PACIFIC GRAPHICS 2019

Shadow Inpainting and Removal Using Generative Adversarial Networks with Slice Convolutions

Jinjiang Wei[1], Chengjiang Long[2†], Hua Zou[1], Chunxia Xiao[1†]
[1]School of Computer Science, Wuhan University, Wuhan, Hubei, China
[2]Kitware Inc., Clifton Park, NY, USA
{weijinjiang, cxxiao, zouhua}@whu.edu.cn, chengjiang.long@kitware.com

Problem

Shadow Image

Input

Computer

Output

Shadow-free Image

Problem

- dark shadow
  - Lost color and texture clue information
  - For human vision, the information in the black hole region is clearly speculated
  - Shadow area and non-shadow area global contextual information are related

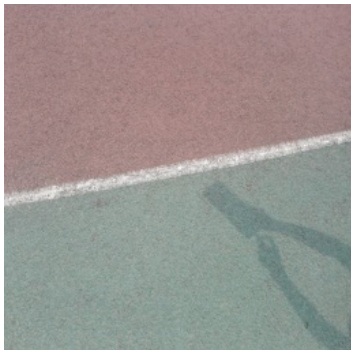
Image Inpainting and global contextual features can handle this problem together!

Inpainting Motivation

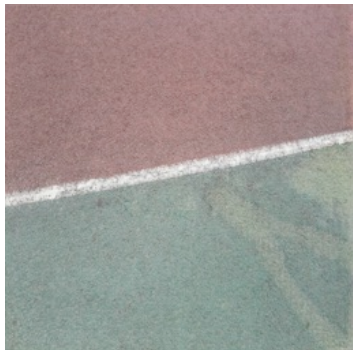
- Dark shadow region without direct color and texture information
- Similar to non-shadow areas
- Global context understanding
- Unlimited inpainting images

Related Work

- Shadow removal by deep neural networks



Input



ST-CGAN

DSC

ST-CGAN: WANG J., LI X., YANG J.: Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal. In CVPR (2018)

DSC: HU X., FU C.-W., ZHU L.:Direction-aware spatial context features for shadow detection. In CVPR (2018)

Our work

- Two-stage GAN inpainting and removal to get coarse and fine removal images separately
- Slice convolutions module to extract long-range contextual spatial information
- Non-local block to explore the local and global contextual spatial information

