Saliency Map-aided Generative Adversarial Network for RAW to RGB Mapping

Yuzhi Zhao[1] Lai-Man Po[1] Tiantian Zhang[1] Zongbang Liao[2] Xiang Shi[2] et al.

[1]City University of Hong Kong [2]Huazhong University of Science and Technology

ICCV 2019 Seoul, Korea

Problem Statement

Learn a **general mapping** from RAW file to RGB format.

1) **Saliency map** implicit data augmentation
2) **Fastest** network across all solutions that produces 28 images per second (224 × 224 resolution)
3) Transform Huawei RAW file to Canon RGB quality
4) Train a **RAW2RGB GAN** on your own dataset

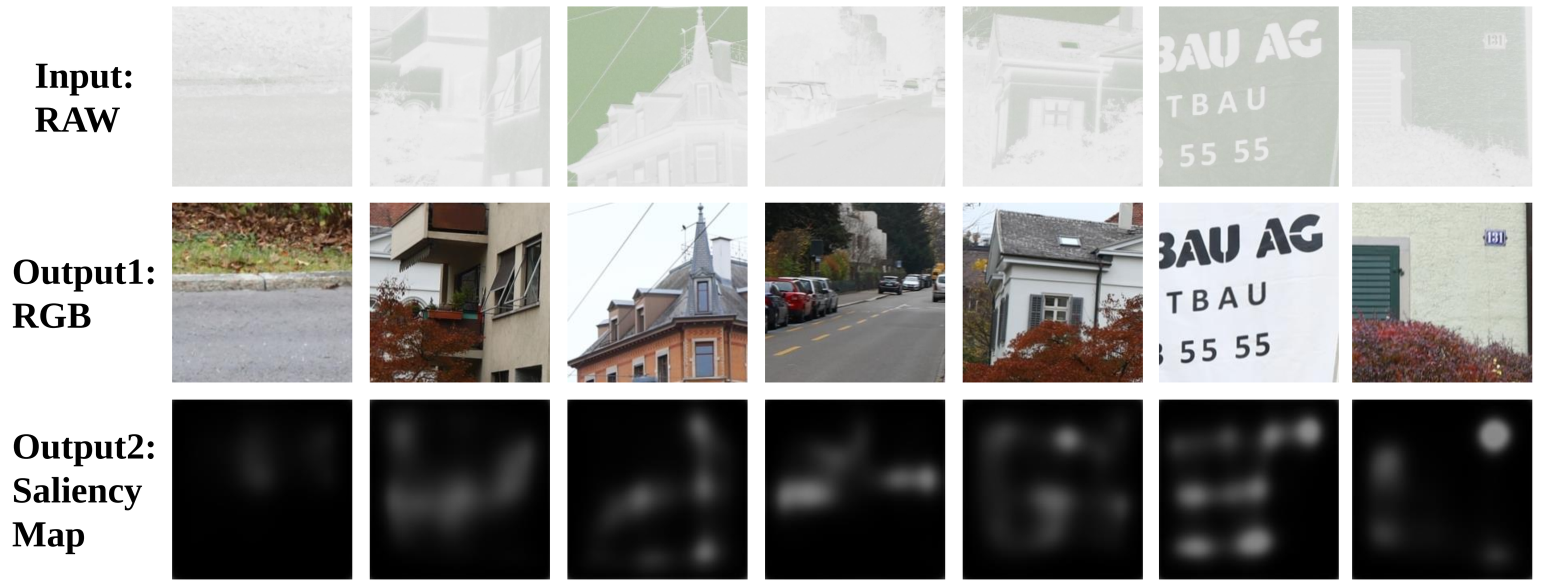
→ Try it: https://github.com/zhaoyuzhi/RAW2RGB-GAN

Saliency Map Data Augmentation (SMDA)

Problem: Traditional data augmentation method (flipping, cropping, rotation) only perform physical transformation.

