Instance Shadow Detection

Tianyu Wang[1,2,*], Xiaowei Hu[1,*], Qiong Wang[2], Pheng-Ann Heng[1,2], and Chi-Wing Fu[1,2]
[1] Department of Computer Science and Engineering, The Chinese University of Hong Kong
[2] Shenzhen Key Laboratory of Virtual Reality and Human Interaction Technology,
Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China

Abstract

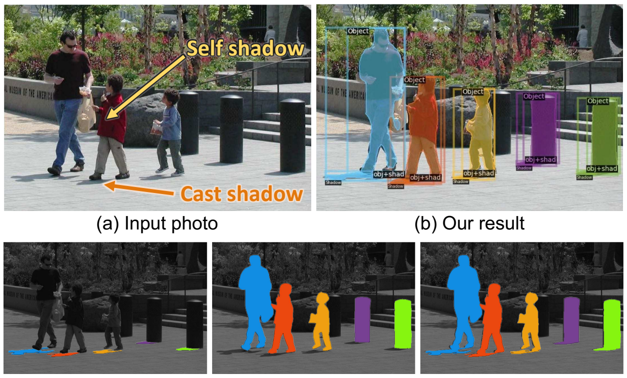
Instance shadow detection is a brand new problem, aiming to find shadow instances paired with object instances. To approach it, we first prepare a new dataset called SOBA, named after Shadow-OBject Association, with 3,623 pairs of shadow and object instances in 1,000 photos, each with individually-labeled masks. Second, we design LISA, named after Light-guided Instance Shadow-object Association, an end-to-end framework to automatically predict the shadow and object instances, together with the shadow-object associations and light direction. Then, we pair up the predicted shadow and object instances and match them with the predicted shadow-object associations to generate the final results. In our evaluations, we formulate a new metric named the shadow-object average precision to measure the performance of our results. Further, we conducted various experiments and demonstrate our method's applicability to light direction estimation and photo editing.

1. Introduction

"*When you light a candle, you also cast a shadow,*"—Ursula K. Le Guin written in A Wizard of Earthsea.

When some objects block the light, shadows are formed. And when we see a shadow, we also know that there must be some objects that create or cast the shadow. Shadows are light-deficient regions in a scene, due to light occlusion, but they carry the shape of the light-occluding objects, as they are projections of these objects onto the physical world. In this work, we are interested in a new problem, *i.e.*, *finding shadows together with their associated objects*.

Concerning shadows, prior works in computer vision and image understanding focus mainly on shadow detection [15, 18, 19, 21, 22, 26, 46, 50, 54] and shadow removal [8, 16, 17, 25, 37, 47]. Our goal in this work is to leverage the remarkable computation capability of deep neural networks to address the new problem of associating shadows and objects—*instance shadow detection*. That is, we want to detect the shadow instances in images, together with the associated object that casts each shadow.



(a) Input photo
(b) Our result
(c) Shadow instance mask
(d) Object instance mask
(e) Shadow-object association mask

Figure 1. Given a photo with shadows (a), the problem of *instance shadow detection* is to detect the individual shadow instances (c) and the individual object instances (d), as well as to associate the shadows with the objects (e) that cast them. (b) shows the prediction results produced by our method on (a).

Being able to find shadow-object associations has the potentials to benefit various applications. For example, for privacy protection, when we remove humans and cars from photos, we can remove objects and associated shadows altogether. In a recent work on removing objects from images for privacy protection [42], the shadows are simply left behind. Also, when we edit photos, say by scaling or translating objects, we can naturally manipulate objects with their associated shadows simultaneously. Further, shadow-object associations give hints to the light direction in the scene, supporting applications such as relighting.

To approach the problem of instance shadow detection, first, we prepare a new dataset called *SOBA*, named after *Shadow OBject Association*. SOBA contains 3,623 pairs of shadow-object associations over 1,000 photos, each with three masks (see Figures 1 (c)-(e)): (i) shadow instance mask, where we label each shadow instance with a unique color; (ii) shadow-object association mask, where we label

[*] Joint first authors

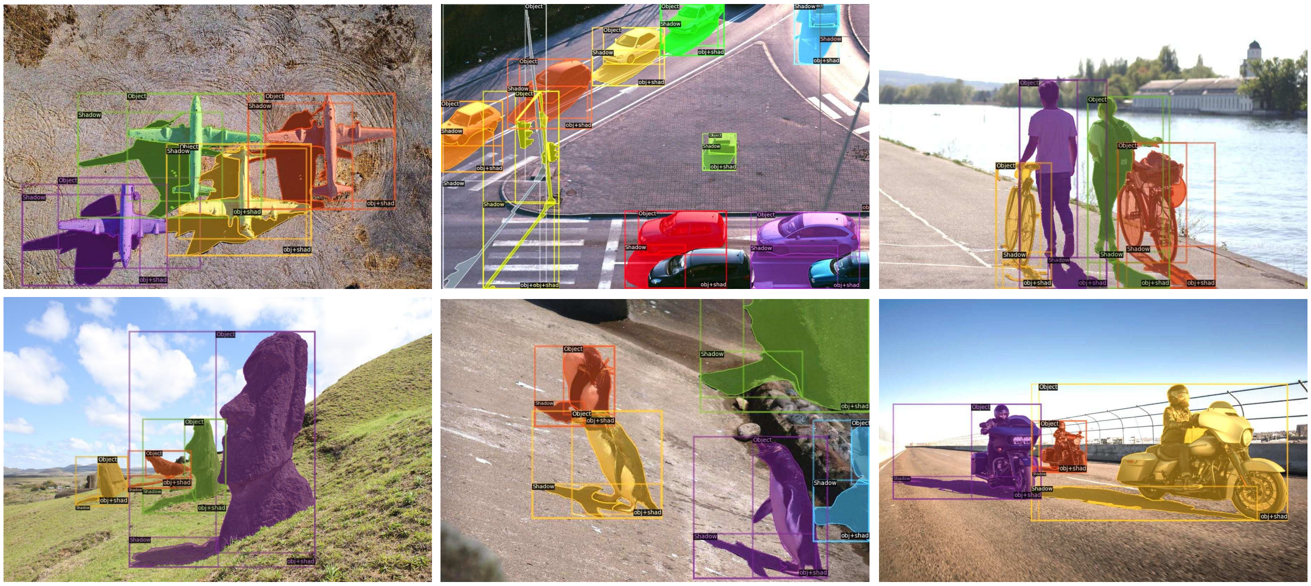


Figure 2. Example images with mask and box labels in our SOBA dataset. Please zoom in for better visualization.

each shadow-object pair with a corresponding unique color; and (iii) object instance mask, which is (ii) minus (i). In general, there are two types of shadows: (i) *cast shadows*, formed on background objects, usually ground, as the projections of the light-occluding objects, and (ii) *self shadows*, formed on the side of the light-occluding objects opposite to the direct light (see Figure 1(a)). In this work, we consider mainly cast shadows, which are object projections, since self shadows are already on the associated objects. See also Figure 2 for example images in our SOBA dataset.

