Deep Image Compositing

Shivangi Aneja, Soham Mazumder

Advisor: Prof. Dr. Matthias Nießner, Technical University of Munich, Germany

Introduction

Problem: Image Compositing, pasting of the foreground object from one image onto the background of another.

Motivation: Due to inconsistencies in appearances (like color, lighting and texture compatibility) of foreground with background makes the composite image unrealistic.

Contribution: We propose a conditional GAN with our custom loss for the task of image compositing that outperforms the current state-of-the-art.

Dataset & Network Architecture

No image compositing benchmark dataset is available.

Dataset Creation : Images from iCoseg[1] serve as the ground truth. To generate the edited images we make use of the method presented in [2].

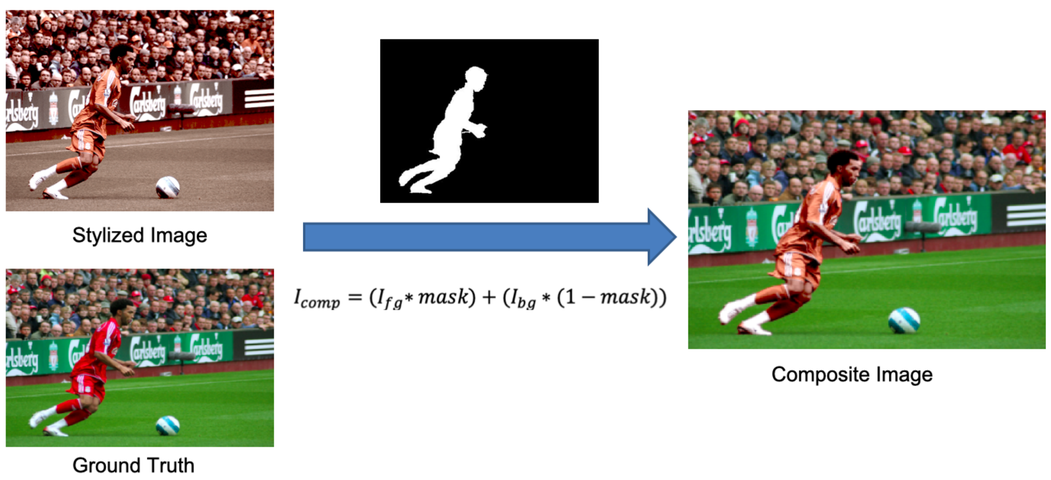

Figure: Image Compositing Pipeline

Generator Network : Simple autoencoder network with skip links that takes composite image as input makes it realistic.

Discriminator Network : Inspired from PatchGAN from Pix2Pix thus acting as a strong regularizer and penalizing more during training.

The novelty of our work lies in training the network with a variety of special losses that helps the network nicely blend the foreground object with the background.

Loss Functions & Metrics

Reconstruction Loss: L1-loss or L2-loss on the RGB images.

Perceptual Loss: Used for Style Transfer.

HSV Loss: Convert RGB images to HSV colour space and then compute channel wise L2 loss for Hue and Saturation channels only.

$$L_{hue} = L_2(I_{generated}[hue] - I_{real}[hue]) \quad (1)$$

$$L_{sat} = L_2(I_{generated}[sat] - I_{real}[sat]) \quad (2)$$

$$L_{hsv} = L_{hue} + L_{sat} \quad (3)$$

$$L_{generator} = L_{adversarial} + \lambda_1 L_{recon} + \lambda_2 L_{perceptual} + \lambda_3 L_{hsv} \quad (4)$$

Peak Signal To Noise Ratio (PSNR) : This is ratio of maximum possible signal power to power of corrupting noise.

Structural Similarity Index (SSIM) : Predicts the perceived quality of image using luminance masking and contrast masking.

Visual Information Fidelity (VIF) : This metric interprets the image quality as "fidelity" with the reference image.

Results

We present our results and compare them with the current state of the art.

