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）

単語の分散表現を用いる DAMSM と GAN との同期的学習を導入した AttnGAN モデルの提案

Proposal of AttnGAN model that introduces synchronous learning between DAMSM and GAN using distributed representation of words

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 high-level intellectual activity unique to human beings. In recent years, with the remarkable development of artificial intelligence (AI), research field has extended to the field of creation by AI. Therefore, attempts to understand creative works by AI, and automatic generation have become interesting and significant in terms of computer engineering. Recently, not only research on single information in the fields of language or image, but also research on multimodal automatic generation that handles information in multiple fields in a complex manner has begun to be actively conducted. Attentional Generative Adversarial Networks (AttnGAN) is one of the methods that is attracting attention because it has characteristics in both creative and multimedia fields.

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

However, word meaning, syntax and grammar are not considered at all in AttnGAN. Considering the cultural background and social context of the word based on the method of natural language processing will lead to the improvement of the attention by DAMSM and the performance of AttnGAN as a whole.

In this paper, we propose a model that uses the feature vector obtained by the distributed representation of words as the input of the text encoder. Furthermore, for the purpose of compensating for the shortness of image data of DAMSM, we propose a model that can utilize the information of the generated image obtained from the generator by enabling synchronous learning with GAN not only pre-learning. Moreover, in order to reduce the instability of text-to-image, we incorporate two distances, Frèchet Inception Distance (FID) and Sliced Wasserstein Distance (SWD) between the real image and the generated image into the loss function. This suppresses the phenomenon that a image far from the real image is unintentionally generated due to some influential words in the input text.