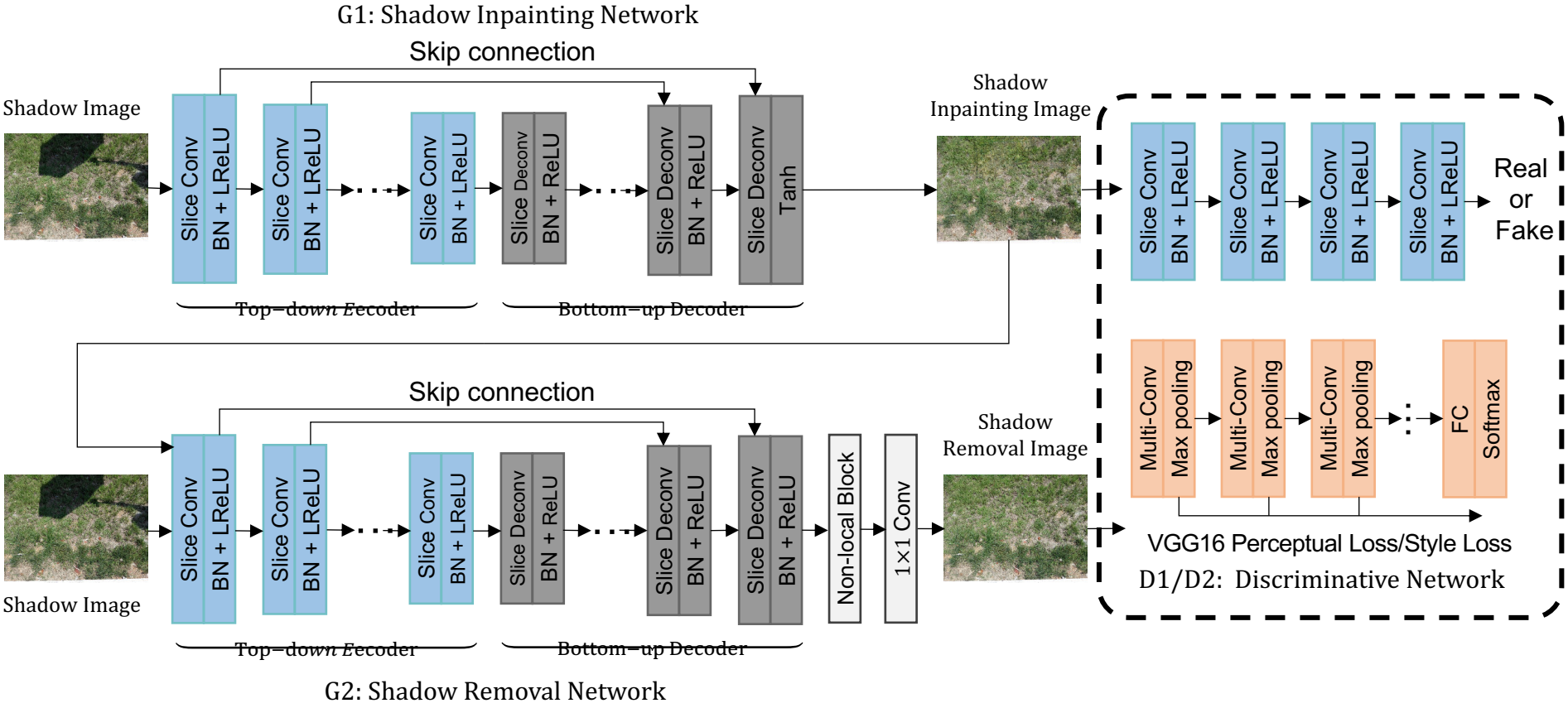
Input

Inpainting

Removal

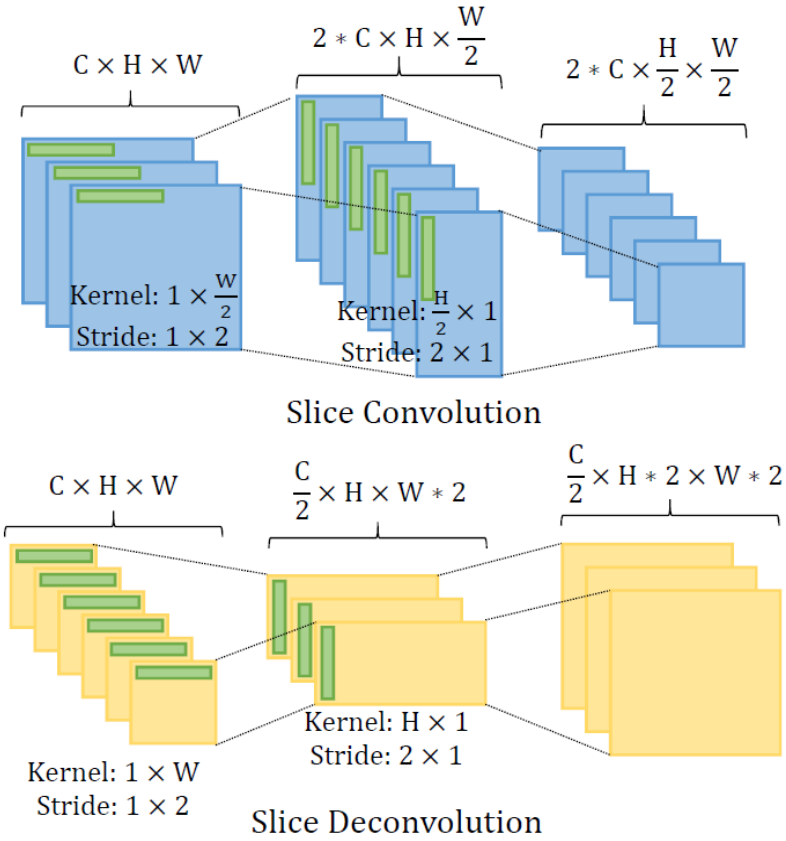
Ground truth

Overview


G1: Shadow Inpainting Network
Skip connection
Shadow Image
Slice Conv
BN + LReLU
Slice Conv
BN + LReLU
Slice Conv
BN + LReLU
Slice Deconv
BN + ReLU
Slice Deconv
BN + ReLU
Slice Deconv
Tanh
Top-down Encoder
Bottom-up Decoder
Shadow
Inpainting Image
Slice Conv
BN + LReLU
Slice Conv
BN + LReLU
Slice Conv
BN + LReLU
Slice Conv
BN + LReLU
Real
or
Fake
Skip connection
Shadow Image
Slice Conv
BN + LReLU
Slice Conv
BN + LReLU
Slice Conv
BN + LReLU
Slice Deconv
BN + ReLU
Slice Deconv
BN + ReLU
Slice Deconv
BN + ReLU
Non-local Block
1×1 Conv
Top-down Encoder
Bottom-up Decoder
Shadow
Removal Image
Multi-Conv
Max pooling
Multi-Conv
Max pooling
Multi-Conv
Max pooling
FC
Softmax
VGG16 Perceptual Loss/Style Loss
D1/D2: Discriminative Network
G2: Shadow Removal Network

