様式第１号その１（第５条第１項の規定による場合）

# 学 位 授 与 申 請 書

2021 年　2 月 10 日

大阪府立大学長　様

大学院　　工学研究科　　電気・情報系専攻

氏　名　置名 一元

大阪府立大学学位規程第５条第１項の規定により

修士（工 学）の学位の授与を申請します。

（注意）

１．この申請書は、２通提出すること。

右肩の日付けは西暦年表示。

（Ａ４）

# 敵対的生成ネットワークによる文からの画像生成の改良手法の提案

# A Proposal of a Method to Improve the Image Generation from Textual data using Generative Adversarial Networks


| 分　野　名 | 知能情報工学分野 | 氏　名 | 置名　一元 |
|---|---|---|---|
| Department | Computer Science and Intelligent Systems | Name | Ichika OKINA |


Creation is a high-level intellectual activity, and it is unique to human beings. In recent years, a research field of creation by Artificial Intelligence (AI) has been extended with the remarkable development of AI and machine learning techniques such as deep learning. And in computer engineering, its importance of understanding and creation by AI has become greater. Nowadays, in this field, research on automatic generation using multimodal information in multiple fields in a complex manner has begun to be actively conducted. Attentional Generative Adversarial Networks (AttnGAN) is one of the text-to-image generation methods based on the deep learning model.

AttnGAN allows attention-driven, multi-stage refinement for a fine-grained text-to-image generation. With a novel attentional generative network, the AttnGAN was been able to synthesize fine-grained details at different subregions of the image by paying attention to the relevant words in the natural language texts. Besides, Deep Attentional Multimodal Similarity Model (DAMSM) was proposed to pre-compute a fine-grained image-text matching loss for the training phase of the generator.

However, important language information such as word meaning, syntax, and grammar was not considered at all in

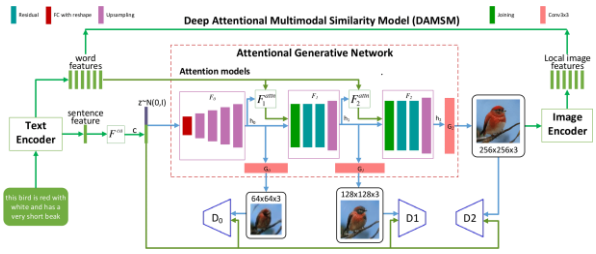


**Figure 1. The architecture of AttnGAN**

AttnGAN. If it considers the cultural background and social context contained with these kinds of language information, the performance of AttnGAN and DAMSM will improve.

In this paper, I propose a method to improve a model. In the method, (1) feature vectors of the distributed representation of words based on large scale language data are inputted to the text encoder. And (2) to compensate for the shortness of image data of DAMSM, DAMSM and the generator of AttnGAN learn synchronously to generate more good images. Moreover, to reduce the instability of text-to-image, two distance criteria, Fréchet Inception Distance (FID) and Sliced Wasserstein Distance (SWD) are adopted between the real image and the generated image into the loss function. These distances will suppress the phenomenon that image far from the real image is unintentionally generated due to some influential words in the input text. The effectiveness of the proposed method was confirmed by the experimental results.