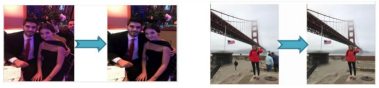
Deep Layer Regeneration: Image Reconstruction

Jesus Mendez (mendezj@stanford.edu)
CS 230 (Deep Learning), Stanford University

INTRODUCTION

Technological advances have not only influenced the ways we connect, but also the way we document our daily activities. Society today connects individuals by an array of social media platforms, all of which its user undergo substantial efforts to perfectly curate their online façade.

PROBLEM STATEMENT



Despite our best efforts in capturing picturesque memories, we cannot always capture them the way we intend to. Sublime scenery ruined by giant herds of peoples and focused sabotaged by our friends can all ruin the perfect moment. The explored challenge is removing objects from photos and replacing them with realistic renderings.

DATA

Dataset: People in Photo Albums [1]

- 37,107 photos, 60,000 instances of 2000 identities
- 17,000 Train, 5,684 Val, 7,868 Test, 6,555 leftover

Figure 1: People in Photo

Dataset: Places365 [2]

- 36,500 256x256 RGB images
- 400 Train, 125 Dev/Test, 35,975 leftover

Figure 2: Places365

MODELS

The network architecture is comprised of the following

- Mask R-CNN – Object Detection
- Image Conversion, resize and input preparation
- Generative Adversarial Network – Image Reconstruction

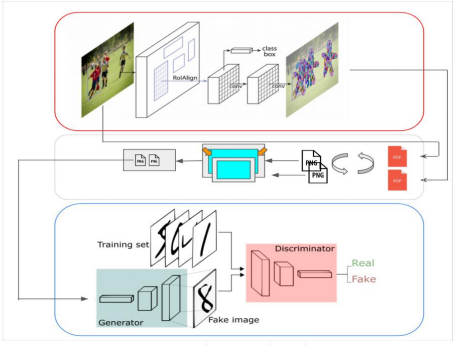


Figure 3: Neural Network Architecture

Pipeline process is as follows.

- Mask R-CNN Object Detection
- Modified colormap changes to specified color
- Contour of object wrapped
- Outputs full image resolution as PDF
- Converted image format
- Image is resized to 256x256
- Images are appended to each other forming 512x256
- Modifications to Images applied
- Outputs real image mask

	backbone	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
MNC [10]	ResNet-101-C4	24.6	44.3	24.8	4.7	25.9	43.6
FCIS [26] +OHEM	ResNet-101-C5-dilated	29.2	49.5	-	7.1	31.3	50.0
FCIS+++ [26] +OHEM	ResNet-101-C5-dilated	33.6	54.5	-	-	-	-
Mask R-CNN	ResNet-101-C4	33.1	54.9	34.8	12.1	35.6	51.1
Mask R-CNN	ResNet-101-FPN	35.7	58.0	37.8	15.5	38.1	52.4
Mask R-CNN	ResNeXt-101-FPN	**37.1**	**60.0**	**39.4**	**16.9**	**39.9**	**53.5**

Figure 4: Instance Segmentation [3]

RESULTS

Figure 5a: Mask R-CNN Results

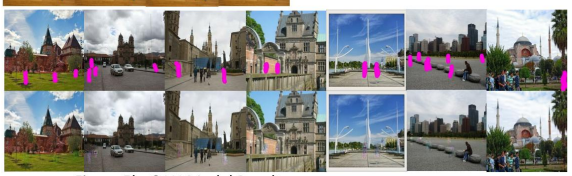


Figure 5b: GAN Model Results

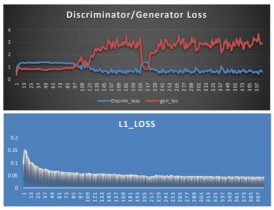


Figure 6: Loss

Discriminator and Generator loss results for GANs network of 50 Epochs. Images inputs (512x256)

CONCLUSION

- Removing unwanted objects successful
- Longer training/data required for increased photo realism
- Random RBG noise mask not required to affect real image distribution
- Friendship memories may vary

FUTURE WORK

- Train GANs Longer
- Add object reflections to Dataset
- Sync entire process

REFERENCE

[1] Ning Zhang, Manohar Paluri, Yaniv Taigman, Rob Fergus, Lubomir Bourdev: DATASET: Beyond Frontal Faces: Improving Person Recognition Using Multiple Cues
[2] Zhou, A. Lapedriza, A. Khosla, A. Oliva, and A. Torralba. Places: A 10 million image database for scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017.*
[3] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, Alexei A. Efros: Image-to-Image Translation with Conditional Adversarial Networks
[4] Kaiming He, Georgia Gkioxari, Piotr Dollar, Ross Girshick: Mask R-CNN