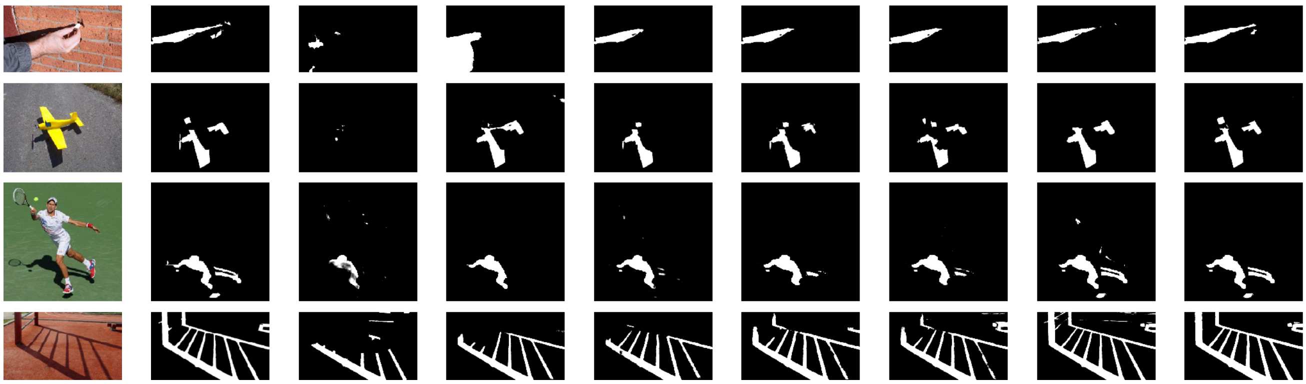


Fig. 5 Visual comparison of hard shadow detection results in SBU dataset: **a** is the input images, **b** is the ground-truth shadow masks, **c** is the results of Guo [7], **d** is the results of ScGAN [25], **e** is the results of CPNet [23], **f** is the results of BDRAR [43], **g** is the results of Zheng [42], **h** is the result of Wang [38], **i** is our results

Fig. 6 Visual comparison of soft shadow detection results in ISTD dataset: **a** is the input images, **b** is the ground-truth shadow masks, **c** is the results of Guo [7], **d** is the results of ScGAN [25], **e** is the results of CPNet [23], **f** is the results of BDRAR [43], **g** is the results of Zheng [42], **h** is the result of Wang [38], **i** is our results

Table 4 Ablation study on Soft-DA at different regions

Methods	Backbone	L_{ce}	L_{d_1}	L_{d_2}	L_f	L_{div}	BER	Shadow	Non-shadow
Soft-DA$_1$	✓	✓					**14.3**	**13.7**	**14.9**
Soft-DA$_2$	✓	✓	✓				12.5	11.7	13.4
Soft-DA$_3$	✓	✓	✓	✓			9.8	9.4	9.9
Soft-DA$_4$	✓	✓	✓	✓	✓		9.3	8.6	10.5
Soft-DA	✓		✓	✓	✓	✓	*8.8*	*7.9*	*9.3*

shadow data). To be specific, we align the feature distributions of the two domains through adversarial domain adaptation to reduce the domain discrepancy. Besides, we develop a novel detector separation scheme to overcome the intention difference between domains. In this way, our proposed method can detect soft shadow well with a few soft shadow annotations. In future, we will combine the above domain adaptation method further to soft shadow removal.