Goal: Add semantic information to system, which is significant for many domain transfer applications

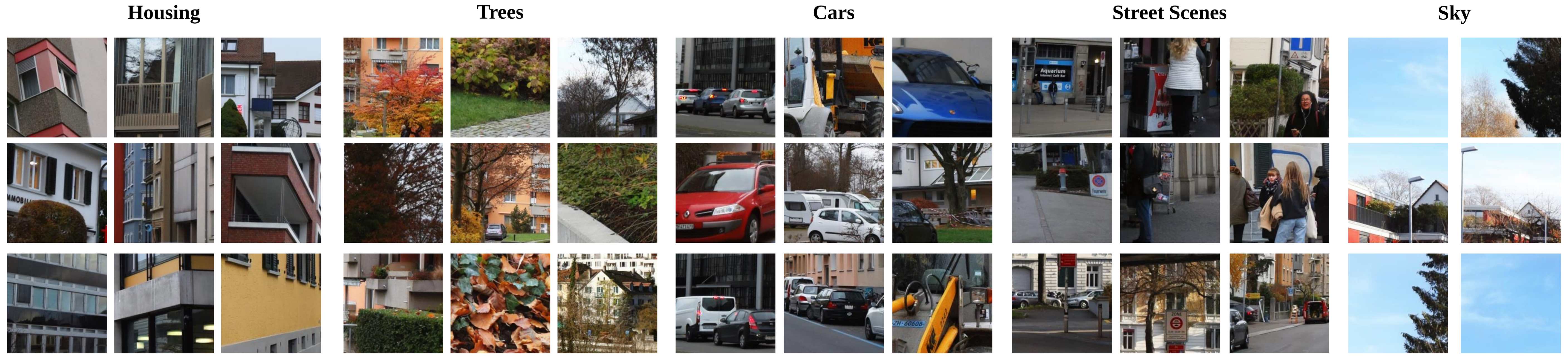
Procedure: A pre-trained Sal-GAN is adopted to generate saliency maps automatically. Then, the generated saliency maps are utilized to scale pixel-level L1 loss as a proxy target.



Analysis: RAW2RGB-GAN is a multi-task system. At backward propagation, the gradients from saliency map prediction branch revises mainstream.

Quantitative Result:
We have done an experiment on ImageNet training set on colorization task. SMDA obviously improve high-level representation. Please see more details through this link:
https://github.com/zhaoyuzhi/Semantic-Colorization-GAN

Qualitative Samples (224 × 224 Patches)



Qantitative Analysis

Baseline: Pix2Pix framework

Stage 1st: only L1 Loss, same epochs

Stage 2nd: L1 Loss, Perceptual Loss, GAN Loss, Attn Loss (Ours)