Next, we design an end-to-end framework called *LISA*, named after *Light-guided Instance Shadow-object Association*, to find the individual shadow and object instances, the shadow-object associations, and the light direction in each shadow-object association. From these predictions, we then use a simple yet effective method to pair the predicted shadow and object instances and to match them with the predicted shadow-object associations.

Third, to quantitatively measure and evaluate the performance of the instance shadow detection results, we formulate a new evaluation metric called SOAP, named after *Shadow-Object Average Precision*. In the end, we further perform a series of experiments to show the effectiveness of our method and demonstrate its applicability to light direction estimation and photo editing.

2. Related Work

Shadow detection. Early works [39, 33, 41] made use of physical illumination and color models, and analyzed the spectral and geometrical properties of shadows. Later, machine learning methods were explored to detect shadows by modeling shadows based on handcrafted features, *e.g.*, texture [53, 43, 12, 45], color [24, 43, 12, 45], T-junction [24], and edge [24, 53, 20], then by using various classifiers, *e.g.*, decision tree [24, 53] and SVM [12, 20, 43, 45], to differentiate shadows and non-shadows. However, physical models and handcrafted features have limited feature representation capability, thus they are not robust in general situations.

Later, convolutional neural networks (CNN) were introduced to detect shadows. Khan *et al.* [21] and Shen *et al.* [40] used CNN to learn high-level features and optimization methods to detect shadows. Vicente *et al.* [46] trained a fully-connected network to predict a shadow probability map, then locally refine the shadows via a patch-CNN.

More recently, end-to-end networks were designed to detect shadows. Nguyen *et al.* [32] built a conditional generative adversarial network with a sensitive parameter to stabilize the network training. Hu *et al.* [16, 19] and Zhu *et al.* [54] explored the spatial context via the direction-aware spatial context module and recurrent attention residual module, respectively. Wang *et al.* [47] and Ding *et al.* [8] jointly detected and removed shadows by using multiple networks or a multi-branch network. To improve the detection performance, Le *et al.* [26] proposed to generate more training samples, while Zheng *et al.* [50] combined the strengths of multiple methods to explicitly revise the results. This work explores a new problem on detecting shadows, namely *instance shadow detection*. Unlike general shadow detection, which finds only a single mask for all shadows in an image, we design a deep architecture to find not just the individual shadows but also the associated objects altogether.

Instance segmentation. Besides, this work relates to the emerging problem of instance segmentation, where the goal is to label pixels of individual foreground objects in the in-

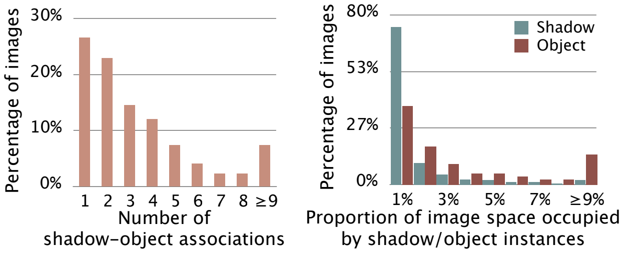


Figure 3. Statistical properties of the SOBA dataset.

put image. Overall, there are two major approaches to the problem: proposal-based and proposal-free approaches.

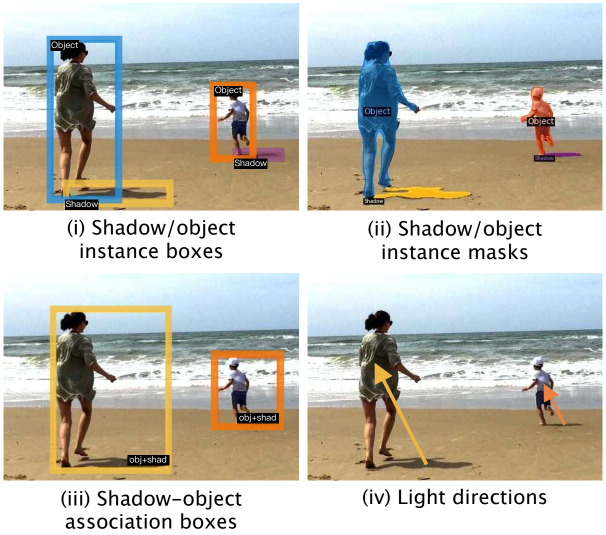
Proposal-based approach generally uses object detectors to propose candidates and classifies the candidates to find object instances, *e.g.*, MNC [6], DeepMask [35], InstanceFCN [6], and SharpMask [36]. Later, FCIS [27] jointly detects and segments the object instances using a fully convolutional network. BAIS [13] models the object shapes and segments the object instances in a boundary-aware manner. MaskLab [4] uses a network with three outputs for box detection, semantic segmentation, and direction prediction, while methods based on Mask R-CNN [14], *e.g.*, [31, 3, 34], achieved great performance by simultaneously detecting the object instances and predicting the segmentation masks.

The proposal-free approach [1, 2, 23, 30] first classifies the image pixels, then group the pixels into object instances. Recently, TensorMask [5] leverages a fully convolutional network for dense mask prediction, while SSAP [9] predicts the object instance labels in just a single pass.

3. SOBA (Shadow OBject Association) Dataset

We collected 1,000 images from the ADE20K [51, 52], SBU [15, 44, 46], ISTD [47], and Microsoft COCO [29] datasets, and also from the Internet using keyword search with shadow plus animal, people, car, athletic meeting, zoo, street, etc. Then, we coarsely label the images to produce the shadow instance masks and shadow-object association masks, and refine them using Apple Pencil; see Figures 1 (c) & (e). Next, we obtain the object instance masks (see Figure 1 (d)) by subtracting each shadow instance mask from the associated shadow-object association mask. Overall, there are 3,623 pairs of shadow-object instances in the dataset images, and we randomly split the images into a training set (840 images, 2,999 pairs) and a testing set (160 images, 624 pairs); see Figure 2 for some examples.

Figure 3 shows some statistical properties of the SOBA dataset. From the histogram shown on the left, we can see that SOBA has a diverse number of shadow-object pairs per image, with around 3.62 pairs per image on average. Also, it contains many challenging cases: 7% of the images have nine or more shadow-object pairs per image. On the other hand, the histogram shown on the right reveals the proportion of image space (horizontal axis) occupied, respectively, by the shadow and object instances in the dataset images. From the plot, we can see that most shadows and objects occupy relatively small areas in the whole images, demonstrating the challenges to detect them.