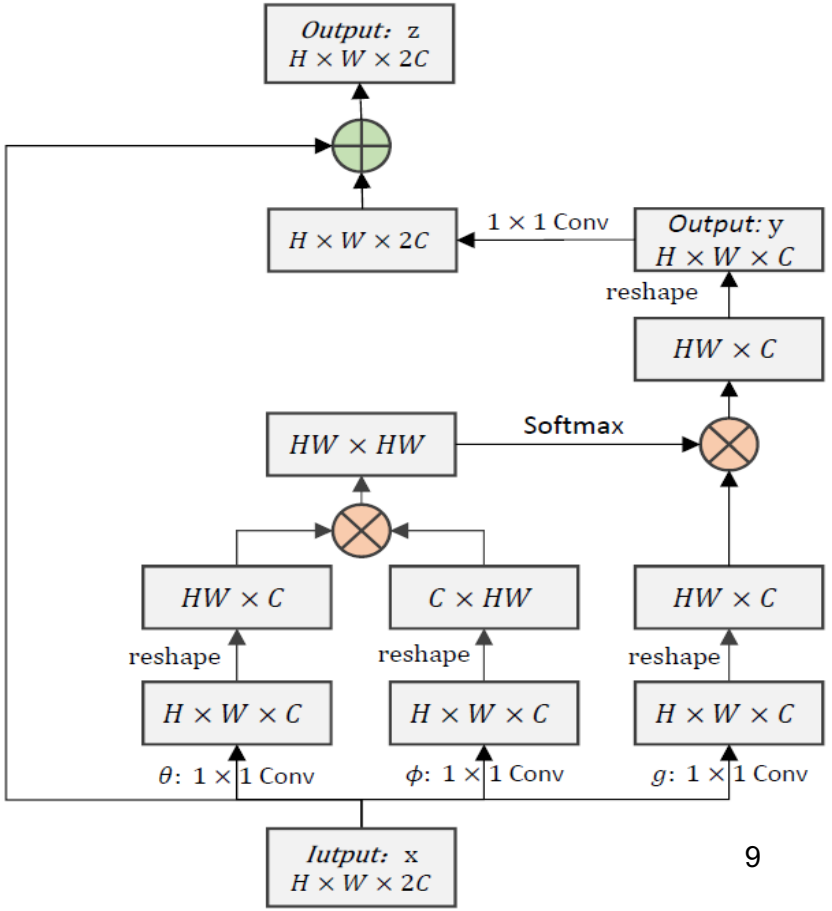
Slice Convolution

- Decomposition and less parameters
- Long-range dependency information



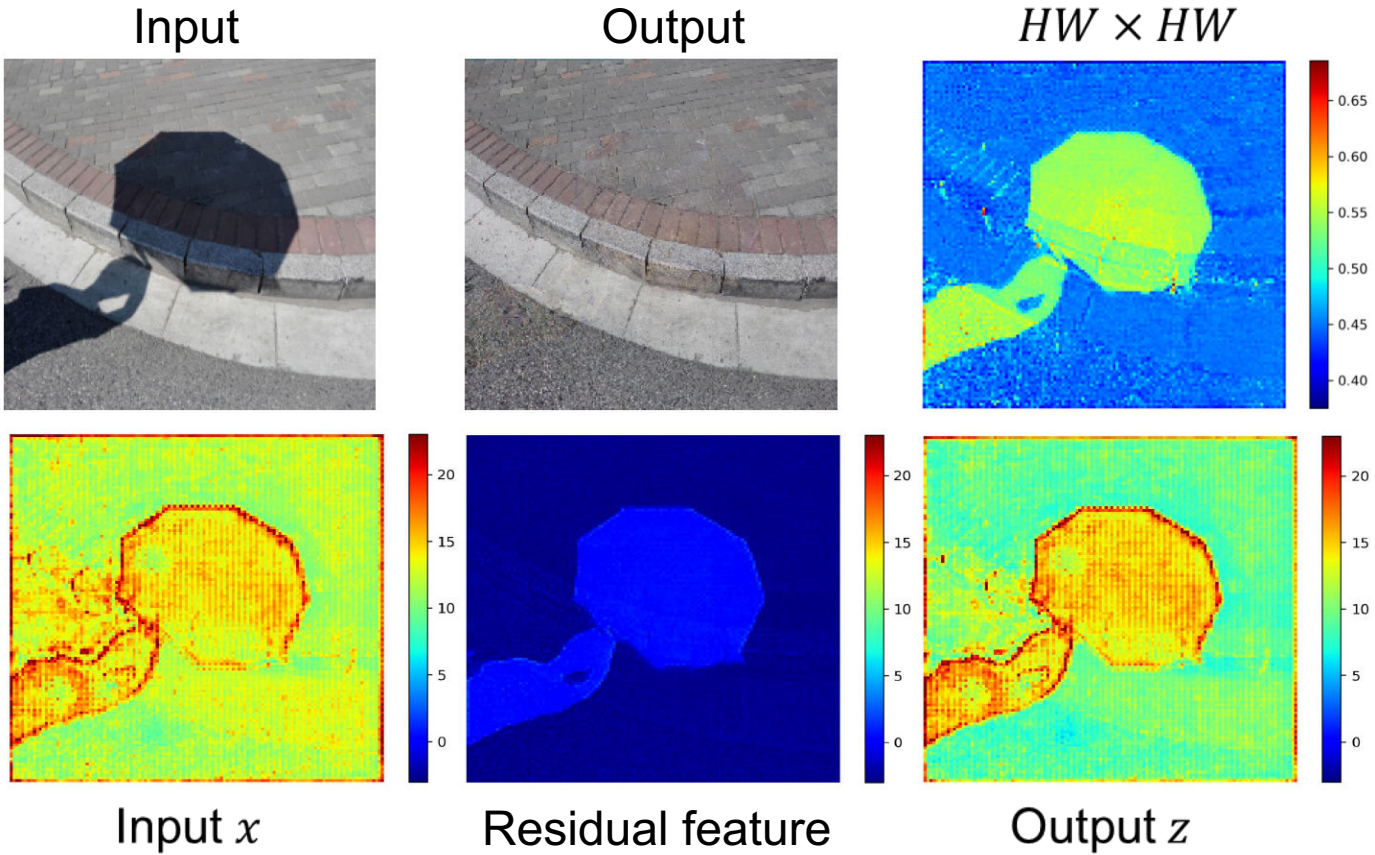
Non-local block

- Global context information to explore both local and global inter-dependencies of pixels



Wang X, Girshick R, Gupta A, et al. Non-local neural networks.

Non-local visualization



Loss Function

- cGAN loss + perceptual loss

$$G^* = \arg\min_G \max_D \mathcal{L}_{cGAN}(G,D) + \mathcal{L}_{total},$$

$$\mathcal{L}_{cGAN}(G,D) = \mathbb{E}_{x,y\sim p_{data}(x,y)}[\log D(x,y)] + \mathbb{E}_{x\sim p_{data}(x)}[\log(1 - D(x,G(x)))],$$

$$\mathcal{L}_{total} = \lambda\mathcal{L}_{L1} + \alpha\mathcal{L}_{perc} + \beta\mathcal{L}_{style} + \gamma\mathcal{L}_{tv},$$