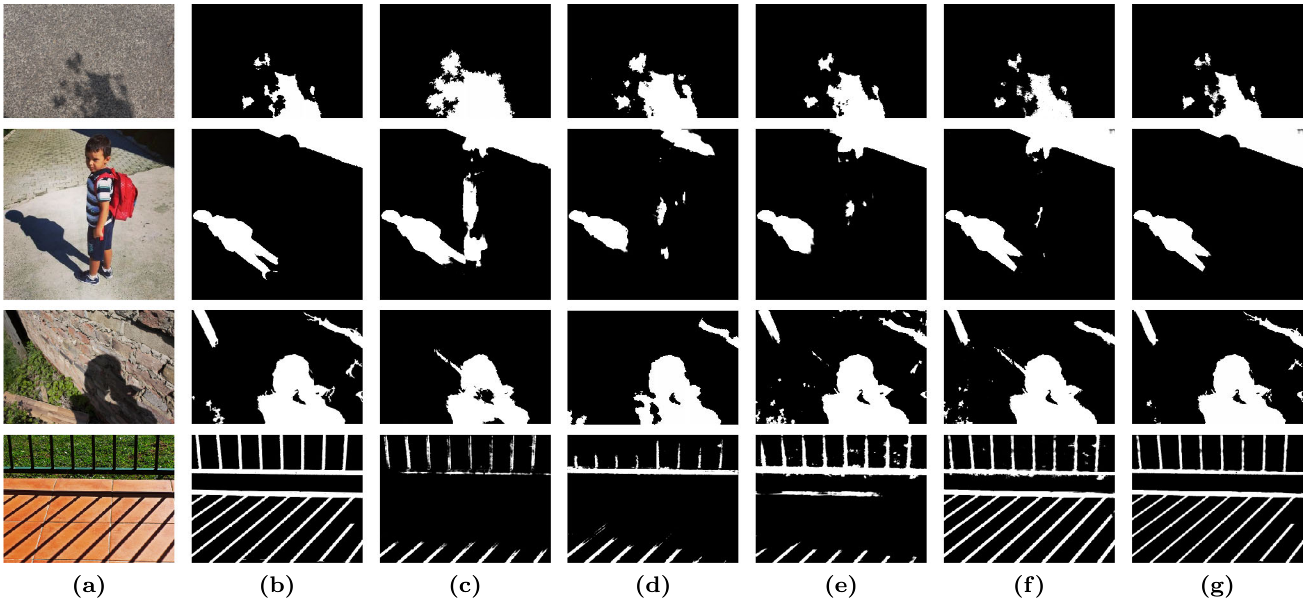


Fig. 7 Shadow detection results: **a** is the input images, **b** is the ground-truth shadow masks, **c** is the results of Soft-DA_1, **d** the results of Soft-DA_2, **e** is the results of Soft-DA_3, **f** is the results of Soft-DA_4, **g** is the results of Soft-DA

Acknowledgements This work is supported by the Natural Science Foundation of Xinjiang Autonomous Region in China (No. 2020D01A48) and the National Social Science Foundation Western Project (No. 20XGL029).

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

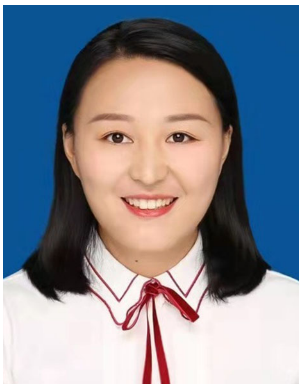
1. Arbel, E., Hel-Or, H.: Shadow removal using intensity surfaces and texture anchor points. IEEE Trans. Pattern Anal. Mach. Intell. **33**(6), 1202–1216 (2010)
2. Bandara, R., Ranathunga, L., Abdullah, N.A.: Deep learned compact binary descriptor with a lightweight network-in-network architecture for visual description. Vis. Comput. (2020). https://doi.org/10.1007/s00371-020-01798-5
3. Ganin, Y., Lempitsky, V.: Unsupervised domain adaptation by backpropagation. In: International Conference on Machine Learning, pp. 1180–1189 (2015)
4. Gong, H., Cosker, D.: Interactive shadow removal and ground truth for variable scene categories. In: BMVC, pp. 1–11 (2014)
5. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: Advances in Neural Information Processing Systems, pp. 2672–2680 (2014)
6. Gryka, M., Terry, M., Brostow, G.J.: Learning to remove soft shadows. ACM Trans. Graph. **34**(5), 153 (2015)
7. Guo, R., Dai, Q., Hoiem, D.: Single-image shadow detection and removal using paired regions. In: CVPR 2011, pp. 2033–2040. IEEE (2011)
8. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
9. Hsu, H.K., Yao, C.H., Tsai, Y.H., Hung, W.C., Tseng, H.Y., Singh, M., Yang, M.H.: Progressive domain adaptation for object detection. In: The IEEE Winter Conference on Applications of Computer Vision, pp. 749–757 (2020)
10. Hu, X., Zhu, L., Fu, C.W., Qin, J., Heng, P.A.: Direction-aware spatial context features for shadow detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7454–7462 (2018)
11. Kán, P., Kafumann, H.: Deeplight: light source estimation for augmented reality using deep learning. The Vis. Comput. **35**(6–8), 873–883 (2019)
12. Karsch, K., Hedau, V., Forsyth, D., Hoiem, D.: Rendering synthetic objects into legacy photographs. ACM Trans. Graph. (TOG) **30**(6), 1–12 (2011)
13. Khan, S.H., Bennamoun, M., Sohel, F., Togneri, R.: Automatic shadow detection and removal from a single image. IEEE Trans. Pattern Anal. Mach. Intell. **38**(3), 431–446 (2015)
14. Khodabandeh, M., Vahdat, A., Ranjbar, M., Macready, W.G.: A robust learning approach to domain adaptive object detection. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 480–490 (2019)
15. Lafarge, M.W., Pluim, J.P., Eppenhof, K.A., Moeskops, P., Veta, M.: Domain-adversarial neural networks to address the appearance variability of histopathology images. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, pp. 83–91. Springer (2017)
16. Lalonde, J.F., Efros, A.A., Narasimhan, S.G.: Estimating natural illumination from a single outdoor image. In: 2009 IEEE 12th International Conference on Computer Vision, pp. 183–190. IEEE (2009)
17. Le, H., Vicente, T.F.Y., Nguyen, V., Hoai, M., Samaras, D.: A+D net: training a shadow detector with adversarial shadow attenuation. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 680–696 (2018)

18. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2980–2988 (2017)
19. Long, M., Zhu, H., Wang, J., Jordan, M.I.: Deep transfer learning with joint adaptation networks. In: International Conference on Machine Learning, pp. 2208–2217 (2017)
20. Luo, Y., Zheng, L., Guan, T., Yu, J., Yang, Y.: Taking a closer look at domain shift: category-level adversaries for semantics consistent domain adaptation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2507–2516 (2019)
21. Luo, Z., Zou, Y., Hoffman, J., Fei-Fei, L.F.: Label efficient learning of transferable representations acrosss domains and tasks. In: Advances in Neural Information Processing Systems, pp. 165–177 (2017)
22. Mao, X., Li, Q., Xie, H., Lau, R.Y., Wang, Z., Paul Smolley, S.: Least squares generative adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2794–2802 (2017)
23. Mohajerani, S., Saeedi, P.: Cpnet: a context preserver convolutional neural network for detecting shadows in single rgb images. In: 2018 IEEE 20th International Workshop on Multimedia Signal Processing (MMSP), pp. 1–5 (2018)
24. Motiian, S., Piccirilli, M., Adjeroh, D.A., Doretto, G.: Unified deep supervised domain adaptation and generalization. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 5715–5725 (2017)
25. Nguyen, V., Yago Vicente, T.F., Zhao, M., Hoai, M., Samaras, D.: Shadow detection with conditional generative adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 4510–4518 (2017)
26. Nielsen, M., Madsen, C.B.: Graph cut based segmentation of soft shadows for seamless removal and augmentation. In: SCIA'07 Proceedings of the 15th Scandinavian Conference on Image Analysis, pp. 918–927 (2007)
27. Okabe, T., Sato, I., Sato, Y.: Attached shadow coding: estimating surface normals from shadows under unknown reflectance and lighting conditions. In: 2009 IEEE 12th International Conference on Computer Vision, pp. 1693–1700. IEEE (2009)
28. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 234–241. Springer (2015)
29. Saito, K., Watanabe, K., Ushiku, Y., Harada, T.: Maximum classifier discrepancy for unsupervised domain adaptation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3723–3732 (2018)
30. Sanchez-Matilla, R., Poiesi, F., Cavallaro, A.: Online multi-target tracking with strong and weak detections. In: European Conference on Computer Vision, pp. 84–99. Springer (2016)
31. Savarese, S., Rushmeier, H., Bernardini, F., Perona, P.: Shadow carving. In: Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001, vol. 1, pp. 190–197. IEEE (2001)
32. Shen, L., Wee Chua, T., Leman, K.: Shadow optimization from structured deep edge detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2067–2074 (2015)
33. Tzeng, E., Hoffman, J., Saenko, K., Darrell, T.: Adversarial discriminative domain adaptation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7167–7176 (2017)
34. Tzeng, E., Hoffman, J., Zhang, N., Saenko, K., Darrell, T.: Deep domain confusion: maximizing for domain invariance. arXiv preprint arXiv:1412.3474 (2014)
35. Vicente, T.F.Y., Hou, L., Yu, C.P., Hoai, M., Samaras, D.: Large-scale training of shadow detectors with noisily-annotated shadow examples. In: European Conference on Computer Vision, pp. 816–832. Springer (2016)
36. Wang, J., Li, X., Yang, J.: Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1788–1797 (2018)
37. Wang, S., Yu, L., Li, K., Yang, X., Fu, C.W., Heng, P.A.: Boundary and entropy-driven adversarial learning for fundus image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 102–110. Springer (2019)
38. Wang, T., Hu, X., Wang, Q., Heng, P.A., Fu, C.W.: Instance shadow detection. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1880–1889 (2020)
39. Wu, T.P., Tang, C.K., Brown, M.S., Shum, H.Y.: Natural shadow matting. ACM Trans. Graph. (TOG) **26**(2), 8-es (2007)
40. Zhang, Y., Chen, H., Wei, Y., Zhao, P., Cao, J., Fan, X., Lou, X., Liu, H., Hou, J., Han, X., et al.: From whole slide imaging to microscopy: deep microscopy adaptation network for histopathology cancer image classification. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 360–368. Springer (2019)
41. Zheng, M., Lei, Z., Zhang, K.: Intelligent detection of building cracks based on deep learning. Image Vis. Comput. **103**, 103987 (2020)
42. Zheng, Q., Qiao, X., Cao, Y., Lau, R.W.: Distraction-aware shadow detection. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5167–5176 (2019)
43. Zhu, L., Deng, Z., Hu, X., Fu, C.W., Xu, X., Qin, J., Heng, P.A.: Bidirectional feature pyramid network with recurrent attention residual modules for shadow detection. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 122–137 (2018)
44. Zhu, X., Chen, Z.: Dual-modality spatiotemporal feature learning for spontaneous facial expression recognition in e-learning using hybrid deep neural network. Vis. Comput. **36**, 1–13 (2019)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Wen Wu received the BSc degree in Electronic and Information Engineering from Wuhan College of Arts & Science in 2016. Then, he received the MSc degree in Technology of Computer Application from Hubei University in 2019. He is currently an assistant at the School of Information Engineering, Xinjiang Institute of Technology, China. His research areas include computer vision, image processing, and deep learning.

