A Generic Deep Architecture for Single Image Reflection Removal and Image Smoothing

Qingnan Fan[1], Jiaolong Yang[2], Gang Hua[2], Baoquan Chen[1,3], David Wipf[2]
[1]Shandong University [2]Microsoft Research [3]Shenzhen Research Institute, Shandong University

Codes and model: https://github.com/fqnchina/CEILNet

Challenges

- Single image reflection removal:
  - Underdetermined, weighted combination of scenes from the two sides of glass window.
    $I = w \bullet L_b + (1 - w) \bullet L_r$
  - Lacking plenty of reflection and clear image pairs for training the deep neural network.
  - Naively mixing two natural images by scaling different layers with weights summed to 1 does not work!
- Image smoothing:
  - Slow running time of traditional image smoothers.
  - Approximation of existing edge-aware filters with deep networks: Unsatisfactory quality (PSNR < 35).

Observation

- Edge information plays a very important role in many low-level vision tasks, such as
  - layer separation (reflection removal)

  - Image filtering (image smoothing)

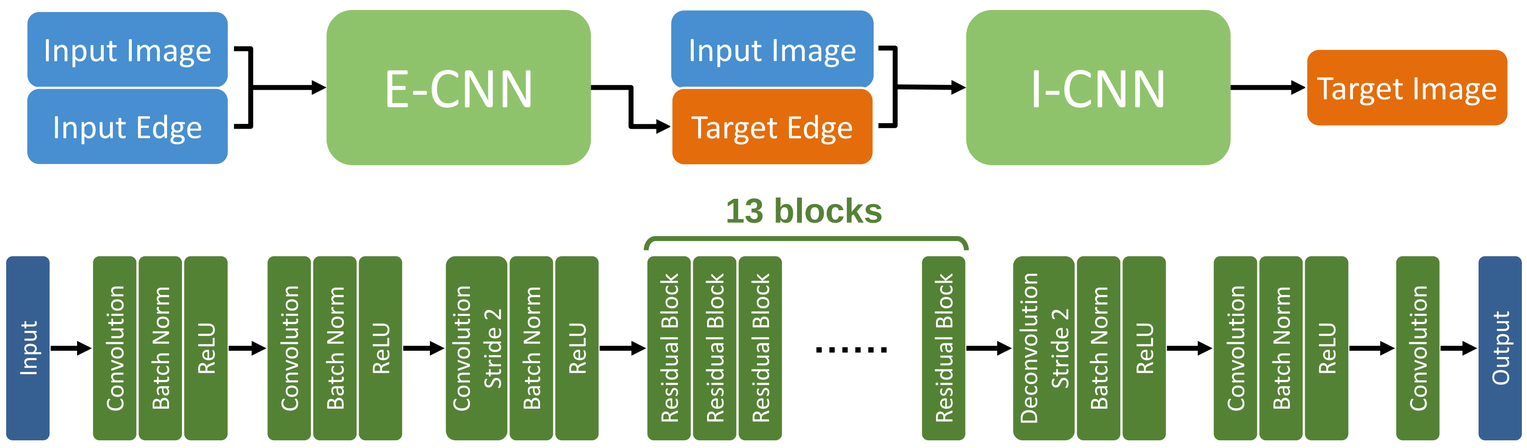
Contribution1: Cascaded Edge and Image Learning Network (CEILNet)

Instead of predicting images directly, we separate the end-to-end FCN into two sub networks:

- Target edge prediction
- Target image reconstruction

Both tasks are much more easier and learned with the similar CNN structure, 32-layer FCN.

