Chapter 10
Text to Image Synthesis Using Multistage Stack GAN

V. Dinesh Reddy
SRM University, India

Yasaswini Desu
SRM University, India

Medarametla Sindhu
SRM University, India

Chilukuri Vamsee
SRM University, India

Neelissetti Girish
SRM University, India

ABSTRACT

Many recent studies on text-to-image synthesis decipher approximately 50% of the problem only. They failed to compute all the imperative details in it. This chapter presents a solution using stacked generative adversarial networks (GAN) to generate lifelike images based on the given text. The stage-I GAN creates a distorted images by depicting the rudimentary/basic colours and shape of a scene predicted on text illustration. Stage-II GAN ends up on generating high-resolution images with naturalistic features using Stage-I findings and the text description as inputs. The output generated by this technique is more credible than many other techniques which are already in use. More importantly, stack GAN produces 256 x 256 images based on the text descriptions, while the existing algorithms produces 128 x 128.

DOI: 10.4018/978-1-6684-6937-8.ch010

Copyright © 2023, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

INTRODUCTION

Text-to-image synthesis describes that we are trying to convert the text descriptions into meaningful and appropriate images. It is one amongst the arduous problems in the computer vision (CV) sector and natural language processing (NLP) sector. We generally observe image captioning, where a caption will be given to an image after processing it. But, here we are trying to approach the problem in the reverse fashion i.e. caption to image mapping. A pictorial representation speaks a thousand words compared to oral or textual descriptions. Oral or textual descriptions can't provide comprehensive information. So, with the advancement of technology, this chapter is grueling towards converting human thoughts (textual descriptions) and ideas into visions. In a real-world scenario, text-to-image synthesis is a back-breaking issue due to the reason that there can be more than one scene that represent a single caption.

Nowadays, we have different neural network models like the Convolutional Neural Network (CNN), Recurrent Convolutional Neural network (RCNN), and many other models which uses the encoder-decoder mechanism. These architectures produce fact-based information. To our knowledge, we cannot generate captions with the help of limited or synthetic images. To address this issue, Generative Adversarial Networks (GANs) came into the picture, where we can generate synthetic images based on the given captions (Dosovitskiy et al. (2015)).

There are various Generative Adversarial Networks like Deep Convolutional GAN (DCGAN) that works on ConvNets. The ConvNets are using a stride without a pooling layer and the neurons in this model are not fully connected. The main drawback of using DCGAN is while converting the descriptions, the model parameters will never converge. Moreover, the generator of the GAN will translate only a few samples, and it is quite sensitive to hyperparameters. So, to overcome these disadvantages of DCGAN, the conditional GAN (CGAN) came into existence, where we can add some parameters for labeling the inputs in both generator and discriminator to classify the input-text correctly (T. Salimans et al.(2016)).

Generative Adversarial Networks are comprised of a generator(G) and discriminator(D), which work parallelly by a competitive goal (T. Salimans et al.(2016)). The generator is designed in such a way that it keeps on generating the sample tests towards the original data dispersal to bypass or dope the discriminator. Whereas the discriminator is designed in such a way that it always tries to identify real data samples over the generated fake samples. We are interested to work on translating the single-sentence text into its equivalent image pixels. For example: *"A white Bird with a black crown and a yellow peak"*. GANs have numerous applications in the real world like photo editing, image quality enhancement, computer-assisted design, etc. But they failed to spawn the high-resolution images using the text descriptions as shown in Figure 1.

14 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/text-to-image-synthesis-using-multistage-stack-gan/318066

Related Content

Blockchain and IoT-Based Diary Supply Chain Management System for Sri Lanka
K. Pubudu Nuwnthika Jayasenaand Poddivila Marage Nimasha Ruwandi Madhunamali (2021). *Blockchain and AI Technology in the Industrial Internet of Things (pp. 246-273).*
www.irma-international.org/chapter/blockchain-and-iot-based-diary-supply-chain-management-system-for-sri-lanka/277329

The Synthesis Between Artificial Intelligence and Editing Stories of the Future
Serhat Erdem (2025). *Transforming Cinema with Artificial Intelligence (pp. 231-250).*
www.irma-international.org/chapter/the-synthesis-between-artificial-intelligence-and-editing-stories-of-the-future/365413

Bridge Crack Recognition Method Based on Yolov5 Neural Network Fused With Attention Mechanism
Yingjun Wu, Junfeng Shi, Wenxue Maand Bin Liu (2024). *International Journal of Intelligent Information Technologies (pp. 1-25).*
www.irma-international.org/article/bridge-crack-recognition-method-based-on-yolov5-neural-network-fused-with-attention-mechanism/361575

Machine Learning-Based Demand Forecasting in Supply Chains
Real Carbonneau, Rustam Vahidovand Kevin Laframboise (2007). *International Journal of Intelligent Information Technologies (pp. 40-57).*
www.irma-international.org/article/machine-learning-based-demand-forecasting/2426

The Fringe Benefit of Industry 4.0 and Industry 5.0 on the Educational Sector: A Comprehensive Bibliometric Review
Vihas Vijayand Divya Lakshmi J. (2025). *Generative AI for Business Analytics and Strategic Decision Making in Service Industry (pp. 359-394).*
www.irma-international.org/chapter/the-fringe-benefit-of-industry-40-and-industry-50-on-the-educational-sector/368895