(i) Shadow/object instance boxes

(ii) Shadow/object instance masks

(iii) Shadow-object association boxes

(iv) Light directions

Figure 4. Example predictions (output) from our LISA framework.

4. Methodology

4.1. Overall Network Architecture of LISA

Compared with shadow detection, the challenges of instance shadow detection are that we have to predict shadow instances rather than just a single mask for all the shadows in the input image. Also, we have to find object instances in the input image and pair them up with the shadow instances. To meet these challenges, we design an end-to-end framework called LISA, named after Light-guided Instance Shadow-object Association. Overall, as shown in Figure 5, LISA takes a single image as input and predicts

(i) a box of each shadow/object instance,

(ii) a mask of each shadow/object instance,

(iii) a box of each shadow-object association (pair), and

(iv) the light direction for each shadow-object association.

Figure 4 shows a set of example outputs. Particularly, LISA predicts the light direction and takes it as guidance to find shadow-object associations, since the light direction is usually consistent with the shadow-object associations.

Figure 5 shows the architecture of LISA, which begins by using a convolutional neural network (ConvNet) to extract semantic features from the input image. Here, we use the feature pyramid network [28] as the backbone ConvNet. Then, we design a two-branch architecture: the top branch predicts the box and mask for each shadow/object instance and the bottom branch predicts the box for each shadow-object association and the associated light direction.

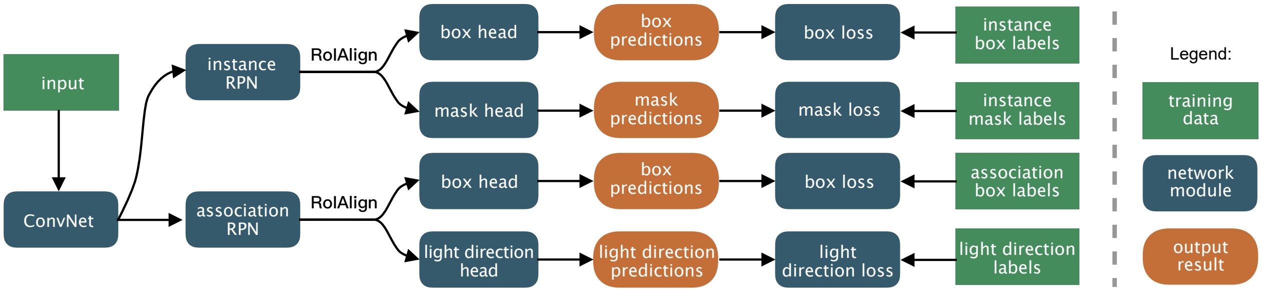


Figure 5. The schematic illustration of our Light-guided Instance Shadow-object Association (LISA) framework.

In detail, the top branch starts with the instance region proposal network (RPN) [38] to find region proposals, which are regions with high probabilities of containing the shadow/object instances. Then, we adopt RoIAlign [14] to extract features for each proposal and leverage the box and mask heads to predict the boxes and masks for the shadow and object instances by minimizing the loss between the prediction results and the supervision signals from the training data. Please refer to Mask R-CNN [14] for the detail. On the other hand, the bottom branch adopts an association RPN to generate region proposals for the shadow-object associations, then uses RoIAlign to extract features for each proposal and adopts the box head to produce the bounding boxes of the shadow-object associations. After obtaining the associations, we can then efficiently obtain the masks of the shadow-object associations by combining the shadow and object masks predicted from the top branch. Note that the parameters in the box head are learned by minimizing the loss between the boxes of the predicted shadow-object associations and the ground-truth associations.

Besides, we design a light direction head in parallel with the box head of the bottom branch to predict an angle that represents the estimated light direction from shadow to object in each association pair. Note that we compute the ground-truth angle θ^g of the light direction by

$$\theta^g \;=\; \mathrm{atan2}(\; y_o^g - y_s^g ,\; x_o^g - x_s^g \;) \;,$$

where (x_s^g, y_s^g) and (x_o^g, y_o^g) are 2D coordinates of the shadow and object instance centroids in the ground-truth image, and $\mathrm{atan2}(y,x)$ is a variation of the $\arctan(y/x)$ function to avoid anomaly and output a full-range polar angle in $(-\pi, \pi]$. The shadow-object association branch and light direction branch share the common feature extraction network and the association RPN. By jointly optimizing the predictions of the light direction and shadow-object association in each region proposal, we can improve the overall performance of instance shadow detection; see the experimental results in Section 5.

4.2. Pairing up Shadow and Object Instances

The raw predictions of LISA include shadow instances, object instances, shadow-object associations, and a light direction predicted per association. Note that, the predicted shadow and object instances are not paired, whereas the predicted shadow-object associations are not separated as shadows and objects. Also, some of these predictions may not be correct, and they may also contradict one another. Hence, we have to analyze these predictions, pair up the predicted shadow and object instances, and match them with the predicted shadow-object associations, so that we can find and output the final paired shadow and object instances.

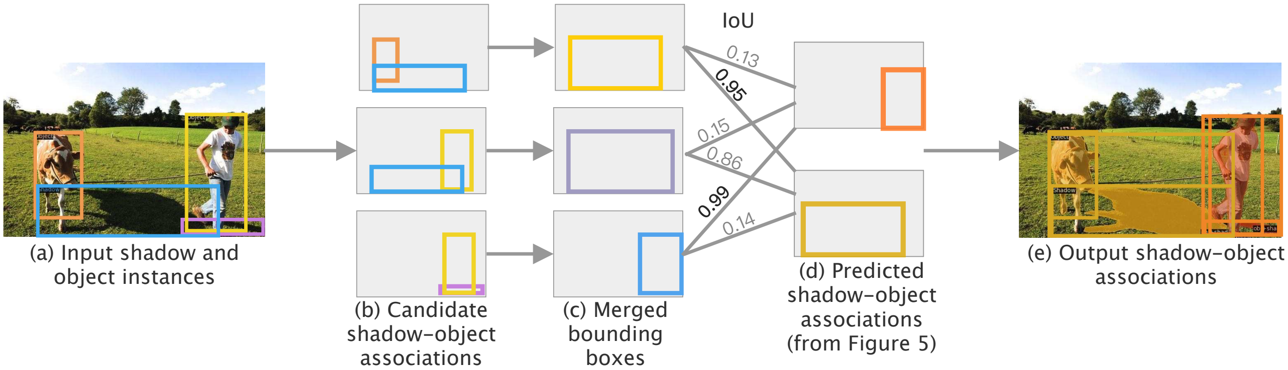
Figure 6 illustrates the procedure, where we first find candidate shadow-object associations (see Figure 6 (b)) by (i) computing the shortest distance between the bounding boxes of every pair of shadow and object instances, and (ii) regarding a pair as a candidate association, if the computed distance is smaller than a threshold, which is empirically set as the height of the associated shadow instance. After that, we construct bounding box B_i for the i-th candidate pair (see Figure 6 (c)) by merging the bounding boxes of the associated shadow and object instances. Given ($x^s_{\min}$,$y^s_{\min}$) and ($x^s_{\max}$,$y^s_{\max}$) as the lower-left and upper-right corners of the shadow instance bounding box, and ($x^o_{\min}$,$y^o_{\min}$) and ($x^o_{\max}$,$y^o_{\max}$) as the lower-left and upper-right corners of the object instance bounding box, the corners of the merged bounding box B_i are given by

