Generatively Removing Photobombers

Thibault Févry
New York University
thibault.fevry@nyu.edu

Mihir Rana
New York University
ranamihir@nyu.edu

Kenil Tanna
New York University
kenil@nyu.edu

1 Introduction

Recent progresses in image segmentation have led to algorithms that can reliably provide bounding boxes around objects or even label every pixel of an image. On another front, the development of generative adversarial networks (GANs) has led to new approaches for image completion. In this project, our goal is to combine the two approaches to enable easy removal of undesired objects from images while reconstructing a realistic background. In short, we want to solve photo-bombing.

2 Methodology

The objective of this project would be to first perform image segmentation and remove certain objects, and then to fill the blank with the background.

Our work will focus on integrating both approaches to provide an end-to-end pipeline where users select a pixel, the object whose pixel is selected is identified and removed and the generative model will fill the image back in.

We expect to train (or use pre-trained models if training is deemed too computationally expensive) our models on the MS-COCO data set [7] and then integrate them on our pipeline. The two pre-trained models will be trained separately as there does not seem to be any benefit from training end-to-end.

Evaluation for both models trained independently will follow standard evaluation metrics in related work (see below). For the end-to-end pipeline, we expect to conduct mostly qualitative evaluation.

3 Related Work

Removing photobombers or specific people from images is a problem that has been worked on from many years. Deterministic algorithms like [1] show very good results by using a patch based method to identify the background and remove objects . Although they do have some shortcomings, as in the algorithm is not able to handle curved structures, sense of depth and is not able to work well when the patches do not have similar structures. In deep learning variational auto encoders (VAE)[6] and generative adversarial networks (GAN) [3] have been quite popular as generative models and have shown quite good results for specific tasks. One such method using GAN was for Semantic Image Inpainting [11] which showed good results on reconstructing faces and showed better results than DC-GAN [8] and VAE for this task. where they showed good results by using a context loss for the remaining image area and a prior loss to penalize the network for generating unrealistic images. Iizuka et al. [5] show a globally and locally consistent training of an adversarial network so that their model is able to fill up novel objects but the shortcomings are that they are not able to fill up complex structures like people and animals but works well with backgrounds. Recently, Ulyanov et al. [9] showed that randomly-initialized neural networks used as handcrafted priors give excellent results in inpainting, and thus that the structure of a generator network is sufficient to capture low-level image statistics prior to any learning, rather than learning realistic image priors from a large number of example images. [10] propose a blind inpainting solution that does not need the information

31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

regarding the region that requires inpainting to be given a priori, with good results for tasks such as text removal.

4 Evaluation Method

4.1 Evaluation for Image Segmentation

Following the method of Mask-RCNN[4], we will evaluate image segmentation methods under the AP at different intersection over union thresholds using Facebook's Detectron [2].

4.2 Evaluation for Image Inpainting

For image inpainting, we will use the standard metric of pixel-wise square difference (only where we apply the mask).

4.3 Evaluation for the End-to-end Model

This will be mostly qualitative analysis. We will compare the input and the output of our model and assess whether the image removal and the image inpainting leave any artifacts or unrealistic features.

References

[1] A. Criminisi, P. Perez, and K. Toyama. Region filling and object removal by exemplar-based image inpainting. *IEEE Transactions on Image Processing*, 13(9):1200–1212, Sept 2004.

[2] Ross Girshick, Ilija Radosavovic, Georgia Gkioxari, Piotr Dollár, and Kaiming He. Detectron. `https://github.com/facebookresearch/detectron`, 2018.

[3] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[4] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross B. Girshick. Mask r-cnn. *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2980–2988, 2017.

[5] Satoshi Iizuka, Edgar Simo-Serra, and Hiroshi Ishikawa. Globally and locally consistent image completion. *ACM Trans. Graph.*, 36(4):107:1–107:14, July 2017.

[6] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[7] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft coco: Common objects in context. In David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars, editors, *Computer Vision – ECCV 2014*, pages 740–755, Cham, 2014. Springer International Publishing.

[8] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *CoRR*, abs/1511.06434, 2015.

[9] Dmitry Ulyanov, Andrea Vedaldi, and Victor S. Lempitsky. Deep image prior. *CoRR*, abs/1711.10925, 2017.

[10] Junyuan Xie, Linli Xu, and Enhong Chen. Image denoising and inpainting with deep neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 341–349. Curran Associates, Inc., 2012.

[11] Raymond A. Yeh, Chen Chen, Teck-Yian Lim, Alexander G. Schwing, Mark Hasegawa-Johnson, and Minh N. Do. Semantic image inpainting with deep generative models. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6882–6890, 2017.