DeshadowNet: A Multi-context Embedding Deep Network for Shadow Removal

Liangqiong Qu[1,2,3], Jiandong Tian[1], Shengfeng He[4], Yandong Tang[1], Rynson W. H. Lau[2]

[1] State Key Laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Sciences; [2] University of Chinese Academy of Sciences; [3] City University of Hong Kong; [4] South China University of Technology

Prior Works and Motivation :

- **Lack of a fully-automatic and end-to-end pipeline**
  Existing works require the prior information of shadow location, but shadow detection itself is a challenging task.
- **Neglect high level semantic information**
- **Require specific operation for penumbra regions**

Our Solution:

- A shadow-free image I_{ns} can be considered as a pixel-wise product of a shadow image I_s and a shadow matte S_m

(a) Shadow image I_s (b) Shadow Matte S_m (c) Shadow-free image I_{ns}

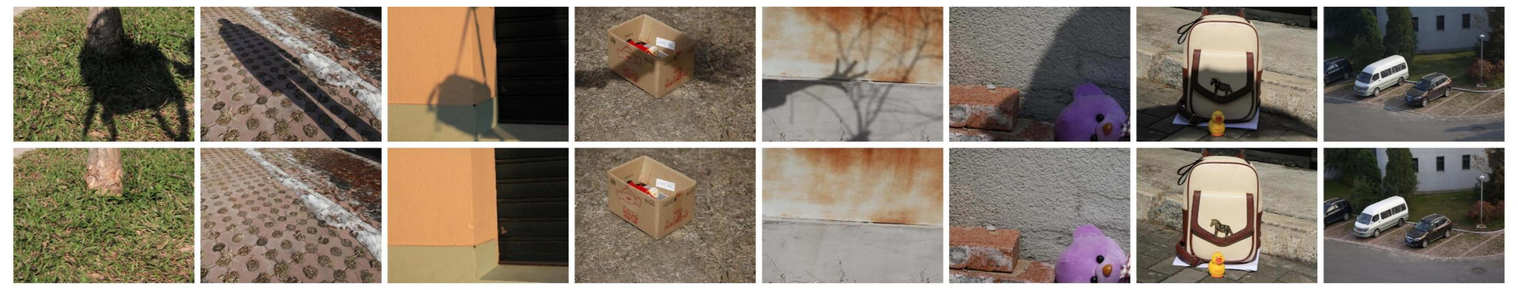
- **Directly estimating a shadow matte** to remove shadows
- We propose a **multi-context embedding** deep network (**DeshadowNet**) to learn the mapping function between the shadow image and its shadow matte as: $S_m = F(I_s, \Theta)$, where Θ represents the learned parameters of DeshadowNet.

Contributions:

- **Design an end-to-end and fully automatic framework**
  Unify detect shadows, classify umbra/penumbra regions, and remove shadows into one step
- **A proposed multi-context embedding network**
  Understand image content from a global perspective and model the precise illumination compensation with local image details
- **Provide a new large scale shadow removal data set (SRD)**