$$\begin{aligned} &\big(\;\min(x^s_{\min}, x^o_{\min}) \;\;,\;\; \min(y^s_{\min}, y^o_{\min}) \;\big) \;, \\ \text{and} \;\; &\big(\;\max(x^s_{\max}, x^o_{\max}) \;,\; \max(y^s_{\max}, y^o_{\max}) \;\big) \;. \end{aligned}$$

In the end, as illustrated in Figure 6 (d), we compute the Intersection over Union (IoU) between the merged boxes and the shadow-object association boxes predicted independently in LISA (see Figure 5), and select those with the highest IoUs as the final shadow-object associations. Then, for each of these associations, we can get back the associated shadow instance and object instance, and pair them as the final outputs; see Figure 6 (e).

4.3. Training Strategies

Loss function. We optimize LISA by jointly minimizing the instance box loss, instance mask loss, association box loss, light direction loss (see Figure 5), and the losses of instance RPN and association RPN. The loss functions of



(a) Input shadow and object instances

(b) Candidate shadow–object associations

(c) Merged bounding boxes

(d) Predicted shadow–object associations (from Figure 5)

(e) Output shadow–object associations

Figure 6. The pair-and-match procedure for pairing the predicted shadow and object instances and efficiently matching them with the predicted shadow-object associations.

boxes, masks, and RPNs follow the formulations in Mask R-CNN [14], whereas the light direction loss L_{light} is formulated by a smooth L_1 loss [10], as follows:

$$L_{light}(\theta^p, \theta^g) \;=\; \begin{cases} 0.5\,(\theta^p - \theta^g)^2 & \text{if } |\theta^p - \theta^g| < 1 \\ |\theta^p - \theta^g| - 0.5 & \text{otherwise,} \end{cases}$$

where θ^p and θ^g are the predicted and ground-truth angles of the light direction, respectively.

Training parameters. We train our LISA framework by following the training strategies of Mask R-CNN implemented on Facebook Detectron2 [48]. Specifically, we adopt the weights of ResNeXt-101-FPN [28, 49] trained on ImageNet [7] to initialize the parameters of the backbone network, and train our framework on two GeForce GTX 1080 Ti GPUs (four images per GPU) for $40k$ training iterations. We set the base learning rate as 1e-4, adopt a warm-up [11] strategy to linearly increase the learning rate to 1e-3 during the first 1,000 iterations, keep the learning rate as 1e-3, and stop the learning after $40k$ iterations. We re-scale the input images, such that the longer side is less than 1,333 and the shorter side is less than 800 without changing the image aspect ratio. Lastly, we randomly apply horizontal flips on the images for data augmentation.

5. Experiments

5.1. Evaluation Metrics

Existing metrics evaluate instance segmentation results by looking at object instances individually. Our problem involves multiple types of instances: shadows, objects, and their associations. Hence, we formulate a new metric called the *Shadow-Object Average Precision* (SOAP) by adopting the same formulation as the traditional average precision (AP) with the intersection over union (IoU) but further considering a sample as true positive (an output shadow-object association), if it satisfies the following three conditions:

(i) the IoU between the predicted shadow instance and ground-truth shadow instance is no less than τ;

(ii) the IoU between the predicted object instance and ground-truth object instance is no less than τ; and

(iii) the IoU between the predicted and ground-truth shadow-object associations is no less than τ.

Table 1. Comparing our full pipeline with two simplified baseline frameworks on the bounding boxes of the final shadow-object associations in terms of SOAP$_{50}$, SOAP$_{75}$, and SOAP.

Method	box SOAP$_{50}$	box SOAP$_{75}$	box SOAP
Baseline 1	40.3	14.0	16.7
Baseline 2	47.8	14.0	19.6
Our full pipeline	**50.5**	**16.4**	**21.8**

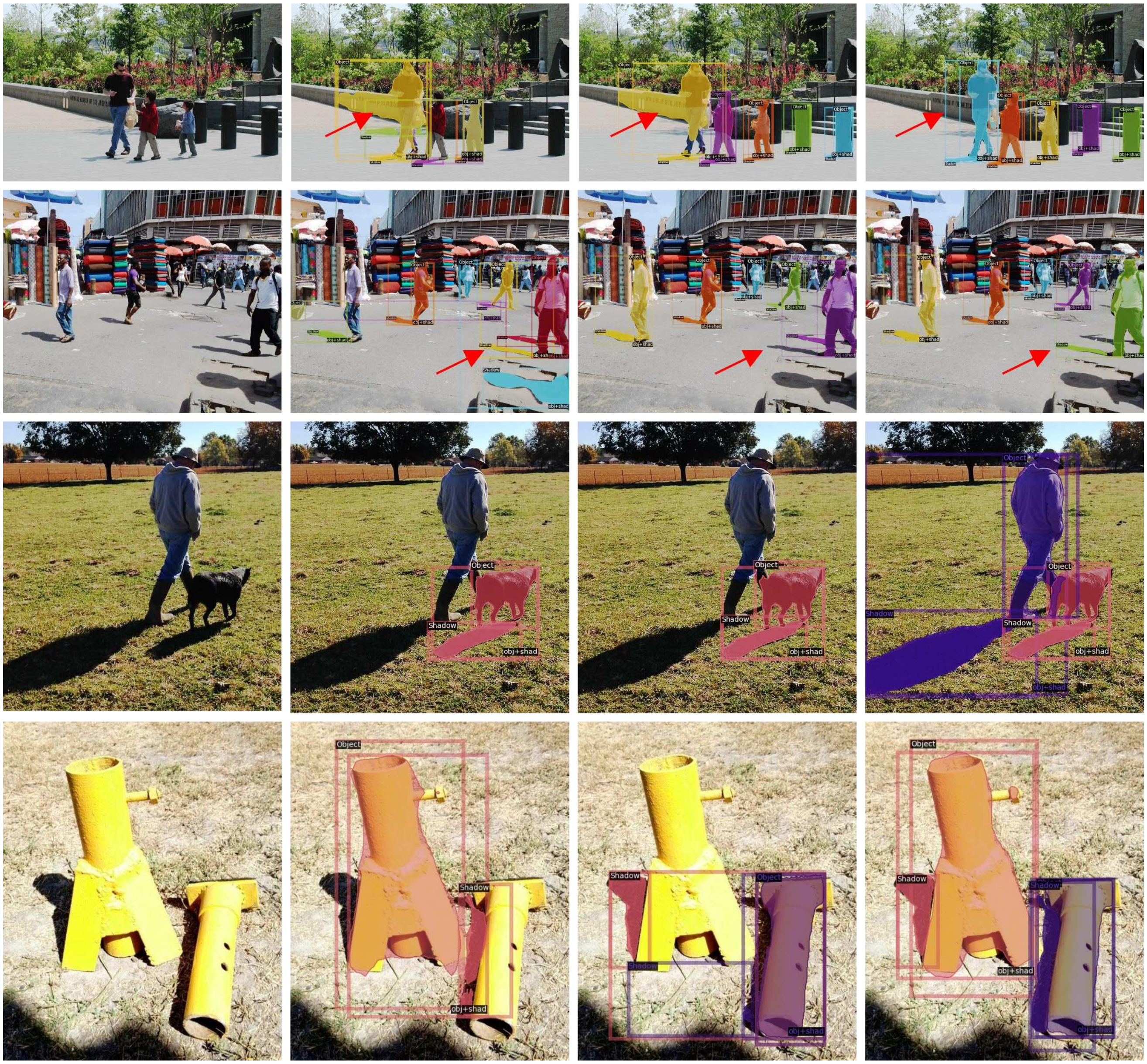
Table 2. Comparing our full pipeline with two simplified baseline frameworks on the masks of the final shadow-object associations in terms of SOAP$_{50}$, SOAP$_{75}$, and SOAP.

