

Bidirectional Feature Pyramid Network with Recurrent Attention Residual Modules for Shadow Detection

Lei Zhu^{1,2,5}, Zijun Deng³, Xiaowei Hu¹, Chi-Wing Fu^{1,5},
Xuemiao Xu⁴, Jing Qin², and Pheng-Ann Heng^{1,5}

¹ The Chinese University of Hong Kong, Hong Kong, China

² The Hong Kong Polytechnic University, Hong Kong, China

³ South China University of Technology, Guangzhou, China

⁴ Guangdong Provincial Key Lab of Computational Intelligence and Cyberspace Information (South China University of Technology), Guangzhou, China

⁵ Shenzhen Key Laboratory of Virtual Reality and Human Interaction Technology, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China

Abstract. This paper presents a network to detect shadows by exploring and combining global context in deep layers and local context in shallow layers of a deep convolutional neural network (CNN). There are two technical contributions in our network design. First, we formulate the recurrent attention residual (RAR) module to combine the contexts in two adjacent CNN layers and learn an attention map to select a residual and then refine the context features. Second, we develop a bidirectional feature pyramid network (BFPN) to aggregate shadow contexts spanned across different CNN layers by deploying two series of RAR modules in the network to iteratively combine and refine context features: one series to refine context features from deep to shallow layers, and another series from shallow to deep layers. Hence, we can better suppress false detections and enhance shadow details at the same time. We evaluate our network on two common shadow detection benchmark datasets: S-BU and UCF. Experimental results show that our network outperforms the best existing method with 34.88% reduction on SBU and 34.57% reduction on UCF for the balance error rate.

1 Introduction

Shadows are regions that receive less illumination than the surroundings, due to lights occluded by associated objects in the scene. To detect shadows in images, early works develop physical models with heuristic priors [1, 2] or take a machine learning approach based on hand-crafted features. However, image priors and hand-crafted features are not effective for extracting high-level semantics.

More recently, methods based on the convolutional neural network (CNN) [3–7] show distinct performance on various shadow detection benchmarks, e.g., [4,

Joint first authors

Corresponding author (cwfu@cse.cuhk.edu.hk)