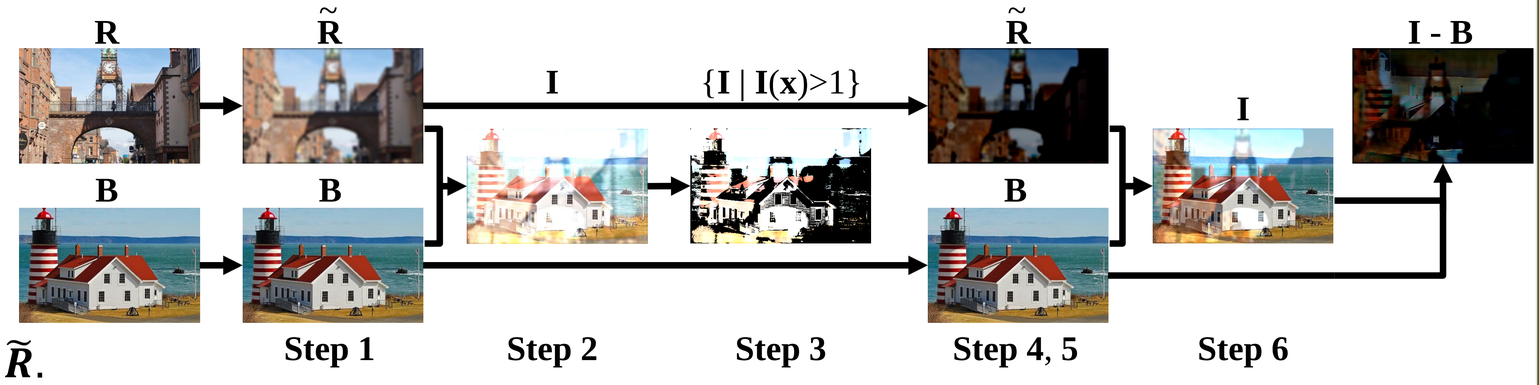
✓ Edge map represents the color difference between the 4 adjacent pixels.



Contribution2: Reflection Image Synthesis Pipeline

1. $\widetilde{\boldsymbol{R}} \leftarrow gauss\_blur_{\sigma}(\boldsymbol{R})$ with $\sigma \sim \mathcal{U}(2,5)$
2. $\boldsymbol{I} \leftarrow \widetilde{\boldsymbol{R}} + \boldsymbol{B}$
3. $m \leftarrow mean(\{\boldsymbol{I}(x,c) | \boldsymbol{I}(x,c) > 1, \forall x, \forall c = 1,2,3\})$
4. $\widetilde{\boldsymbol{R}}(x,c) \leftarrow \widetilde{\boldsymbol{R}}(x,c) - \gamma \cdot (\mathrm{m} - 1), \forall \mathrm{x}, \forall \mathrm{c}; \gamma\ set\ as\ 1.3$
5. $\widetilde{\boldsymbol{R}} \leftarrow Clip_{[0,1]}(\widetilde{\boldsymbol{R}})$
6. $\boldsymbol{I} \leftarrow Clip_{[0,1]}(\boldsymbol{B} + \widetilde{\boldsymbol{R}})$

Note m is adaptively-computed, and subtracted by $\widetilde{\boldsymbol{R}}$.

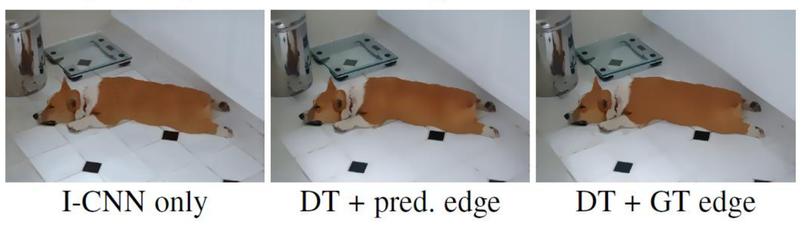


Ablation Study of Deep Network

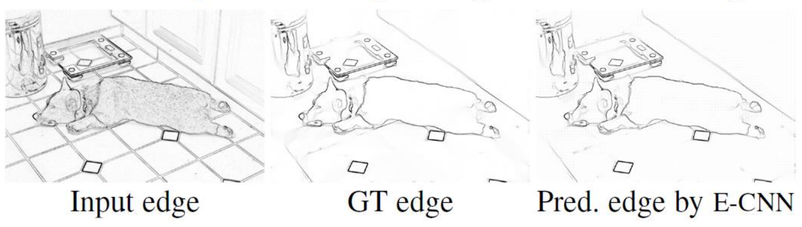
Demonstration of necessity of I-CNN by replacing it with traditional method Domain Transform (DT) , and importance of E-CNN by using I-CNN only.

	MSE	PSNR	SSIM
DT + input image edge	124.41	27.38	0.806
DT + pred. edge by E-CNN	51.26	31.17	0.964
DT + GT edge	45.67	31.66	0.971
I-CNN only	37.79	32.58	0.969
I-CNN only (64 layers)	31.86	33.33	0.973
I-CNN with input edge (64 layers)	22.50	34.86	0.979
CEILNet	**13.34**	**37.10**	**0.989**

