EE 367 Project Proposal: Image Inpainting with DCGAN

Qiwen Fu (qiwenfu), You Guan (you17), Yuxin Yang (yuxiny)

February 17, 2018

1 Motivation

Image inpainting has always been a challenging and ongoing field of exploration for consumers. With the rising demand of taking nice pictures, lots of effort has been invested in building better tools for users to take the pictures they want. But sometimes, users could take undesirable pictures when their object of interest is obstructed by the unrelated. For example in Fig.1, our teammate, who is our object of interest, is blocked by a photographer who accidentally blocked the scene by her finger. Or, sometimes, we want to take a picture of the main quad in sun, but blocked by crowded students or visitors whom we want to get rid of from our picture (Fig.2).

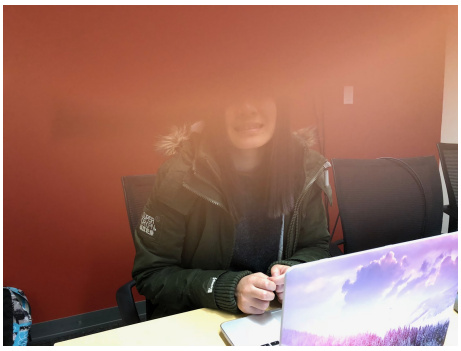

Figure 1: Our teammate You is blocked by photographer's finger.

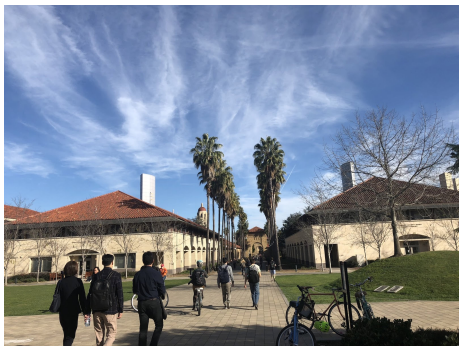

Figure 2: Crowds blocking the view of main quad.

Motivated by these could-be-improved pictures, in our project, we are going to attempt to tackle the problem of getting ride of these undesirable objects in these pictures by using generative and adversarial nets.

2 Related Work

The state-of-the-art learning based method requires specific information about the holes in the training phase. And there are some existing CNN methods trained to generate an arbitrary missing segment based on its surroundings in the image with an algorithm driven by context-based pixel prediction[4] Pathak experimented with two different loss models: a standard pixel-wise reconstruction loss, and a reconstruction plus an adversarial loss, with the second one giving better result than the first one. But their final results were still unsatisfactory as the generated fillings for the missing segment were mostly blurry and distorted. And we think it was because of that, as we mentioned in the motivation, this method extracts information from only one single image. This defect motivates us to implement a method that would generate the missing content by conditioning on the set of available data. Another inspiration from this work is that we could try different loss models if we are not getting satisfactory results.

Based on those state of art works, there were some recent break-through. In 2016, Yeh and his group from UIUC claiming that their model outperforms the state-of-the-art methods by innovatively implementing GAN. And their outcome achieves pixel-level photorealism.

Yeh and his team proposes to use deep generative models for image inpainting [5]. They conduct experiments on three datasets, including the CelebFaces Attributes Dataset (CelebA), the Street View House Numbers (SVHN) and the Stanford Cars Dataset. They show generative results by GANs with high image

quality, compared with TV inpainting [1], LR inpainting [3] and nearest neighbor inpainting [2], which are three relatively mature methods to take care of this problem.

However, the dataset they used are too specific with a relatively simple latent spaces, making the inpainting task easy. This defect motivates us to research deeper into the use of GANs in inpainting. We are going to experiment on a more diverse dataset, such as CIFAR-10, and try to train a model that is more general.

3 Project Overview

In this project, we intend to implement the recently published Deep Conditional Generative and Adversarial Network (DCGAN) on image inpainting. GANs incorporates two neural networks - a generator which takes in the incomplete image as the input and generates a fake full image with the missing segment filled, and a discriminator which tries to distinguish generated images from original full images. The generator and the discriminator are trained at the same time and trying to beat each other.

Base on Yeh's work, we intend to overcome their shortcoming of using some too specific datasets with a relatively simple latent spaces. With all of the existing work considered, we intend to design a new model to fulfill such tasks like inpainting for some more general purposes. One application would be to develop mobile Apps to erase tourists from the landscape we care about. Some other applications might include restoring old or corrupted photographs, or lowering the cost of creating huge sceneries in movies and so on.

For our dataset, we plan to use landscape dataset as original image, and randomly crop out a patch for training purpose. It is better demonstrated in Fig.3, 4.

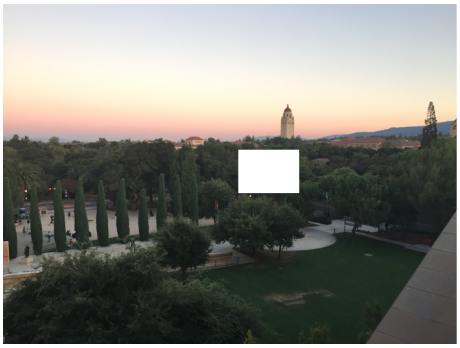

Figure 3: Training sample created from our dataset by randomly dropping a patch from the original image.

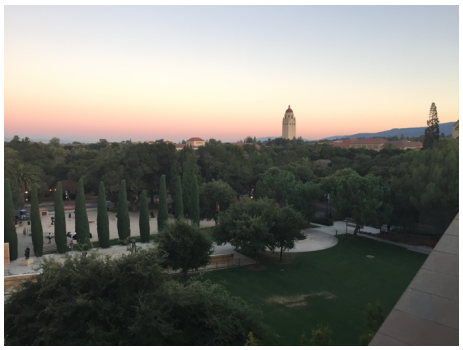

Figure 4: Ground truth from our dataset. We will use this image as a target we try to achieve by our GAN.

One thing we are concerned about is the computational time for our model. We hope to coordinate with the class staff more to obtain some GPU resources in order to accelerate the process of training our deep neural net.

4 Milestones

2/13: Proposal due.
2/13 - 2/20: Obtain and prepare dataset, read more on related literature.
2/20 - 3/6: Implement and train the neural networks. Optimize result.
3/6 - 3/13: Finalize work, prepare the poster and write project report.
3/13: Poster session.
3/16: Project report and code due.

References

[1] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo. An augmented lagrangian approach to the constrained optimization formulation of imaging inverse problems. *IEEE Transactions on Image Processing*, 20(3):681–695, March 2011.

[2] J. Hays and A. A. Efros. Scene completion using millions of photographs. In *ACM SIGGRAPH 2007 Papers*, SIGGRAPH '07, New York, NY, USA, 2007. ACM.

[3] C. Lu, J. Tang, S. Yan, and Z. Lin. Generalized nonconvex nonsmooth low-rank minimization. *CoRR*, abs/1404.7306, 2014.

[4] D. Pathak, P. Krähenbühl, J. Donahue, T. Darrell, and A. A. Efros. Context encoders: Feature learning by inpainting. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2536–2544, June 2016.

[5] R. A. Yeh, C. Chen, T. Lim, M. Hasegawa-Johnson, and M. N. Do. Semantic image inpainting with perceptual and contextual losses. *CoRR*, abs/1607.07539, 2016.