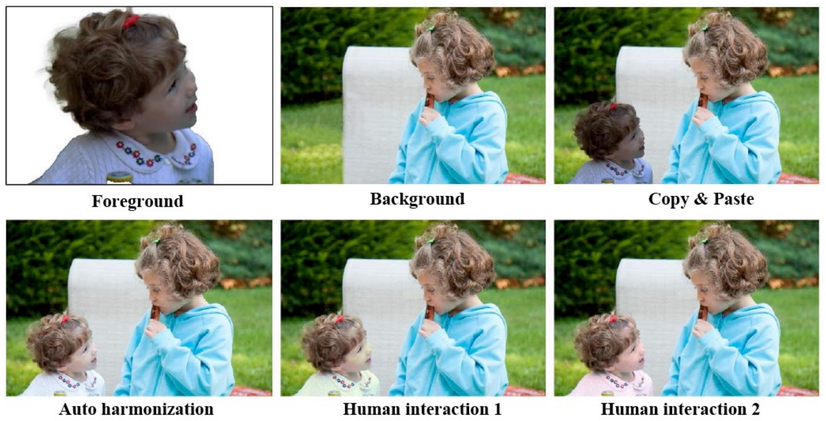
DCCF: Deep Comprehensible Color Filter Learning Framework for High-Resolution Image Harmonization

Ben Xue[1], Shenghui Ran[2], Quan Chen[2], Rongfei Jia[2], Binqiang Zhao[2], Xing Tang[2]
[1]Peking University & [2]Alibaba Inc, Tao Technology

Paper, code are available:
https://github.com/rockeyben/DCCF

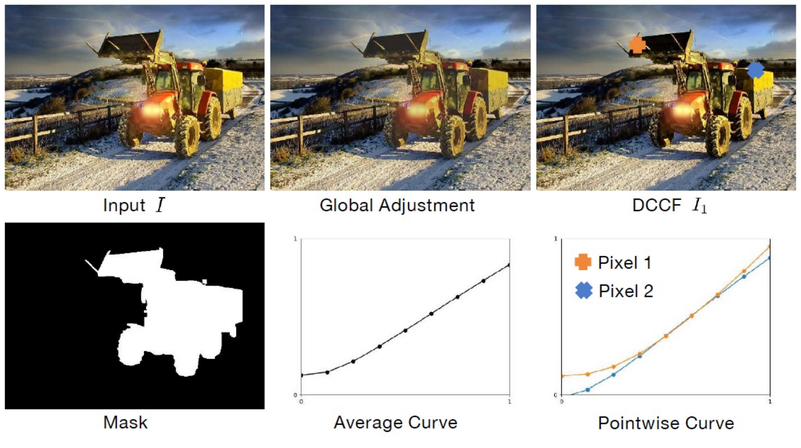
Comprehensible Image Processing



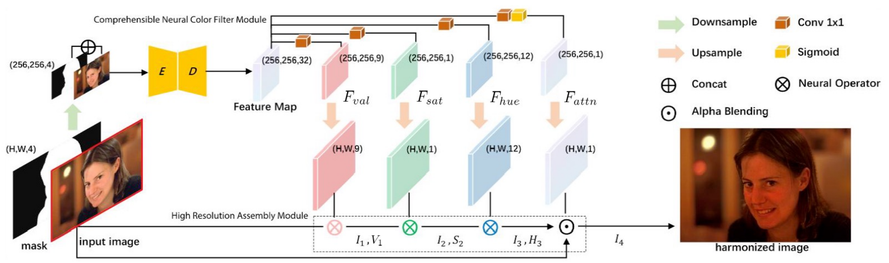
Our contributions:
- Make deep learning based color harmonization practical for real-world high-resolution images.
- Design four types of novel neural filter (i.e. hue, saturation, value and attentive rendering filters) that are comprehensible for human. Further interaction is possible.
- Achieve SOTA performance on iHarmony4 (Cong et al., CVPR'20) dataset at original resolution by 7.63% / 1.69% relative improvements on MSE / PSNR.