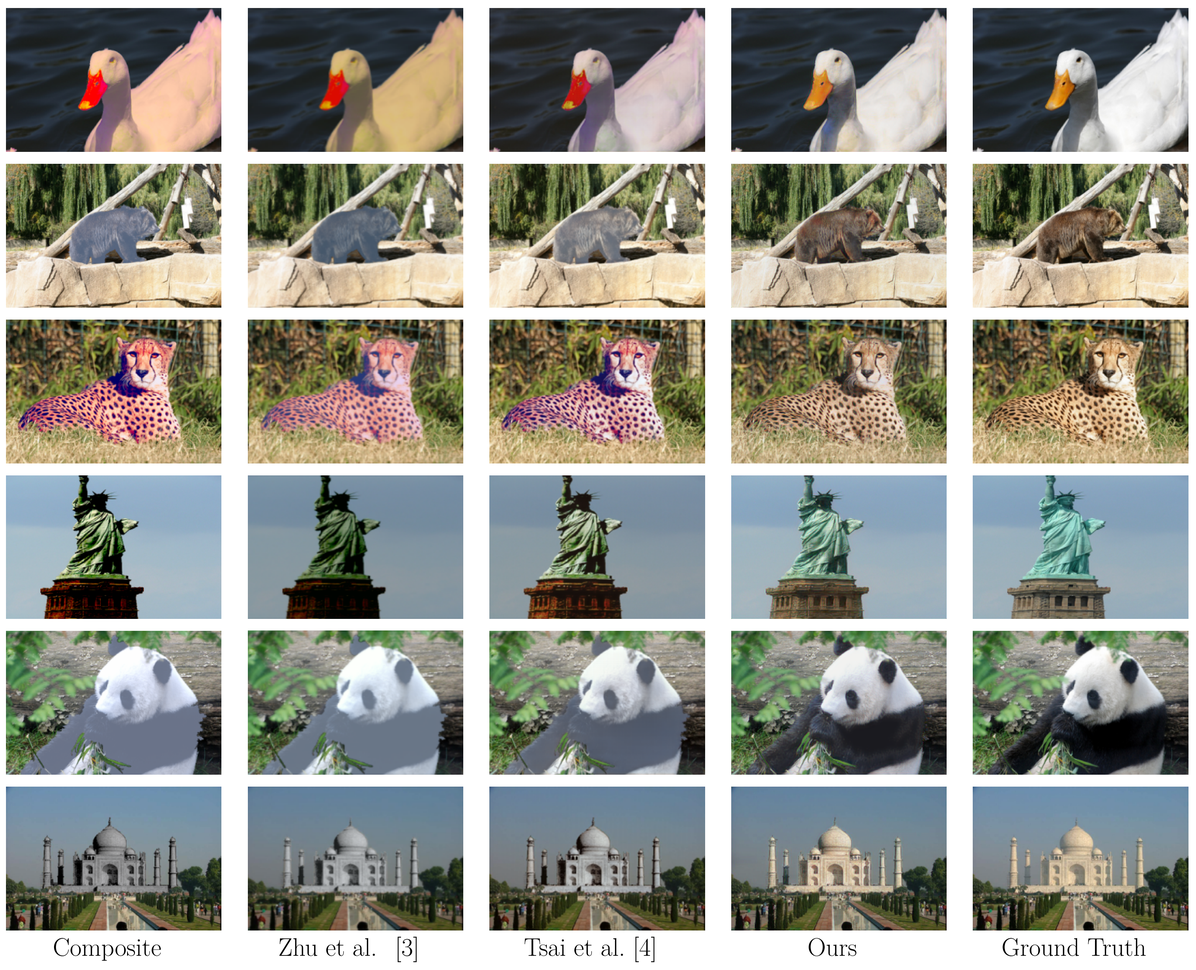

Figure: Comparison of results on the synthesized dataset.

Configuration	MSE	PSNR	SSIM	VIF
Ours (No Patch GAN)	1043.65	20.94	0.88	0.61
Ours (Patch GAN)	**430.163**	**22.189**	**0.935**	**0.898**
Tsai et al. [4]	2176.581	15.824	0.903	0.839
Zhu et al. [3]	3291.268	14.347	0.824	0.85

Table: Quantitative Results

Conclusion

We used GAN to generates images with a high degree of realism. A definite advantage of our network compared to the current state-of-the-art [4] is that our network does not require foreground masks to generate realistic images.

Figure: Generated images which are realistic but not similar to ground truth.(Composite, Generated, Ground Truth)

Experimental results in figure above show that our network is able to learn a diverse range of filters and generates images that appear visually realistic.

References

[1] Dhruv Batra, Adarsh Kowdle, Devi Parikh, Jiebo Luo, and Tsuhan Chen. Icoseg: Interactive co-segmentation with intelligent scribble guidance. *2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010.

[2] Joon-Young Lee, Kalyan Sunkavalli, Zhe Lin, Xiaohui Shen, and In So Kweon. Automatic content-aware color and tone stylization. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[3] Jun-Yan Zhu, Philipp Krahenbuhl, Eli Shechtman, and Alexei A Efros. Learning a discriminative model for the perception of realism in composite images. *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015.

[4] Yi-Hsuan Tsai, Xiaohui Shen, Zhe Lin, Kalyan Sunkavalli, Xin Lu, and Ming-Hsuan Yang. Deep image harmonization. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jul 2017.