Perceptual Losses

$$\mathcal{L}_{L1} = \frac{1}{C \times H \times W} \left\| I_{out} - I_{gt} \right\|_1$$

Difference between output and GT each pixel

$$\mathcal{L}_{perc} = \frac{1}{C_p \times H_p \times W_p} \sum_{h=1}^{H_p} \sum_{w=1}^{w_p} \left\| \Psi_p^{I_{out}} - \Psi_p^{I_{gt}} \right\|_1$$

Feature difference between generated image input to backbone network and GT input to backbone network

$$\mathcal{L}_{style} = \frac{1}{C_p \times H_p \times W_p} \sum_{h=1}^{H_p} \sum_{w=1}^{W_p} \left\| \left(\Psi_p^{I_{out}}\right)^T \left(\Psi_p^{I_{out}}\right) - \left(\Psi_p^{I_{gt}}\right)^T \left(\Psi_p^{I_{gt}}\right) \right\|_1$$

Same way, but Gram matrix first computed

$$\mathcal{L}_{tv} = \frac{1}{C \times H \times W} \sum_{(i,j) \in R} \left\| I_{out}^{i,j+1} - I_{out}^{i,j} \right\|_1 + \frac{1}{C \times H \times W} \sum_{(i,j) \in R} \left\| I_{out}^{i+1,j} - I_{out}^{i,j} \right\|_1$$

Difference between each pixel and neighbor pixel in generated image

Training and testing

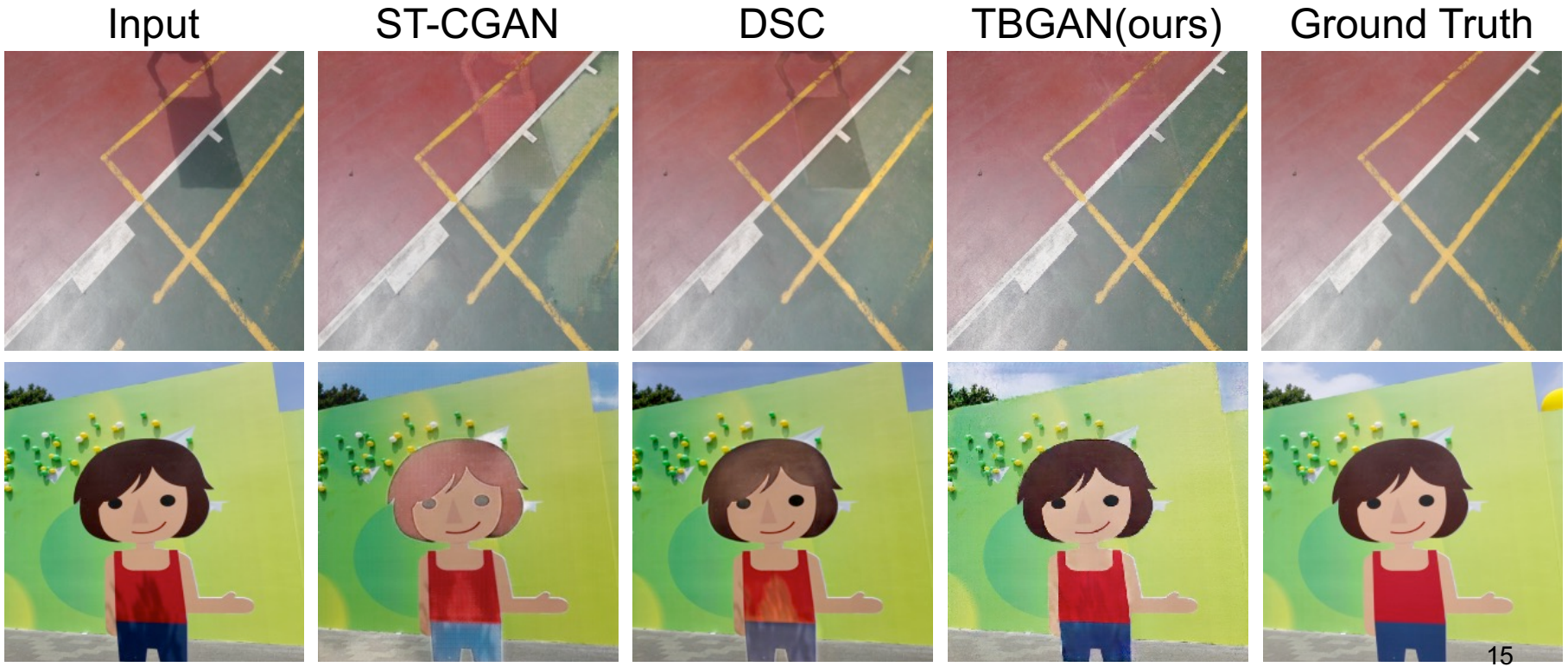
- Datasets:
  - ISTD dataset triples of shadow images, shadow masks, shadow-free images. 135 different shadow environments.
  - Inpainting finetune dataset Places365.
- Evaluation index
  - Root Mean Square Error (RMSE) in Lab space
  - Statistic shadow area and non-shadow area separately

