Supplementary Material for CANet: A Context-Aware Network for Shadow Removal

Zipei Chen[1], Chengjiang Long[2*], Ling Zhang[3], Chunxia Xiao[1*†]
[1]School of Computer Science, Wuhan University, Wuhan, Hubei, China
[2]JD Finance America Corporation, Mountain View, CA, USA
[3]Wuhan University of Science and Technology, Wuhan, Hubei, China
czpp19@whu.edu.cn, cjfykx@gmail.com, zhling@wust.edu.cn, cxxiao@whu.edu.cn

Abstract

In this supplementary material, we first elaborate the details about our network architecture in Section 1. Then, we provide additional explanation and more comparison results about our dual-head contextual patch matching (CPM) module in Section 2. Finally, more visual shadow removal comparison results are given in Section 3. Note that we did not include all the material in the main paper due to the space limit.

1. Details about Network Architectures

CPM module. The architecture of our CPM module is shown in Table 1, which divides into three parts: feature extractor, pair type classifier, and correlation degree regressor. The feature extractor is used to extract 256-dimensional feature representations for input patch pairs. It is designed with 4 convolutional layers, 3 residual blocks, and a bottleneck layer. The obtained representations are feed into the classifier and regressor separately. Both the classification head and the regression head contain 3 fully connected layers, and classification head also have a softmax layer.

CANet. In Figure 1, we illustrate the detailed network architecture of our proposed CANet. Each orange rectangles in the network is the feature map of the corresponding layer, and the number in the rectangles is their channel number. Note that the "DenseBlock", "Transition layer" are followed as the original version of DenseNet [5].

*This work was co-supervised by Chengjiang Long and Chunxia Xiao.
†Corresponding author.

Table 1. The architecture of our CPM module. It contains a feature extractor, a pair type classifier, and a correlation degree regressor.

	Layer	Output Size	Operation
Feature extractor	Conv1	$16 \times 16 \times 64$	Conv(3×3 stride 2)
	Res1	$16 \times 16 \times 64$	Res-blocks(3×3)
	Conv2	$8 \times 8 \times 96$	Conv(3×3 stride 2)
	Res2	$8 \times 8 \times 96$	Res-blocks(3×3)
	Conv3	$4 \times 4 \times 96$	Conv(3×3 stride 2)
	Res3	$4 \times 4 \times 96$	Res-blocks(3×3)
	Conv4	$4 \times 4 \times 64$	Conv(3×3 stride 1)
	Bottleneck	256	FC
Classifier	FC1	256	FC
	FC2	128	FC
	FC3	3	FC
	Softmax	3	Softmax
Regressor	FC1	256	FC
	FC2	128	FC
	FC3	1	FC

2. Better Understanding for CPM Module and Comparison Results

2.1. Effectiveness of Light-unaware Images

As shown in Figure 2, with the supplementary light-unaware image, we can largely eliminate the difference between shadow regions and non-shadow regions, which effectively avoids the matching errors caused by shadows, as illustrated in Figure 6(a).

Also, from Figure 3, we can observe that there is a larger difference between shadow image and light-unaware image in shadow regions while a smaller difference in the non-shadow region. It suggests that the shadow image and light-unaware image pair can provide an indication to distinguish shadow patches from non-shadow patches, which can be used to perform our pair type classifier.

2.2. Large-scale Training Dataset for CPM

To train our CPM module, we collect a large-scale training collection from the existing shadow benchmark

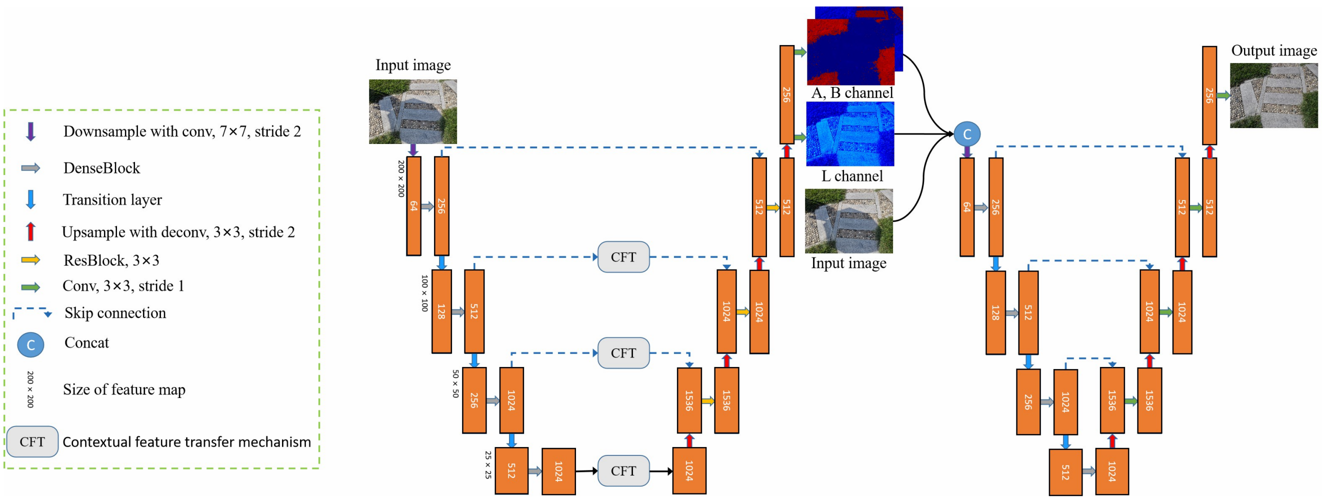


Figure 1. The network architecture of our CANet.