Method	mask SOAP$_{50}$	mask SOAP$_{75}$	mask SOAP
Baseline 1	41.0	10.0	16.7
Baseline 2	48.1	12.5	20.1
Our full pipeline	**50.9**	**14.4**	**21.6**

We follow [29] to report the evaluation results by setting τ as 0.5 (SOAP$_{50}$) or 0.75 (SOAP$_{75}$), and also report the average over multiple τ [0.5:0.05:0.95] (SOAP). Moreover, since we can obtain the bounding boxes as well as the masks for the shadow instances, object instances, and shadow-object associations, we further report SOAP$_{50}$, SOAP$_{75}$, and SOAP in terms of both bounding boxes and masks.

5.2. Results

Evaluation. To evaluate the LISA framework, we set up (i) Baseline 1, which adopts only the top branch of LISA to predict bounding boxes and masks of the shadow and object instances, then merges them to form shadow-object associations based on the proximity between the shadow and object instances; and (ii) Baseline 2, which removes the light direction head in LISA when predicting the shadow-object associations, but still adopts the procedure to pair-and-match the shadow and object instances (Section 4.2).



(a) Input image (b) Baseline 1 (c) Baseline 2 (d) Full pipeline

Figure 7. Visual comparison of instance shadow detection results produced by our full pipeline and two other baseline frameworks.

Tables 1 and 2 report the quantitative comparison results in terms of the bounding boxes and masks in the final detected shadow-object associations. Comparing different rows in the results, we can see that Baseline 2 clearly improves over Baseline 1, demonstrating that we can obtain better shadow-object associations in our deep end-to-end framework by independently predicting also the shadow-object associations and then pairing the shadow and object instances and matching them with the predicted shadow-object associations. Moreover, by further predicting the light direction and taking it as the guidance to jointly optimize the framework, our full pipeline LISA achieves the best performance for all the evaluation metrics.

Figure 7 shows visual comparison results for Baseline 1, Baseline 2, and our full pipeline. The first column shows the input images, whereas the second, third, and fourth columns show the results produced by the two baselines and our full pipeline. By comparing Baseline 1 with Baseline 2, we can see that further learning to detect the shadow-object associations independently in the deep framework helps to discover more shadow-object pairs, as shown in the third and fourth rows in Figure 7. Moreover, after taking the light direction as guidance (Baseline 2 vs. full pipeline), our method improves the performance in various challenging cases, *e.g.*,