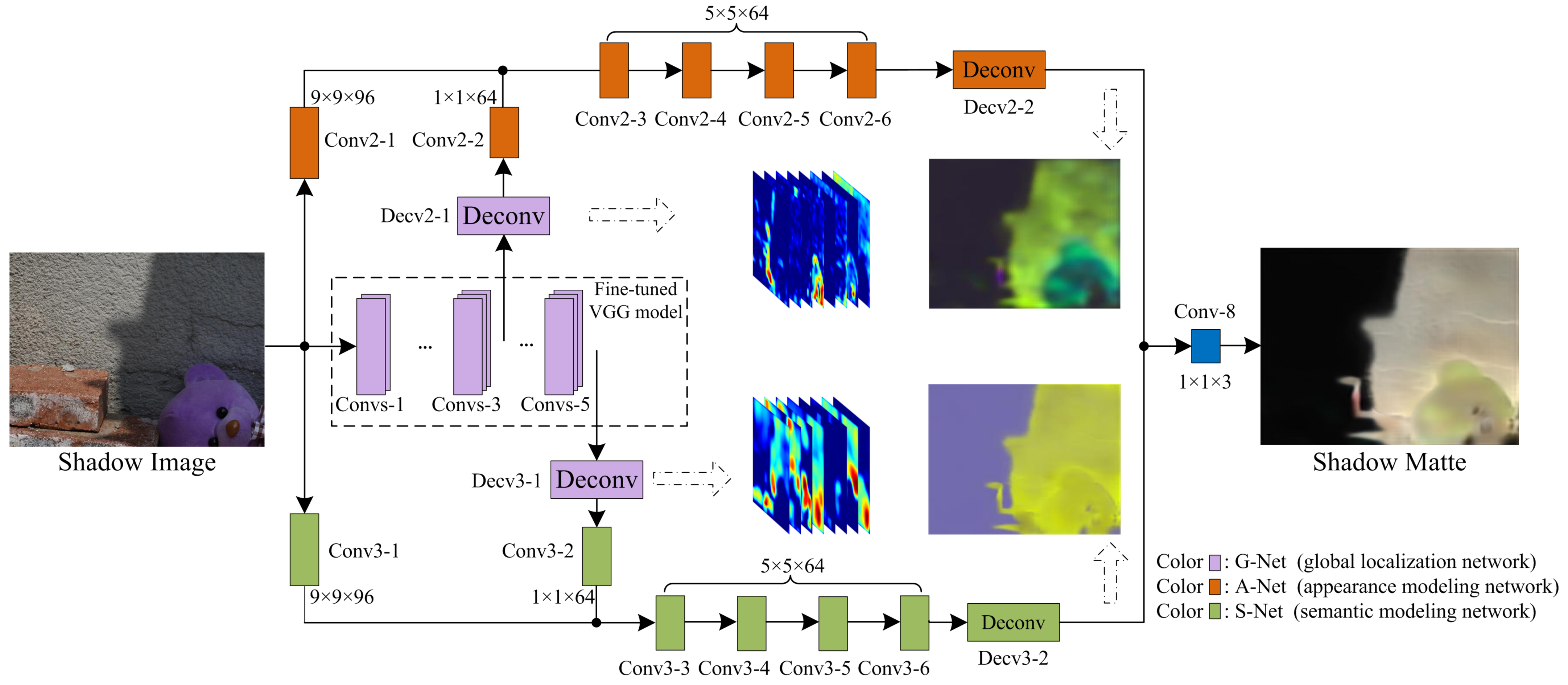
Shadow images

GT



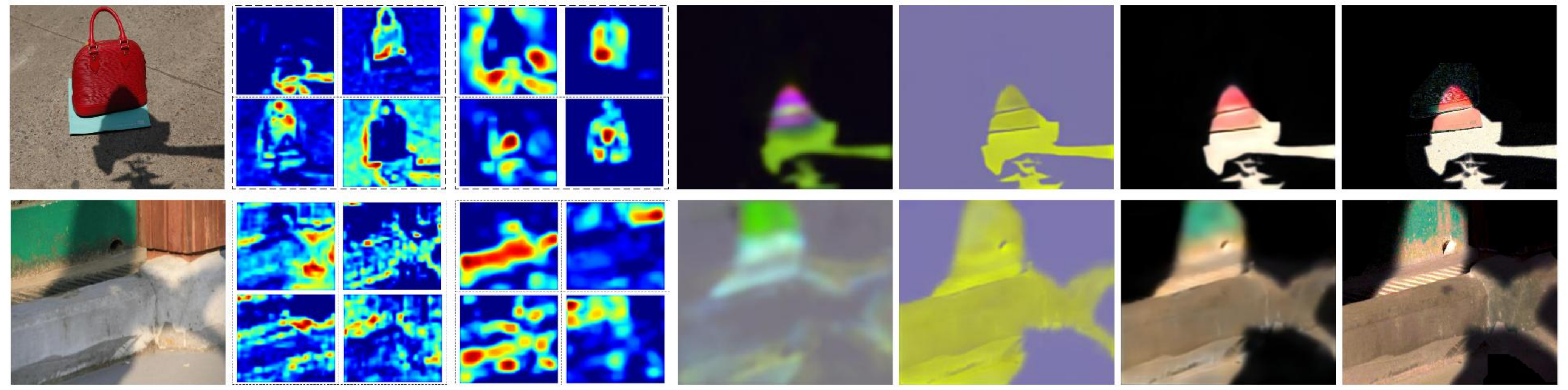
Architecture of DeshadowNet :

- Contain three cooperative sub-networks: global localization network (G-Net), appearance modeling network (A-Net), semantic modeling network (S-Net)



- **G-Net**: describe the global structure and high-level semantic information
- **A-Net**: acquire the appearance information from shallower layer of G-Net, combined with local image detail, to predict shadow matte in coarse scale and help model the appearance of shadow matte
- **S-Net**: extract the semantic information from deeper layer of G-Net, combined with local image detail, to predict shadow matte in fine scale and help encode semantic information of shadow matte

- **Visualization of the intermediate results of DeshadowNet**



From left to right: shadow images, example feature maps of the shallower and deeper layer of G-Net, output of A-Net, output of S-Net, output of DeshadowNet, and the ground-truth shadow mattes

Quantitative Comparison with RMSE Error

Dataset	Different regions	Original	Guo *et al.* [15]	Yang *et al.* [33]	Gong *et al.* [12]	Gryka *et al.*[14]	Khan *et al.* [19]	Ours
UIUC [15]	Shadow	42	13.9	21.6	11.8	13.9	12.1	**9.6**
	Non-shadow	**4.6**	5.4	20.3	4.9	7.6	5.1	4.8
	All	13.7	7.4	20.6	6.6	9.1	6.8	**5.9**
LRSS [14]	Shadow	44.45	31.58	23.35	22.27	-	-	**14.21**
	Non-shadow	**4.1**	4.87	19.35	4.39	-	-	4.17
	All	17.73	13.89	20.70	10.43	-	-	**7.56**
SRD	Shadow	42.38	29.89	23.43	19.58	-	-	**11.78**
	Non-shadow	**4.56**	6.47	22.26	4.92	-	-	4.84
	All	14.41	12.60	22.57	8.73	-	-	**6.64**

Qualitative Results (the numbers in images are RMSE error)

Images GT Guo [15] Yang[33] Gong[12] Ours

References

[12] H. Gong, et al. Interactive shadow removal and ground truth for variable scene categories. BMVC, 2014
[14] M. Gryka, et al. Learning to remove soft shadows. ACM TOG, 2015.
[15] R. Guo, et al. Paired regions for shadow detection and removal. PAMI, 2012.
[19] S. H. Khan, et al. Automatic shadow detection and removal from a single image. PAMI, 2016.
[33] Q. Yang, et al. Shadow removal using bilateral filtering. TIP, 2012.