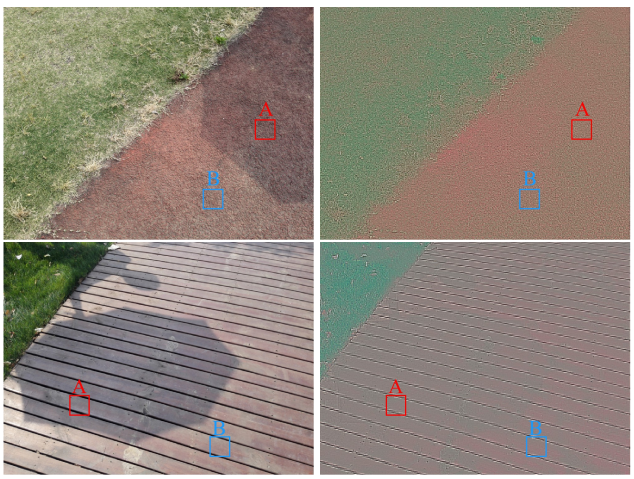
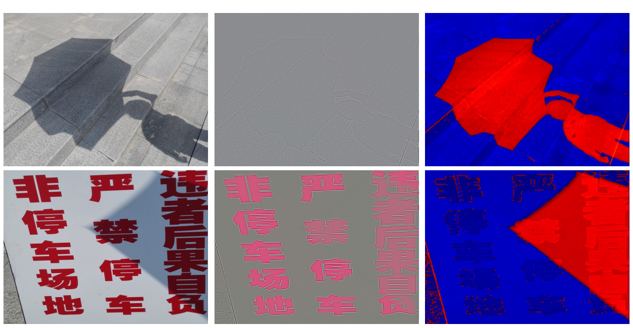


Figure 2. From left to right are: input shadow image; and the proposed light-unaware image, which can eliminate the difference between region A and B caused by shadow.



(a) (b) (c)

Figure 3. The illustration of the difference between shadow image and light-unaware image, from left to right are: (a)input shadow image; (b)proposed light-unaware image and (c)the difference between them.

Figure 4. The illustration of our first way to collect matched patch pairs.

datasets: ISTD [8] and SRD [7]. The collected training dataset contains more than 360,000 and 600,000 patch pairs separately (50% match pairs and 50% non-match pairs). These patch pairs are collected from two ways: (1) we select a shadow patch in the shadow image and a matched non-shadow patch in the shadow-free image, which has the same position as the shadow patch, as illustrated in Figure 4; (2) we select a shadow patch from shadow regions and find another matched patch from non-shadow regions in the shadow image. Specially, we randomly select two patches from shadow and non-shadow regions in the shadow image and calculate the cosine similarity between the two patches in the corresponding shadow-free image. We choose the pairs with cosine similarity higher than 0.95 as the matching pair and less than 0.6 as the non-match pair, as shown in Figure 4. Due to the lack of shadow mask ground-truth in SRD dataset [5], we firstly use the results of the latest shadow detection method DSD [11], and then manually choose the correct results as the approximate ground-truth during the process of dataset collecting.

2.3. Results of Our Dual-head CPM Module

To further explain the superiority of our CPM module, we give more comparison results of our CPM module with the traditional-match method like SIFT [6] and MatchNet [3] by asking two questions: (1) *Does the matched patch*

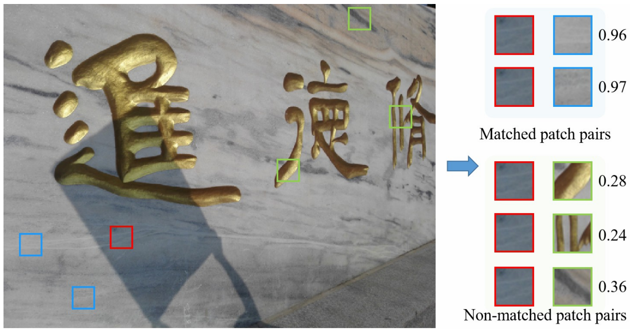
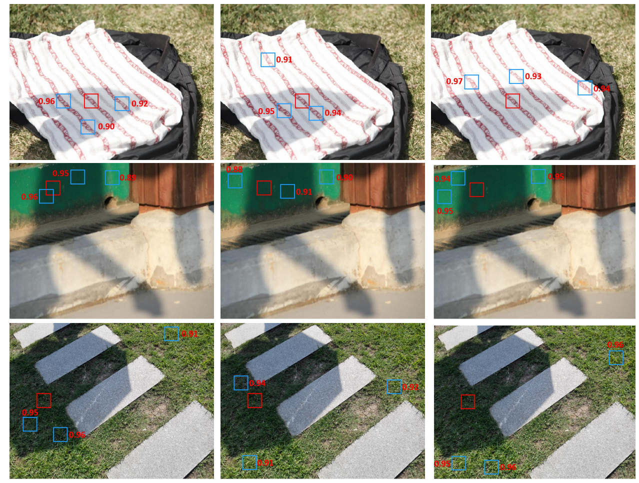


Figure 5. The illustration of our second way to collect matched and non-matched patch pairs.



