Z-GAN

FOR ENHANCING GENERATIVE ADVERSARIAL NETWORK-BASED T2I SYNTHESIS USING TRANSFORMER MODEL

E.M. Yenuli Boyagoda

A dissertation submitted in partial fulfillment of the requirement for
Bachelor of Science (Honours) degree in Computer Science

School of Computing

Informatics Institute of Technology, Sri Lanka
in collaboration with
University of Westminster, UK

2023

ABSTRACT

This research thesis focuses on the task of using Generative Adversarial Networks (GANs) to produce realistic images from text descriptions. The primary goal of this work is to address the research gap in the field of text-to-image synthesis, which is the inability of existing models to generate intact objects that are consistent with the given textual description, which causes the synthetic images to deviate from reality.

This objective is accomplished by the suggested method, which combines a pre-trained SBERT (Sentence-Bidirectional Encoder Representations from Transformers) model with a conditional GAN architecture to produce high-quality images that are semantically accurate to the input text description. The proposed method utilizes a GAN model, which uses supervised learning techniques to give useful feedback to train the generator and improve the consistency of the generated images. The thesis presents an in-depth review of the proposed approach utilizing a variety of evaluation metrics. Both quantitative and qualitative evaluations are used in the evaluation to show that the proposed methods effectively contribute to text-to-image synthesis models in terms of improvements in text information utilization and images with rich semantic consistency that are compatible with the input text descriptions.

Additionally, the thesis explores the drawbacks and difficulties of the proposed approach and demonstrates potential future research pathways to get around this issue. The findings of this study have significance for many applications, such as generative art, computer vision, enhancing human communication, and many more. Overall, this thesis advances the field of text-to-image synthesis by presenting a novel architecture to generate images with rich semantic consistency from textual descriptions while resolving the problems associated with producing synthetic images and text utilization in GAN models.

Keywords: Text-to-Image Generation, Generative Adversarial Networks, Computer Vision, Neural Networks, Natural Language Processing

Subject Descriptors:

- Software and its engineering → Software creation and management → Software development process management → Software development methods
- Computing methodologies → Artificial intelligence → Computer vision
- Computing methodologies → Machine learning → Machine learning approaches → Neural networks