Shuping Zhang received the BSc degree in Applied Mathematics from Shihezi University in 2003. Then, she received the MSc degree in Computer Science from Xinjiang University in 2006. She is currently an associate professor at the School of Information Engineering, Xinjiang Institute of Technology, China. Her research areas include computer vision, image processing, and deep learning.



Mi Tian received the BSc degree in Physical Education from Hunan College of Arts & Sciences in 2016. Then, she received the MSc degree in Physical Education from Shenyang Sport University in 2019. She is currently an assistant at the School of Information Engineering, Xinjiang Institute of Technology, China. Her research areas include computer vision, image processing, and deep learning.

Daoqiang Tan received the BSc degree in Internet of Things Engineering from the Hubei Institute of Business and Technology in 2017. He is currently an MSc candidate at the Hubei University. His research areas include computer vision, image processing, and deep learning.

Xiantao Wu received the BSc degree in Electronic and Information Engineering from Wuhan College of Arts & Science in 2016. She is currently an MSc candidate at the Xinjiang University. Her research areas include computer vision, image processing, and deep learning.

Yi Wan received the PhD degree from Southwest Jiaotong University in 2006. He is currently a Professor at the School of Electrical and Electronic Engineering, Wenzhou University, China. His research areas include computer vision, image processing, and deep learning.