Title: Photo-realistic Image Synthesis Conditioned on Text Descriptions using Stacked Generative Adversarial Networks (StackGAN)

Abstract:
The problem of generating high-quality images from text descriptions is a difficult one in computer vision, yet it has many applications. However, the samples that are generated by the existing text-to-image models fail to add details and vivid object parts. Moreover, it is very difficult to train GAN to generate high-resolution photo-realistic images from text descriptions. Training instability makes it even harder to generate high-resolution images using GAN models. Our research proposes a 256 x 256 Photo-realistic Image Synthesis using Stacked Generative Adversarial Network (GAN). We decompose the hard problem in two stages through a sketch refinement process. The Stage-I generator generates a low-resolution image by drawing the object's rough shape and basic colors from the provided text description and painting the background with a random noise vector. The Stage-II generator, which is based on Stage-I results and text descriptions as inputs, corrects faults and adds attractive features to them, resulting in a more realistic high-resolution image. Also, to improve the stability of training conditional-GAN this paper introduces a novel Conditioning Augmentation technique that encourages smoothness in the latent conditioning manifold. The experiments done and comparing with the state-of-the-arts on benchmark datasets concludes that our proposed model achieves significant improvements on generating high-quality Photo-realistic images.