Deep Image Harmonization by Bridging the Reality Gap

Junyan Cao, Wenyan Cong, Li Niu, Jianfu Zhang, Liqing Zhang

Shanghai Jiao Tong University

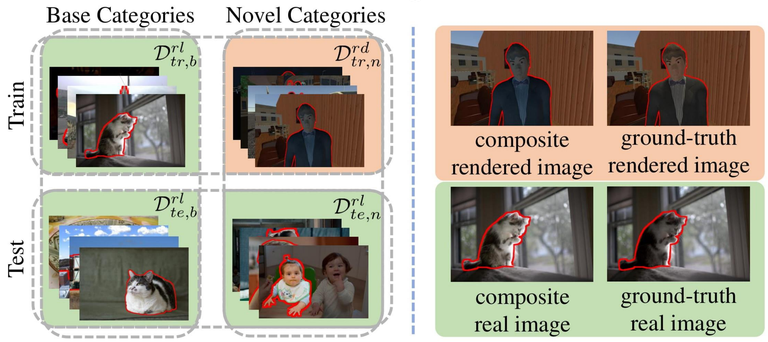
BMVC 2022

Dataset and code are available at:
https://github.com/bcmi/Rendered-Image-Harmonization-Dataset-RdHarmony

Introduction

Image harmonization aims to diminish the discrepancy by adjusting the appearance of the composite foreground according to the background.

- **Datasets**: We split the categories of the composite images into two disjoint sets, *i.e.*, novel categories with only a few or no training pairs, and base categories with abundant training pairs. Our RdHarmony contains rendered training pairs from 11 novel categories, which is produced by 3D rendering techniques, aiming to enrich the real dataset of novel categories.



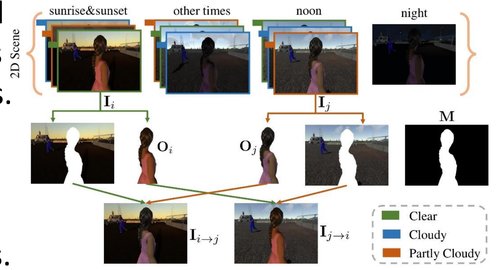
- **CharmNet design**: The main challenge in our problem is to bridge the gap between rendered image domain and real image domain. Our CharmNet is a cross-domain harmonization network, applying to align two domains when training with a combination of real images and rendered images.

Conclusion

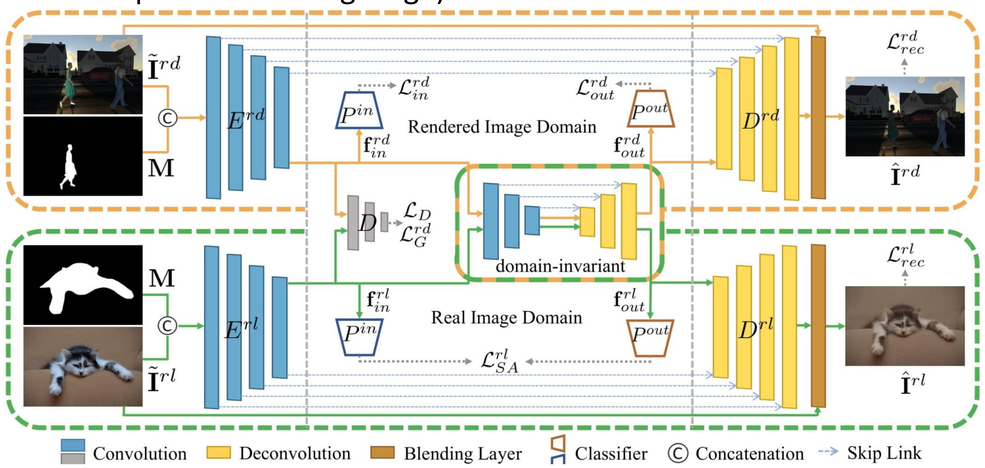
- We investigate on the cross-domain and cross-category issues in image harmonization.
- We have contributed and released the first large-scale rendered image harmonization dataset RdHarmony.
- We propose the first cross-domain image harmonization network CharmNet with novel network architecture and style aggregation loss.