High-Resolution Processing

Learn pixel-level image filters at low-resolution, then upsample them to high-resolution to make subtle local adjustments.
(e.g. value filters: learnable curves)



DCCF: Deep Comprehensible Color Filter



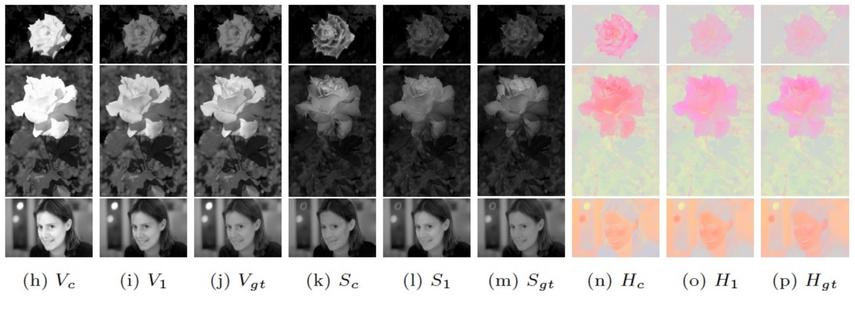
Value filters: $f_{val}(x;\phi,V_{min}) = V_{min} + \sum_{i=1}^{m} \phi_i * \max(x - \frac{i-1}{m}, 0)$

Saturation filter: $f_{sat}(x;\sigma) = x + (x - C_{med}) * clip(\sigma)$

Hue filter: $f_{col}(x;\boldsymbol{\Delta}) = \boldsymbol{R}x + t = \begin{bmatrix}\delta_{11} & \delta_{12} & \delta_{13}\\ \delta_{21} & \delta_{22} & \delta_{23}\\ \delta_{31} & \delta_{32} & \delta_{33}\end{bmatrix}\begin{bmatrix}x_R\\ x_G\\ x_B\end{bmatrix} + \begin{bmatrix}\delta_{14}\\ \delta_{24}\\ \delta_{34}\end{bmatrix}$

Attentive rendering filter: $I_4 = I * \alpha + W_{ref} * I_3 * (1 - \alpha)$

Intermediate Result Visualization



Quantitative Result

High-resolution performance comparison *State-of-the-art*

Method	Entire Dataset MSE ↓	Entire Dataset PSNR ↑	HCOCO MSE ↓	HCOCO PSNR ↑	HAdobe5k MSE ↓	HAdobe5k PSNR ↑	HFlickr MSE ↓	HFlickr PSNR ↑	Hday2night MSE ↓	Hday2night PSNR ↑
Input image	177.99	31.22	73.03	33.53	354.46	27.63	270.99	28.20	113.07	33.91
iDIH-HRNet [28]	-	-	19.96	38.25	-	-	93.50	32.42	71.01	35.77
iDIH-HRNet [28]+BU	43.56	34.98	34.40	35.45	37.82	35.47	104.69	30.91	50.87	37.41
iDIH-HRNet [28]+GF [14]	35.47	36.00	25.93	36.70	34.51	36.03	85.05	32.01	**49.90**	**37.67**
iDIH-HRNet [28]+BGU [2]	26.85	37.24	18.53	37.90	26.71	37.50	66.26	33.19	51.96	37.23
DCCF	**24.65**	**37.87**	**17.07**	**38.66**	**23.34**	**37.75**	**64.77**	**33.60**	55.76	37.40

High-resolution efficiency comparison *Good balance for complexity*

Method	1024 × 1024 T-C (ms) ↓	1024 × 1024 T-G (ms) ↓	1024 × 1024 Mem (MB) ↓	2048 × 2048 T-C (ms) ↓	2048 × 2048 T-G (ms) ↓	2048 × 2048 Mem (MB) ↓	3072 × 3072 T-C (ms) ↓	3072 × 3072 T-G (ms) ↓	3072 × 3072 Mem (MB) ↓
iDIH-HRNet [28]	420	231	1641	41040	907	4233	139768	2042	8551
iDIH-HRNet [28] + GF [14]	642	80.2	983	2001	160	1513	10181	391	2483
iDIH-HRNet [28] + BGU [2]	9932	-	2893	20803	-	4042	29836	-	8173
DCCF-iDIH-HRNet	762	104	1259	3289	286	2607	6517	545	4845

T-C: time on CPU; T-G: time on GPU

Qualitative Result

DCCF learns harmonized global appearance while maintaining high-resolution details.

Human Interaction

θ: hue angle of user's color intention
α: blending ratio for human and network's decision