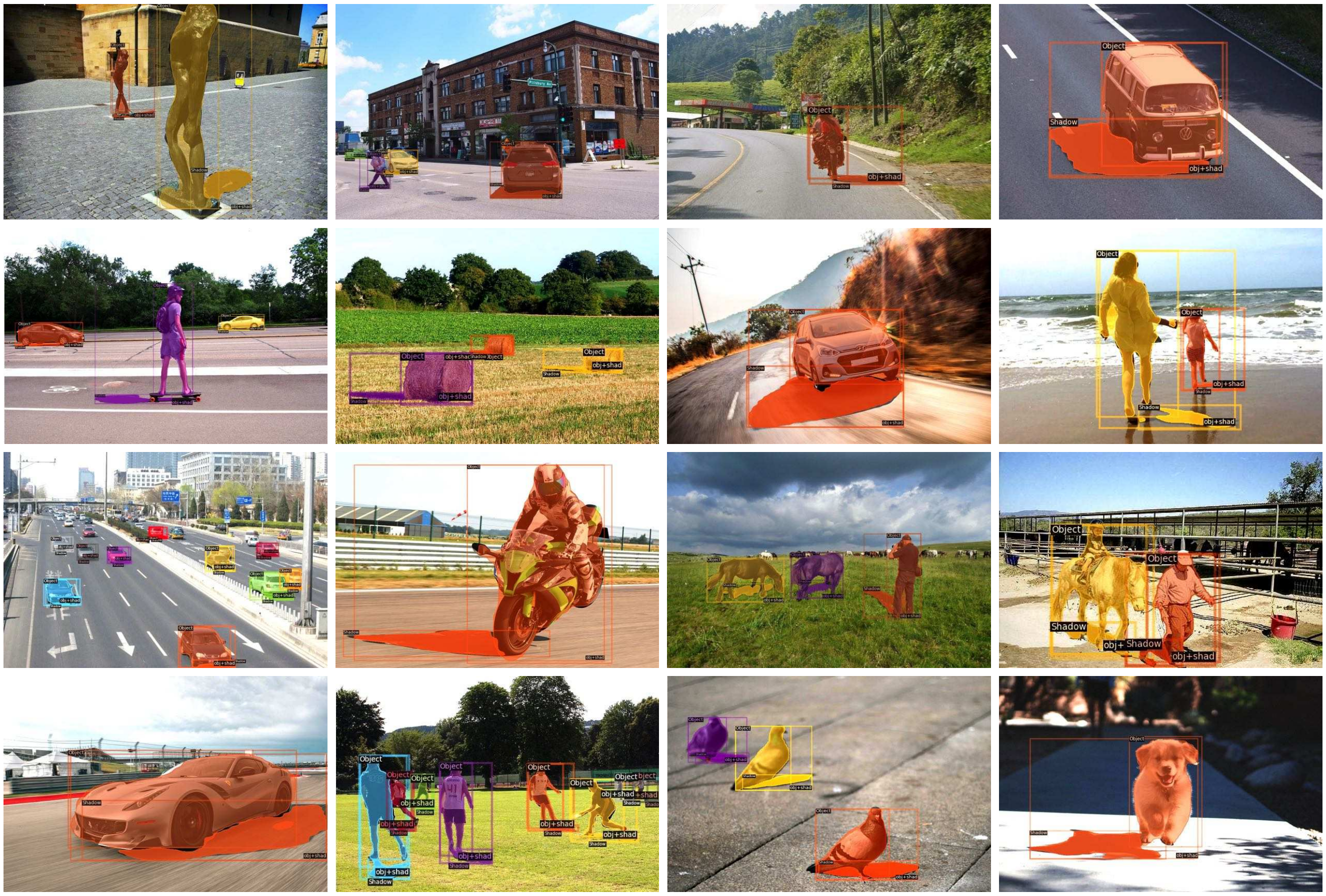


Figure 8. Instance shadow detection results produced by our method over a wide variety of photos and objects.

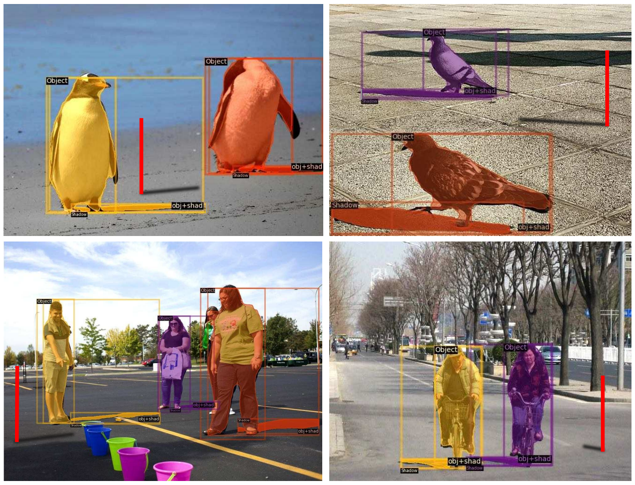


Figure 9. Example images, where we estimate the light directions and incorporate virtual red posts with simulated shadows.

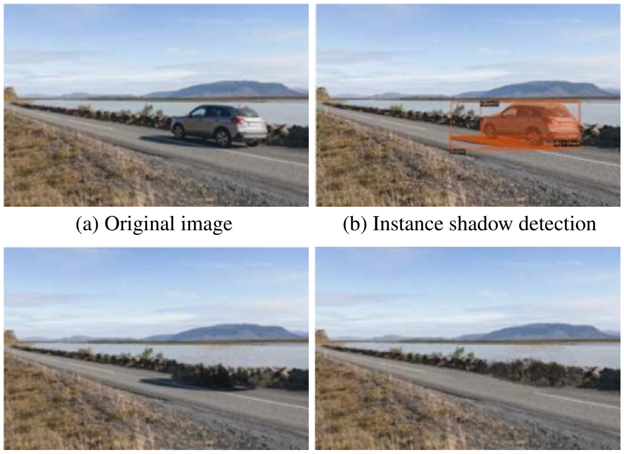
when there is large but irrelevant shadow region nearby (see the first row), when there are multiple shadow instances connect with a single object instance (see the second row), when the centers of the shadow and object instances are far from each other (see the third row), and when there are multiple shadow regions near a single object instance (see the last row). Please see Figure 8 and supplemental material for more instance shadow detection results produced by our method on various types of images and objects.

6. Applications

Below, we present application scenarios to demonstrate the applicability of the results produced by our method.

Light direction estimation. First, instance shadow detection helps to estimate the light direction in a single 2D image, and we connect the centers of the bounding boxes of the shadow and object instances in each shadow-object association pair as the estimated light direction. Figure 9 shows some example results, where for each photo, we estimate the light direction and render a virtual red post with a simulated shadow on the ground based on the estimated light direction. From the results, we can see that the virtual shadows with the red posts look consistent with the real shadows cast by other objects, thus demonstrating the applicability of our detection results.

Photo editing. Another application to demonstrate instance shadow detection is photo editing, where we can



(a) Original image (b) Instance shadow detection

(c) An example result in [42] (d) Enhanced by our result

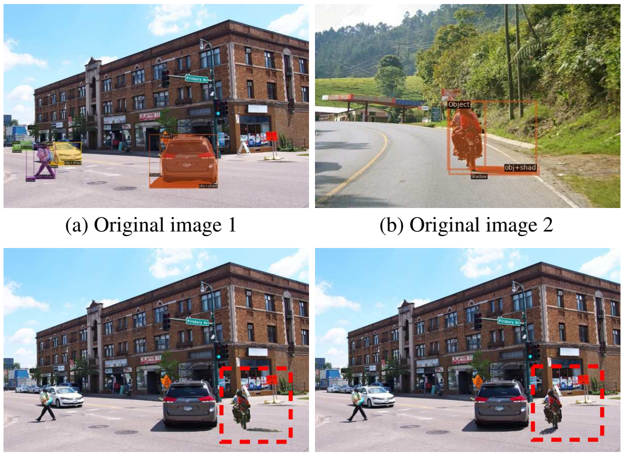
Figure 10. Instance shadow detection enables us to easily remove objects (*e.g.*, vehicle) with their associated shadows altogether.

remove not only the object instances but also their associated shadows altogether. For privacy protection, Uittenbogaard *et al.* [42] presents a method to automatically remove specific objects in street-view photos; see Figure 10 (c) for a result, where it can successfully remove the vehicle. However, the shadow cast by the vehicle remains on the ground. With the help of our instance shadow detection result (Figure 10 (b)), we can remove the vehicle with its shadow altogether, as shown in Figure 10 (d).

Further, we can more efficiently transfer an object together with its shadow from one photo to another photo. Figure 11 presents an example, we cut the motorcycle with its shadow from (b) and paste them into (a) in smaller sizes. Clearly, if we simply paste the motorcycle and shadow to (a), the shadow is not consistent with the real shadows in the target photo; see (c). Thanks to instance shadow detection, which outputs individual masks for both object and shadow instances, as well as light directions. Therefore, we can achieve light-aware photo editing by making use of the estimated light direction in both photos to adjust the shadow images when transferring the motorcycle from one photo to the other; see (d).