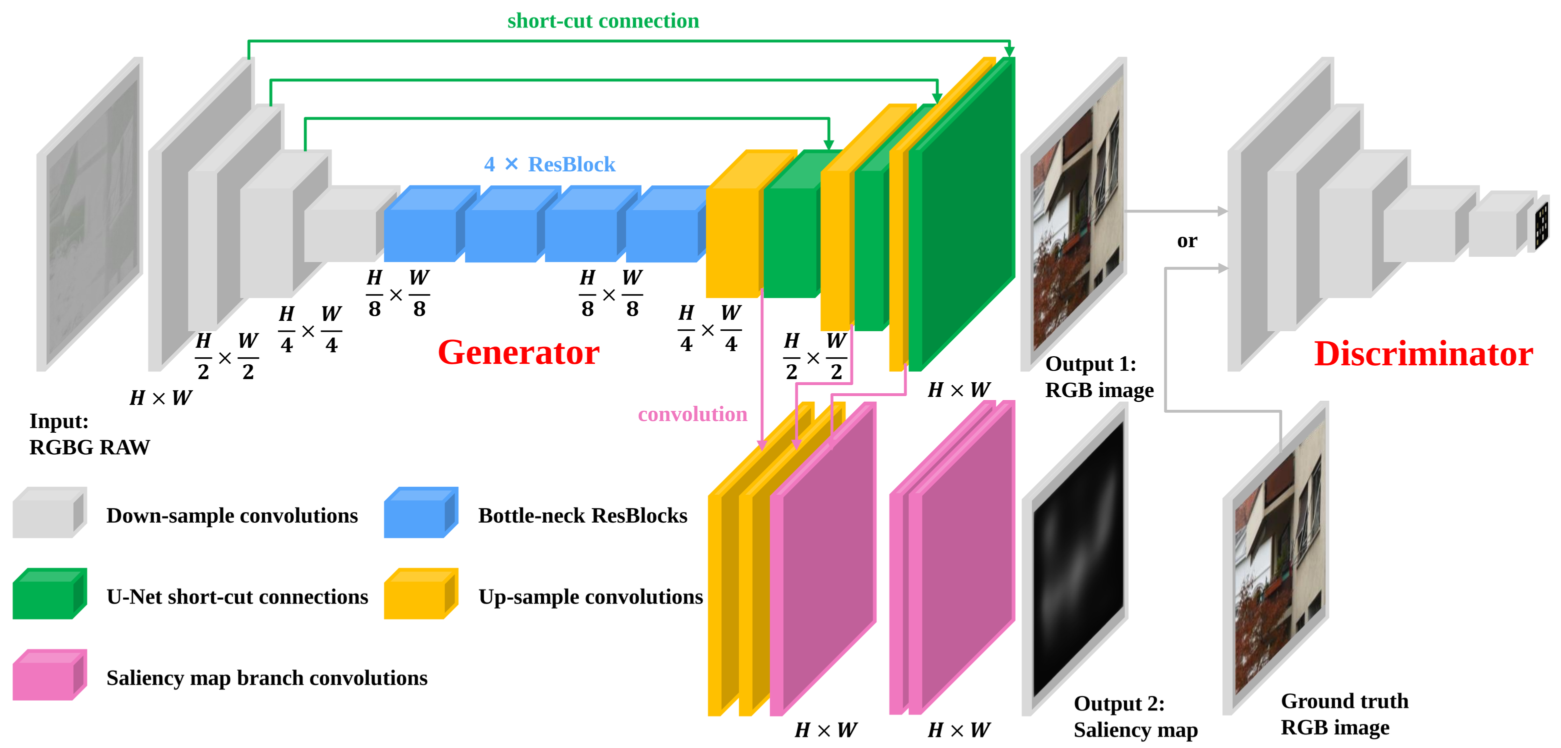
Method	PSNR	SSIM
Pix2Pix 1st	19.123838	0.712934
Pix2Pix 2nd	19.491088	0.727142
Ours L1	22.244904	0.798523
Ours 1st	22.067571	0.798382
Ours 2nd	**22.455825**	**0.798674**

Test Time: For 224 × 224 patches, it takes 0.03571s (27.98 images / second) to transfer an image on single 1080 Ti.

Full Resolution Testing Result: It takes only 0.5 second for rendering approximately 2000 × 1500 image.

RAW2RGB-GAN Architecture

There are three main parts: mainstream G_1, saliency map prediction branch G_2, and patch-based discriminator D.



Pixel L1 Loss: $L_1 = \mathbb{E}[\|G_1(x) - y\|_1]$; Attn Loss: $L_A = \mathbb{E}[\|G_1(x) \odot G_2(x) - y \odot s\|_1]$;

G Loss: $L_G = \frac{1}{2}\mathbb{E}[(D(G_1(x)) - 1)^2]$; Percep Loss: $L_p = \mathbb{E}\left[\|\phi_l(G(x)) - \phi_l(c)\|_1\right]$;

D Loss: $L_D = \frac{1}{2}\mathbb{E}[(D(G_1(x)) - 1)^2] - \mathbb{E}[(D(G_1(x)))^2]$

Failure Cases

1) Unreasonable blurry
2) Little color bleeding

In the future, multiple dataset and advanced network architecture can enhance mapping quality

Conclusion

We achieve 21.91 PSNR on ZRR testing set, and there are main contributions of proposed RAW2RGB-GAN.

1) The saliency map data augmentation (SMDA) enhances the training of network and has been demonstrated in other image translation tasks like colorization.
2) A GAN-based solution to automatically transform RAW file of phone to Canon DSLR camera RGB quality.
3) The fastest framework that generates nearly 30FPS.