Unsupervised Representation Learning with Deep Convolutional Generative A DVERSARIAL N ETWORKS

Liangjie Cao

July 22, 2018

Abstract

In this paper we introduce a generative parametric model capable of producing high quality samples of natural images. Our approach uses a cascade of convo- lutional networks within a Laplacian pyramid framework to generate images in a coarse-to-fine fashion. At each level of the pyramid, a separate generative convnet model is trained using the Generative Adversarial Nets (GAN) approach [1]. Samples drawn from our model are of significantly higher quality than alternate approaches. In a quantitative assessment by human evaluators, our CIFAR10 samples were mistaken for real images around 40% of the time, compared to 10% for samples drawn from a GAN baseline model. We also show samples from models trained on the higher resolution images of the LSUN scene dataset.

1. Introduction

Building a good generative model of natural images has been a fundamental problem within computer vision. However, images are complex and high dimensional, making them hard to model well, despite extensive efforts. Given the difficulties of modeling entire scene at high-resolution, most existing approaches instead generate image patches. In contrast, in this work, we propose an approach that is able to generate plausible looking scenes at 32 x 32 and 64 64. To do this, we exploit the multi-scale structure of natural images, building a series of generative models, each of which captures image structure at a particular scale of a Laplacian pyramid. This strategy breaks the original problem into a sequence of more manageable stages. At each scale we train a convolutional network-based generative model using the Generative Adversarial Networks (GAN) approach of Goodfellow *et al.* [1]. Samples are drawn in a coarse-to-fine fashion, commencing with a low-frequency residual image. The second stage samples the band-pass structure at the next level, conditioned on the sampled residual. Subsequent levels continue this process, always conditioning on the output from the previous scale, until the final level is reached. Thus drawing samples is an efficient and straightforward procedure: taking random vectors as input and running forward through a cascade of deep convolutional networks (convnets) to produce an image.

2. Laplacian Generative Adversarial Networks (LAPGAN)

Their proposed approach combines the conditional GAN model with a Laplacian pyramid representation. The model is best explained by first considering the sampling procedure. Following training (explained below), we have a set of generative convnet models G_0 , . . . , G_K , each of which cap- tures the distribution of coefficients h_k for natural images at a different level of the Laplacian pyramid. Sampling an image is akin to the reconstruction procedure, except that the generative models are used to produce the h_k's:

$$\hat{I}_k = \mu(\hat{I}_{k+1}) + \hat{h}_k = \mu(\hat{I}_{k+1}) + G_k(z_k, \mu(\hat{i}_{k+1})) \quad (1)$$

The recurrence starts by setting $\hat{I}_{k+1}$ = 0 and using the model at the final level G_K to generate a esidual image $\hat{I}_k$ using noise vector z_k : $\hat{I}_k = G_K(z_K)$. Note that models at all levels except the final are conditional generative models that take an upsampled version of the current image $\hat{I}_{k+1}$ as a conditioning variable, in addition to the noise vector z_k . Fig 1 shows this procedure in action for a pyramid with K = 3 using 4 generative models to sample a 64 x 64 image.

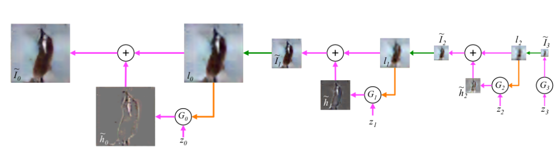

Figure 1. The sampling procedure for our LAPGAN model

References

[1] I. J. 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