7. Conclusions and Limitations

In this paper, we presented instance shadow detection, which targets to find shadow instances and object instances, and pair them up together. Also, we presented three technical contributions to approach the problem. First, we prepare SOBA, a new dataset of 1,000 images and 3,623 pairs of shadow-object associations, where we provide the input photos together with a set of three instance masks. Second, we develop LISA, an end-to-end deep framework, to predict boxes and masks of individual shadow and object instances, as well as boxes of shadow-object associations and the associated light directions; from these predictions, we further match the shadow and object instances, and pair them up to match with the predicted shadow-object associations and light directions for producing the output shadow-object pairs. Third, we formulate SOAP, a new evaluation metric for quantitatively measuring the instance shadow detection results, enabling us to perform various experiments to compare with baseline frameworks. In the end, we also demonstrate the applicability of our results on light direction estimation and photo editing.



(a) Original image 1 (b) Original image 2

(c) Naïve cut-and-paste (d) Light-aware shadow

Figure 11. When we cut-and-paste objects from one photo to the other, instance shadow detection results enable us not only to extract object and shadow instances together, but also to adjust the shadow shape based on the estimated light direction.

As the first attempt to detect shadow-object instances, we admit that there are many possible methods that can be explored to improve the detection performance. Besides methodologies, we did not consider the overlap between shadow instances associated with different objects. Also, we did not consider cast shadows formed on some other object instances. There are many open problems and unexplored situations for instance shadow detection.

In the future, we plan to first improve the performance of instance shadow detection by simultaneously leveraging multiple training data from the current datasets prepared for shadow detection and instance segmentation. By exploring semi- or weakly-supervised methods to learn to detect instance shadows, we could combine the strengths and knowledge from various data to better the performance of instance shadow detection. Last, we will also explore more applications based on the shadow-object association results.

Acknowledgments. This work was supported by the Research Grants Council of the Hong Kong Special Administrative Region (Project no. CUHK 14203416), the CUHK Research Committee Direct Grant for Research 2018/19, and the Science and Technology Key Project of Guangdong Province, China (2019B010149002).

References

[1] Anurag Arnab and Philip H. S. Torr. Pixelwise instance segmentation with a dynamically instantiated network. In *CVPR*, pages 441–450, 2017. 3

[2] Min Bai and Raquel Urtasun. Deep watershed transform for instance segmentation. In *CVPR*, pages 5221–5229, 2017. 3

[3] Kai Chen, Jiangmiao Pang, Jiaqi Wang, Yu Xiong, Xiaoxiao Li, Shuyang Sun, Wansen Feng, Ziwei Liu, Jianping Shi, Wanli Ouyang, Chen Change Loy, and Dahua Lin. Hybrid task cascade for instance segmentation. In *CVPR*, pages 4974–4983, 2019. 3

[4] Liang-Chieh Chen, Alexander Hermans, George Papandreou, Florian Schroff, Peng Wang, and Hartwig Adam. MaskLab: Instance segmentation by refining object detection with semantic and direction features. In *CVPR*, pages 4013–4022, 2018. 3

[5] Xinlei Chen, Ross Girshick, Kaiming He, and Piotr Dollár. TensorMask: A foundation for dense object segmentation. In *ICCV*, pages 2061–2069, 2019. 3

[6] Jifeng Dai, Kaiming He, and Jian Sun. Instance-aware semantic segmentation via multi-task network cascades. In *CVPR*, pages 3150–3158, 2016. 3

[7] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. ImageNet: A large-scale hierarchical image database. In *CVPR*, pages 248–255, 2009. 5

[8] Bin Ding, Chengjiang Long, Ling Zhang, and Chunxia Xiao. ARGAN: Attentive recurrent generative adversarial network for shadow detection and removal. In *ICCV*, pages 10213–10222, 2019. 1, 2

[9] Naiyu Gao, Yanhu Shan, Yupei Wang, Xin Zhao, Yinan Yu, Ming Yang, and Kaiqi Huang. SSAP: Single-shot instance segmentation with affinity pyramid. In *ICCV*, pages 642–651, 2019. 3

[10] Ross Girshick. Fast R-CNN. In *ICCV*, pages 1440–1448, 2015. 5

[11] Priya Goyal, Piotr Dollár, Ross Girshick, Pieter Noordhuis, Lukasz Wesolowski, Aapo Kyrola, Andrew Tulloch, Yangqing Jia, and Kaiming He. Accurate, large minibatch SGD: Training ImageNet in 1 hour. *arXiv preprint arXiv:1706.02677*, 2017. 5

[12] Ruiqi Guo, Qieyun Dai, and Derek Hoiem. Single-image shadow detection and removal using paired regions. In *CVPR*, pages 2033–2040, 2011. 2

[13] Zeeshan Hayder, Xuming He, and Mathieu Salzmann. Boundary-aware instance segmentation. In *CVPR*, pages 5696–5704, 2017. 3

[14] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask R-CNN. In *ICCV*, pages 2961–2969, 2017. 3, 4, 5

[15] Le Hou, Tomás F. Yago Vicente, Minh Hoai, and Dimitris Samaras. Large scale shadow annotation and detection using lazy annotation and stacked CNNs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019. to appear. 1, 3

[16] Xiaowei Hu, Chi-Wing Fu, Lei Zhu, Jing Qin, and Pheng-Ann Heng. Direction-aware spatial context features for shadow detection and removal. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019. to appear. 1, 2

[17] Xiaowei Hu, Yitong Jiang, Chi-Wing Fu, and Pheng-Ann Heng. Mask-ShadowGAN: Learning to remove shadows from unpaired data. In *ICCV*, pages 2472–2481, 2019. 1

[18] Xiaowei Hu, Tianyu Wang, Chi-Wing Fu, Yitong Jiang, Qiong Wang, and Pheng-Ann Heng. Revisiting shadow detection: A new benchmark dataset for complex world. *arXiv preprint arXiv:1911.06998*, 2019. 1

[19] Xiaowei Hu, Lei Zhu, Chi-Wing Fu, Jing Qin, and Pheng-Ann Heng. Direction-aware spatial context features for shadow detection. In *CVPR*, pages 7454–7462, 2018. 1, 2

[20] Xiang Huang, Gang Hua, Jack Tumblin, and Lance Williams. What characterizes a shadow boundary under the sun and sky? In *ICCV*, pages 898–905, 2011. 2

[21] Salman Hameed Khan, Mohammed Bennamoun, Ferdous Sohel, and Roberto Togneri. Automatic feature learning for robust shadow detection. In *CVPR*, pages 1939–1946, 2014. 1, 2

[22] Salman Hameed Khan, Mohammed Bennamoun, Ferdous Sohel, and Roberto Togneri. Automatic shadow detection and removal from a single image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(3):431–446, 2016. 1