Input image | GT image | CEILNet



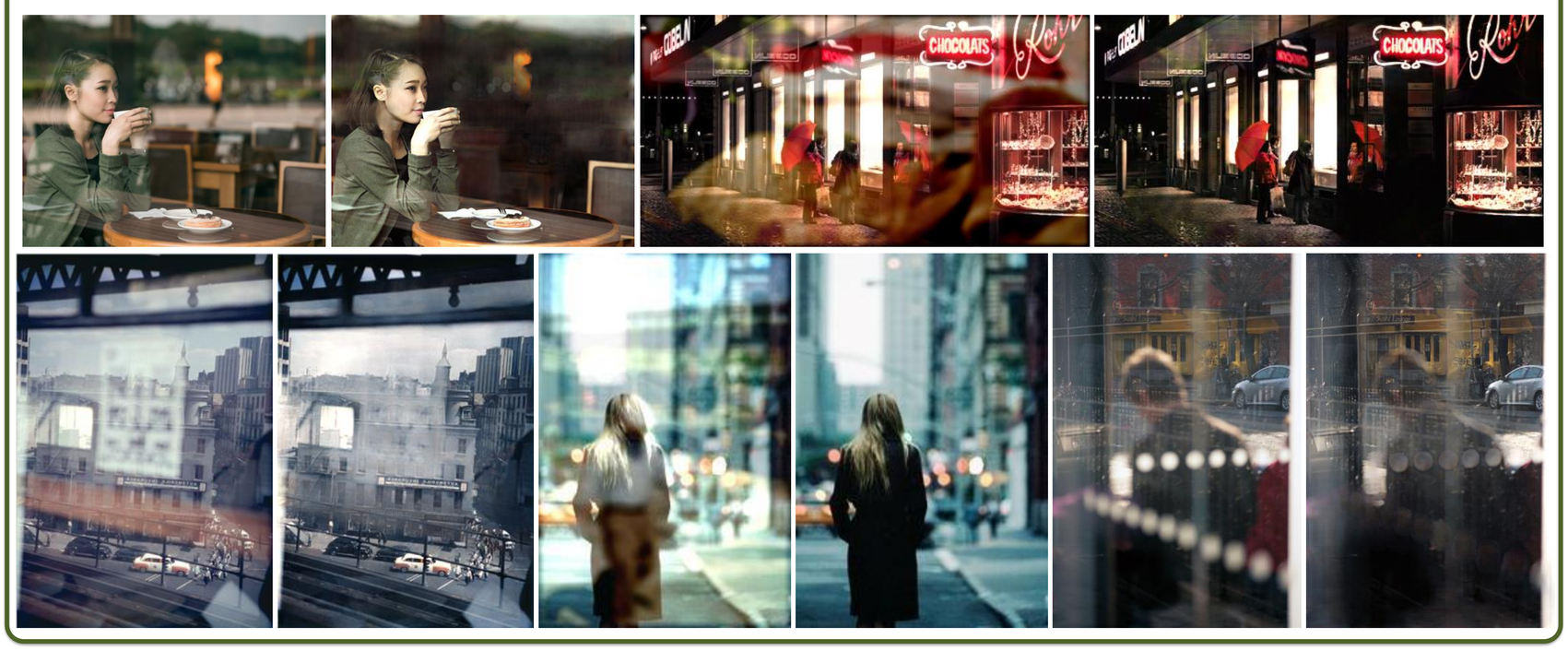
I-CNN only | DT + pred. edge | DT + GT edge



Input edge | GT edge | Pred. edge by E-CNN

Performance Evaluation of Reflection Removal Task

- Guess which is real and which is synthetic?



Performance Evaluation of Image Smoothing Task

Quality		BLF	IBLF	L_0	RGF	RTV	WLS	WMF	L_1	Ave.
PSNR	Xu15	35.02	32.97	31.66	32.49	35.68	33.92	29.62		32.62
	Ours	**43.76**	**38.18**	**37.10**	**42.05**	**44.03**	**41.39**	**39.70**	**36.99**	**40.40**
SSIM	Xu15	0.976	0.962	0.966	0.950	0.974	0.963	0.960		0.964
	Ours	**0.995**	**0.989**	**0.989**	**0.991**	**0.994**	**0.994**	**0.989**	**0.982**	**0.990**

Test time (s)	BLF	IBLF	RGF	L_0	WMF	RTV	WLS	L_1	Xu15	Liu16	Ours
QVGA (320×240)	0.03	0.11	0.22	0.17	0.62	0.41	0.70	32.18	0.23	0.07	0.03
VGA (640×480)	0.12	0.40	0.73	0.66	2.18	1.80	3.34	212.07	0.76	0.14	0.12
720p (1280×720)	0.34	0.97	1.87	2.43	4.98	5.74	13.26	904.36	2.16	0.33	0.35