Deep Learning - Generative Adversarial Text to Image Synthesis

Alex Valle (4854159) Gabriele Berruti (4791919)

Abstract

The paper that has been chosen for analysis is a document from 2016 that showcases the effort of six researchers from the University of Michigan and Saarbrücken (Germany). The document shows how the use of GANs (Generative Adversarial Networks) allows for advancements in the generation of synthetic images starting from a textual description. It compares the method proposed by the research with previous architectures that, although far from the described goal, are capable of obtaining valid textual feature representations. In particular, the paper aims to demonstrate the effectiveness of the model in generating images of birds and flowers based on a precise textual description of them.

Introduction

In 2016, the ability of an AI system to generate realistic and coherent images from textual descriptions (such as "small red bird with a blue beak") was a current issue and far from being achieved. It should be noted that it's necessary to use natural language and domain-specific attributes to describe the image to be generated.

Related Papers Architectures

This paper is based on 3 documents that serve as a starting point and tools for the proposed model:

- `(Farhadi et al., 2009; Kumar et al., 2009; Parikh & Grauman, 2011; Lampert et al., 2014)`: 3 Useful Papers for Encoding Distinctive Features of Objects into Vectors (such as attributes used to distinguish between different classes of objects)
- `(Fu et al., 2014 ; Akata et al., 2015)`: 2 Papers on "zero-shot" recognition, that is, recognizing objects that have never been seen during the model's training.
- `(Yan et al., 2015).`: And in Yan's paper, they discuss conditional image generation in a manner similar to the method proposed here.
- `(Reed et al., 2016)`: Reed's paper presents highly discriminative and generic "zero-shot" text representations, which are learnt automatically from words and characters.

Datasets

- `Caltech-UCSD birds database (Wah et al., 2011)` : Dataset used in related Papers previously described

How to reach the goal ?

The difficulty of translating words into images may be divided into two subproblems.

- `1.` : First, learn a feature vector from a specific text based on the visualization we want to obtain

.
- **2.** : Given these features through the use of a certain architecture, create a realistic and coherent image.