[23] Alexander Kirillov, Evgeny Levinkov, Bjoern Andres, Bogdan Savchynskyy, and Carsten Rother. InstanceCut: from edges to instances with multicut. In *CVPR*, pages 5008–5017, 2017. 3

[24] Jean-François Lalonde, Alexei A. Efros, and Srinivasa G. Narasimhan. Detecting ground shadows in outdoor consumer photographs. In *ECCV*, pages 322–335, 2010. 2

[25] Hieu Le and Dimitris Samaras. Shadow removal via shadow image decomposition. In *ICCV*, pages 8578–8587, 2019. 1

[26] Hieu Le, Tomás F. Yago Vicente, Vu Nguyen, Minh Hoai, and Dimitris Samaras. A+D Net: Training a shadow detector with adversarial shadow attenuation. In *ECCV*, pages 662–678, 2018. 1, 2

[27] Yi Li, Haozhi Qi, Jifeng Dai, Xiangyang Ji, and Yichen Wei. Fully convolutional instance-aware semantic segmentation. In *CVPR*, pages 2359–2367, 2017. 3

[28] Tsung-Yi Lin, Piotr Dollar, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In *CVPR*, pages 2117–2125, 2017. 3, 5

[29] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft COCO: Common objects in context. In *ECCV*, pages 740–755, 2014. 3, 5

[30] Shu Liu, Jiaya Jia, Sanja Fidler, and Raquel Urtasun. SGN: Sequential grouping networks for instance segmentation. In *CVPR*, pages 3496–3504, 2017. 3

[31] Shu Liu, Lu Qi, Haifang Qin, Jianping Shi, and Jiaya Jia. Path aggregation network for instance segmentation. In *CVPR*, pages 8759–8768, 2018. 3

[32] Vu Nguyen, Tomás F. Yago Vicente, Maozheng Zhao, Minh Hoai, and Dimitris Samaras. Shadow detection with conditional generative adversarial networks. In *ICCV*, pages 4510–4518, 2017. 2

[33] Alexandros Panagopoulos, Chaohui Wang, Dimitris Samaras, and Nikos Paragios. Illumination estimation and cast shadow detection through a higher-order graphical model. In *CVPR*, pages 673–680, 2011. 2

[34] Chao Peng, Tete Xiao, Zeming Li, Yuning Jiang, Xiangyu Zhang, Kai Jia, Gang Yu, and Jian Sun. MegDet: A large mini-batch object detector. In *CVPR*, pages 6181–6189, 2018. 3

[35] Pedro O. Pinheiro, Ronan Collobert, and Piotr Dollár. Learning to segment object candidates. In *NeurIPS*, pages 1990–1998, 2015. 3

[36] Pedro O. Pinheiro, Tsung-Yi Lin, Ronan Collobert, and Piotr Dollár. Learning to refine object segments. In *ECCV*, pages 75–91, 2016. 3

[37] Liangqiong Qu, Jiandong Tian, Shengfeng He, Yandong Tang, and Rynson W.H. Lau. DeshadowNet: A multi-context embedding deep network for shadow removal. In *CVPR*, pages 4067–4075, 2017. 1

[38] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. In *NeurIPS*, pages 91–99, 2015. 4

[39] Elena Salvador, Andrea Cavallaro, and Touradj Ebrahimi. Cast shadow segmentation using invariant color features. *Computer Vision and Image Understanding*, 95(2):238–259, 2004. 2

[40] Li Shen, Teck Wee Chua, and Karianto Leman. Shadow optimization from structured deep edge detection. In *CVPR*, pages 2067–2074, 2015. 2

[41] Jiandong Tian, Xiaojun Qi, Liangqiong Qu, and Yandong Tang. New spectrum ratio properties and features for shadow detection. *Pattern Recognition*, 51:85–96, 2016. 2

[42] Ries Uittenbogaard, Clint Sebastian, Julien Vijverberg, Bas Boom, Dariu M. Gavrila, and Peter H.N. de With. Privacy protection in street-view panoramas using depth and multi-view imagery. In *CVPR*, pages 10581–10590, 2019. 1, 8

[43] Tomás F. Yago Vicente, Minh Hoai, and Dimitris Samaras. Leave-one-out kernel optimization for shadow detection. In *ICCV*, pages 3388–3396, 2015. 2

[44] Tomás F. Yago Vicente, Minh Hoai, and Dimitris Samaras. Noisy label recovery for shadow detection in unfamiliar domains. In *CVPR*, pages 3783–3792, 2016. 3

[45] Tomás F. Yago Vicente, Minh Hoai, and Dimitris Samaras. Leave-one-out kernel optimization for shadow detection and removal. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(3):682–695, 2018. 2

[46] Tomás F. Yago Vicente, Le Hou, Chen-Ping Yu, Minh Hoai, and Dimitris Samaras. Large-scale training of shadow detectors with noisily-annotated shadow examples. In *ECCV*, pages 816–832, 2016. 1, 2, 3

[47] Jifeng Wang, Xiang Li, and Jian Yang. Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal. In *CVPR*, pages 1788–1797, 2018. 1, 2, 3

[48] Yuxin Wu, Alexander Kirillov, Francisco Massa, Wan-Yen Lo, and Ross Girshick. Detectron2. https://github.com/facebookresearch/detectron2, 2019. 5

[49] Saining Xie, Ross Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He. Aggregated residual transformations for deep neural networks. In *CVPR*, pages 1492–1500, 2017. 5

[50] Quanlong Zheng, Xiaotian Qiao, Ying Cao, and Rynson W.H. Lau. Distraction-aware shadow detection. In *CVPR*, pages 5167–5176, 2019. 1, 2

[51] Bolei Zhou, Hang Zhao, Xavier Puig, Sanja Fidler, Adela Barriuso, and Antonio Torralba. Scene parsing through ADE20K dataset. In *CVPR*, pages 633–641, 2017. 3

[52] Bolei Zhou, Hang Zhao, Xavier Puig, Tete Xiao, Sanja Fidler, Adela Barriuso, and Antonio Torralba. Semantic understanding of scenes through the ADE20K dataset. *International Journal of Computer Vision*, 127(3):302–321, 2019. 3

[53] Jiejie Zhu, Kegan G. G. Samuel, Syed Z. Masood, and Marshall F. Tappen. Learning to recognize shadows in monochromatic natural images. In *CVPR*, pages 223–230, 2010. 2

[54] Lei Zhu, Zijun Deng, Xiaowei Hu, Chi-Wing Fu, Xuemiao Xu, Jing Qin, and Pheng-Ann Heng. Bidirectional feature pyramid network with recurrent attention residual modules for shadow detection. In *ECCV*, pages 121–136, 2018. 1, 2