

Semantic Image Inpainting with Deep Generative Models

Liangjie Cao

July 24, 2018

Abstract

Semantic image inpainting is a challenging task where large missing regions have to be filled based on the available visual data. Existing methods which extract information from only a single image generally produce unsatisfactory results due to the lack of high level context. In this paper, we propose a novel method for semantic image inpainting, which generates the missing content by conditioning on the available data. Given a trained generative model, the authors search for the closest encoding of the corrupted image in the latent image manifold using our context and prior losses. This encoding is then passed through the generative model to infer the missing content. In this method, inference is possible irrespective of how the missing content is structured, while the state-of-the-art learning based method requires specific information about the holes in the training phase. Experiments on three datasets show that our method successfully predicts information in large missing regions and achieves pixel-level photorealism, significantly outperforming the state-of-the-art methods

1. Introduction

Semantic inpainting [4] refers to the task of inferring arbitrary large missing regions in images based on image semantics. Since prediction of high-level context is required, this task is significantly more difficult than classical inpainting or image completion which is often more concerned with correcting spurious data corruption or removing entire objects. Numerous applications such as restoration of damaged paintings or image editing [1] benefit from accurate semantic inpainting methods if large regions are missing. However, inpainting becomes increasingly more difficult if large regions are missing or if scenes are complex.

Classical inpainting methods are often based on either local or non-local information to recover the image. Most existing methods are designed for single image inpainting. This assumption is hard to satisfy, if the missing region is large and possibly of arbitrary shape. Consequently, in this case, these methods are unable to recover the missing in-

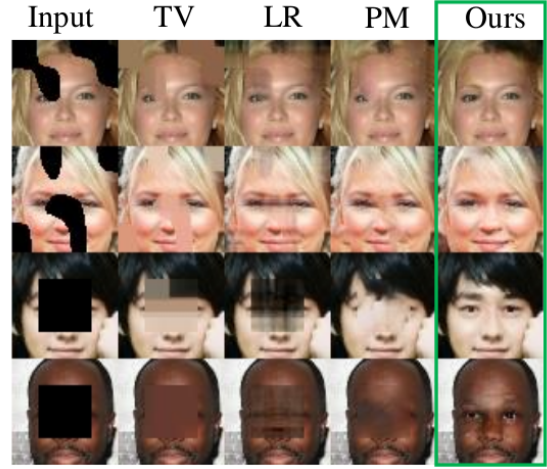


Figure 1. Semantic inpainting results by TV, LR, PM and their method. Holes are marked by black color

formation. Fig. 1 shows some challenging examples with large missing regions, where local methods fail to recover the nose and eyes.

2. Related Work

2.1. GAN

(GANs) are a framework for training generative parametric models, and have been shown to produce high quality images [2, 3, 5]. This framework trains two networks, a generator, G , and a discriminator D . G maps a random vector Z , sampled from a prior distribution p_Z , to the image space while D maps an input image to a likelihood. The purpose of G is to generate realistic images, while D plays an adversarial role, discriminating between the image generated from G , and the real image sampled from the data distribution p_{data} .

The G and D networks are trained by optimizing the loss function is familiar to us:

$$\min_G \max_D V(G, D) = E_{h \sim p_{data}(h)} [\log(D(h))] + E_{Z \sim p_Z(h)} [\log(1 - (D(Z)))] \quad (1)$$

where h is the sample from the p_{data} distribution; Z is a random encoding on the latent space.

- [5] A. Radford, L. Metz, and S. Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. In *ICLR*, 2016. 1

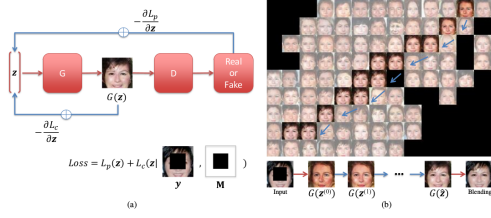


Figure 2. The proposed framework for inpainting. (a) Given a GAN model trained on real images, they iteratively update z to find the closest mapping on the latent image manifold, based on the designed loss functions. (b) Manifold traversing when iteratively updating z using back-propagation. $z^{(0)}$ is random initialed; $z^{(k)}$ denotes the result in k -th iteration; and \hat{z} denotes the final solution

3. Semantic Inpainting by Constrained Image Generation

To fill large missing regions in images, this method for image inpainting utilizes the generator G and the discriminator D , both of which are trained with uncorrupted data. After training, the generator G is able to take a point z drawn from p_Z and generate an image mimicking samples from p_{data} . They hypothesize that if G is efficient in its representation then an image that is not from p_{data} (e.g., corrupted data) should not lie on the learned encoding manifold, z . Therefore, we aim to recover the encoding \hat{z} “closest” to the corrupted image while being constrained to the manifold, as illustrated in Fig. 2; they visualize the latent manifold, using t-SNE on the 2-dimensional space, and the intermediate results in the optimization steps of finding \hat{z} . After \hat{z} is obtained, they can generate the missing content by using the trained generative model G .

References

- [1] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester. Image inpainting. In *ACMSIG*, 2000. 1
- [2] E. L. Denton, S. Chintala, R. Fergus, et al. Deep generative image models using a laplacian pyramid of adversarial networks. In *NIPS*, 2015. 1
- [3] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *NIPS*, 2014. 1
- [4] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros. Context encoders: Feature learning by inpainting. In *CVPR*, 2016. 1