Dataset Construction & Method

We combine different weathers and time-of-the-day of rendered images then define 10 representative styles. We exchange the foregrounds of ground-truth rendered images in the same 2D scene with each other, producing rendered training triplets.



Our CharmNet consists of three stages (i.e., the domain specific encoding stage, the domain-invariant encoding-decoding stage and the domain specific decoding stage).



- **Domain specific encoding stage**: Two specific encoders project images from real image domain and rendered image domain into the same feature space. A domain discriminator is used to further align two domains.
- **Domain-invariant encoding-decoding stage**: Through the domain-invariant encoder-decoder, the knowledge of harmonization is transferred from rendered image domain to real image domain.
- **Domain specific decoding stage**: Two specific decoders project the domain-invariant features back to the input domain, to generate the harmonized output in each domain.

- **Style aggregation (SA) loss**: We design the SA loss to facilitate cross domain knowledge transfer, aiming to enforce the style distribution to be more concentrated after harmonization.

Results

We focus on human harmonization in this work.

Quantitative Comparison

Cross-category and cross-domain issues.

#	1	2	3	4
Training data	-	$\mathcal{D}^{rl}_{tr,n}$	$\mathcal{D}^{rl}_{tr,b}(sub)$	$\mathcal{D}^{rd}_{tr,n}(sub)$
fMSE↓	931.74	386.36	486.54	1100.86
PSNR↑	33.29	35.71	34.64	31.25

Col. 3 *v.s.* col. 2: performance drop on cross-category harmonization.
Col. 4 *v.s.* col. 2: performance drop on cross-domain harmonization.

Results of models trained on various training data.

#	Training data	Method	$\mathcal{D}^{rl}_{te,n}$ MSE↓	$\mathcal{D}^{rl}_{te,n}$ fMSE↓	$\mathcal{D}^{rl}_{te,n}$ PSNR↑	$\mathcal{D}^{rl}_{te,b}$ MSE↓	$\mathcal{D}^{rl}_{te,b}$ fMSE↓	$\mathcal{D}^{rl}_{te,b}$ PSNR↑
1	-	Input Composite	155.74	931.74	33.29	177.34	1505.92	31.15
2	$\mathcal{D}^{rl}_{tr,b}$	DoveNet [4]	69.06	458.15	35.01	65.87	674.57	33.94
3		Hao *et al.* [17]	52.42	407.26	35.43	58.21	582.79	34.54
4		iDIH [38]	46.67	417.71	35.21	53.03	566.96	34.77
5		RainNet [28]	58.40	525.36	34.97	59.54	701.07	34.44
6		Guo *et al.* [15]	49.92	416.21	35.58	52.31	577.16	35.01
7	$\mathcal{D}^{rd}_{tr,n}$ & $\mathcal{D}^{rl}_{tr,b}$	two-stage training	43.06	383.39	35.67	49.56	533.10	34.98
8		dataset fusion	38.31	368.50	35.64	44.15	485.94	35.27
9		UNIT [29]	55.09	458.72	34.74	56.21	607.14	34.38
10		CycleGAN [54]	51.82	474.78	34.64	56.30	592.33	34.52
11		CUT [34]	48.59	427.25	35.14	52.81	572.53	34.72
12		CharmNet	**30.83**	**296.40**	**36.60**	**39.41**	**432.19**	**35.83**
13	$\mathcal{D}^{rl}_{tr,n}$ & $\mathcal{D}^{rl}_{tr,b}$	iDIH [38]	27.71	259.17	37.12	36.17	422.98	36.04

Qualitative Comparison

Example results on novel category

Example results on base categories