(a) (b) (c)

Figure 6. Patch matching results. From left to right are: (a) Traditional-match; (b) MatchNet; (c) our Dual-head CPM module.

pair include a shadow patch and a non-shadow patch? and (2) *How about the accuracy of the similarity prediction?*

Does the matched patch pair include a shadow patch and a non-shadow patch? The CPM module is designed to transfer contextual information from non-shadow regions to shadow regions. To guarantee the accuracy of the predicted matched patch pair, we have to ensure that the matched patch pair contains a shadow patch and a non-shadow patch. Table 2 summarizes the proportion that the matched patch for a shadow patch is from non-shadow regions of different matching methods. Note that the larger value is better. We can clearly observe that our method can find the correct matched patch type for shadow patches, abandoning the undesirable matched patch from the table.

Table 2. The proportion that the matched patch of a shadow patch is from non-shadow regions on ISTD [8] and SRD [7] dataset.

Method	ISTD	SRD
Traditional-match	46.82%	37.21%
MatchNet [3]	63.75%	56.44%
Dual-head CPM	**92.68%**	**90.35%**

How about the accuracy of the similarity prediction? We use the Mean Square Error (MSE) between the predicted correlation score and the ground truth as the metric to evaluate our correlation regressor's accuracy. Note that the smaller value is better. Table 3 reports the quantitative evaluation results, where we can see that our dual-head CPM module outperforms the other methods. We also provide some qualitative results in Figure 6, which evidently demonstrates the superiority of our CPM module.

Table 3. The quantitative comparison results of correlation score prediction on ISTD [8] and SRD [7] dataset in terms of MSE.

Method	ISTD	SRD
Traditional-match	1.42	1.55
MatchNet [3]	0.68	0.81
Dual-head CPM	**0.32**	**0.37**

3. More Visual Shadow Removal Results

In this section, we provide more visual shadow removal comparison results in Figure 7. Here, we compare our CANet with six state-of-the-art methods, *i.e.*, Guo [2], Zhang [10], ST-CGAN [8], ARGAN [1], DSC [4] and RIS-GAN [9].

References

[1] Bin Ding, Chengjiang Long, Ling Zhang, and Chunxia Xiao. Argan: Attentive recurrent generative adversarial network for shadow detection and removal. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, October 2019. 3

[2] Ruiqi Guo, Qieyun Dai, and Derek Hoiem. Single-image shadow detection and removal using paired regions. In *IEEE Conference on Computer Vision Pattern Recognition*, pages 2033–2040. IEEE, 2011. 3

[3] Xufeng Han, Thomas Leung, Yangqing Jia, Rahul Sukthankar, and Alexander. C. Berg. Matchnet: Unifying feature and metric learning for patch-based matching. In *IEEE Conference on Computer Vision Pattern Recognition*, 2015. 2, 3

[4] 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. 3

[5] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *IEEE Conference on Computer Vision Pattern Recognition*, 2017. 1, 2

[6] David G Lowe. Object recognition from local scale-invariant features. In *Proceedings of the seventh IEEE international conference on computer vision*, volume 2, pages 1150–1157. Ieee, 1999. 2

Figure 7. Shadow removal results. From left to right are: (a) input images; shadow removal results of (b) Guo, (c) Zhang, (d) ST-CGAN, (e) DSC, (f) ARGAN, (g) RIS-GAN, (h) our CANet; and (i) the corresponding shadow-free ground truth images.

[7] Liangqiong Qu, Jiandong Tian, Shengfeng He, Yandong Tang, and Rynson W. H. Lau. Deshadownet: A multi-context embedding deep network for shadow removal. In *IEEE Conference on Computer Vision Pattern Recognition*, July 2017. 2, 3

[8] Jifeng Wang, Xiang Li, and Jian Yang. Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal. In *IEEE Conference on Computer Vision Pattern Recognition*, June 2018. 2, 3

[9] Ling Zhang, Chengjiang Long, Xiaolong Zhang, and Chunxia Xiao. Ris-gan: Explore residual and illumination with generative adversarial networks for shadow removal. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(7):12829–12836, 2020. 3

[10] Ling Zhang, Qing Zhang, and Chunxia Xiao. Shadow remover: Image shadow removal based on illumination recovering optimization. *IEEE Transactions on Image Processing*, 24(11):4623–4636, 2015. 3

[11] Quanlong Zheng, Xiaotian Qiao, Ying Cao, and Rynson W.H. Lau. Distraction-aware shadow detection. In *IEEE Conference on Computer Vision Pattern Recognition*, June 2019. 2