Results - RMSE

- Comparison in ISTD dataset

	Shadow	Non-shadow	All
Guo [GDH12]	18.95	7.46	9.3
Gong [GC14]	14.98	7.29	8.53
Global/Local-GAN [ISSI17]	13.46	7.67	8.82
Pix2Pix-HD [WLZ*18]	10.63	6.73	7.37
Deshadow [QTH*17]	12.76	7.19	7.83
ST-CGAN [WLY18]	10.33	6.93	7.47
DSC [HFZ*18]	**9.22**	6.39	6.67
Two-stage TBGANs	10.14	5.91	6.70
Two-stage TBGANs+finetune	9.83	**5.58**	**6.39**

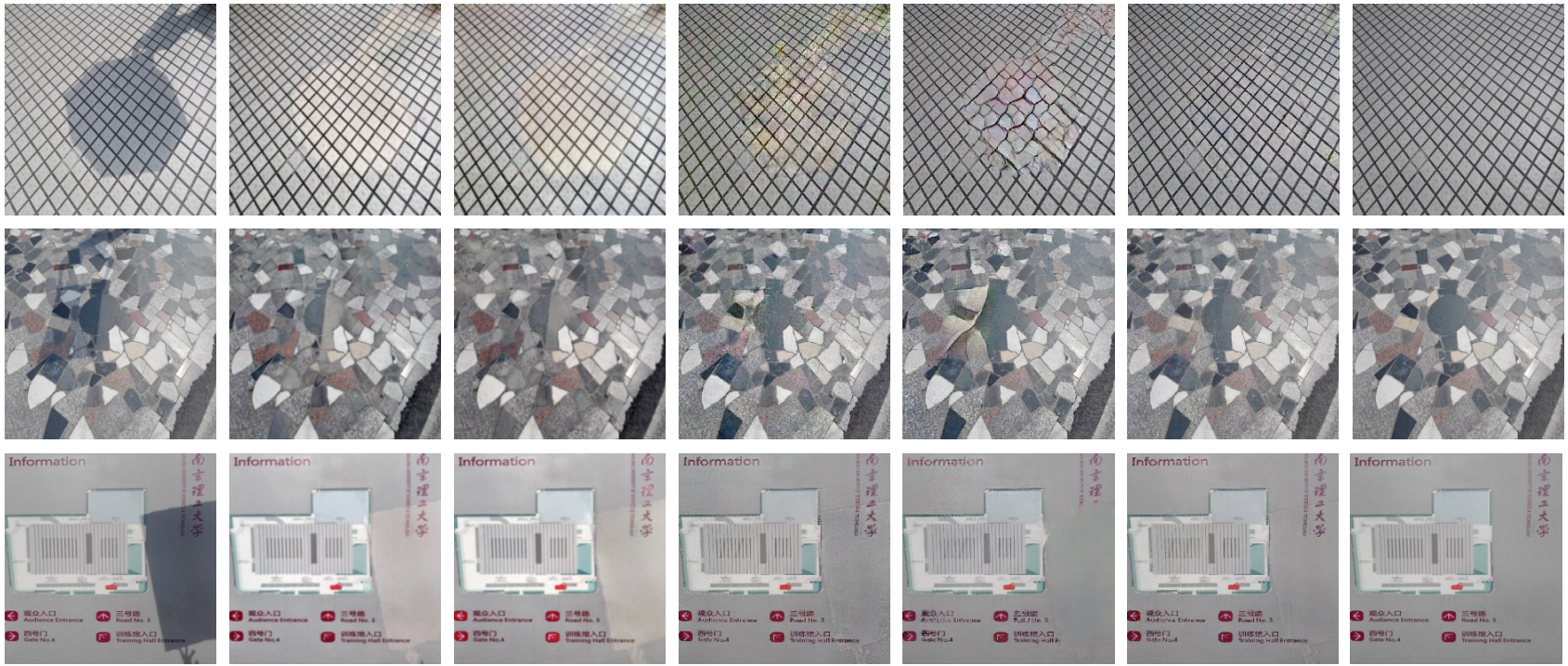
Visual Comparison Results



Ablation

	Shadow	Non-shadow	All
w/o TDBU&Inpainting	18.72	16.33	16.77
w/o TDBU	17.72	16.16	16.51
w/o Non-local block	10.33	5.89	6.79
1st-stage TBGAN	12.37	**5.58**	7.44
Two-stage TBGANs	**10.14**	5.91	**6.70**

Ablation

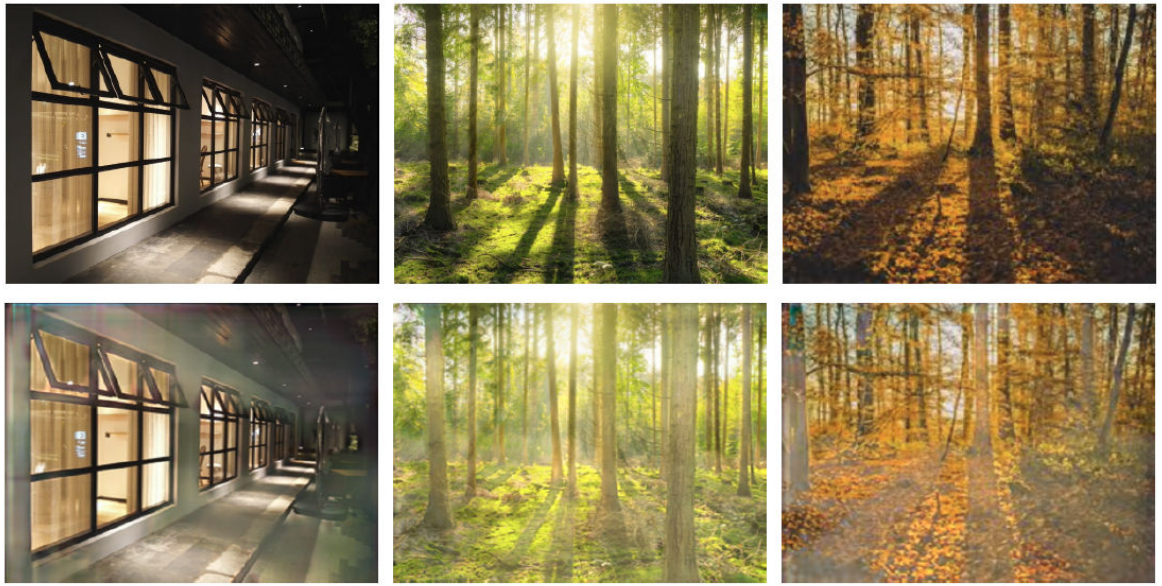


(a) (b) (c) (d) (e) (f) (g)

(a) input images, (b) the results without slice convolution and the 1-stage TBGAN for shadow inpainting, (c) the results without slice convolution, (d) the results without Non-local block, (e) the results of the 1st-stage TBGAN, (f) the results of our proposed two-stage TBGANs, and (g) the corresponding ground-truth shadow-free images, respectively.

Limitations

- Failure cases



- Complex boundaries and complicated shapes
- Shadow area occupies a large proportion

Conclusion and Future work

- Video shadow detection and removal
- Shadow removal for complex scenes
- Explore more real-world applications with slice convolutions

Paper QR Code:
http://www.chengjianglong.com/publications/SCShadow_CGF.pdf

Thank you!