Text-to-Image를 위한 아동 손그림 학습 모델 생성 연구

이은채[O], 문미경[*]
[O]동서대학교 소프트웨어학과,
[*]동서대학교 소프트웨어학과
e-mail: tldjs3651@naver.com[O], mkmoon@dongseo.ac.kr[*]

Study on Generation of Children's Hand Drawing Learning Model for Text-to-Image

Eunchae Lee[O], Mikyeong Moon[*]
[O]Dept. of Software, Dongseo University,
[*]Dept. of Software, Dongseo University

● 요 약 ●

인공지능 기술은 점차 빠른 속도로 발전되며 응용 분야가 확대되어 창작 산업에서의 역할도 커져 예술, 영화 및 기타 창조적인 산업에도 영향을 주고 있다. 이러한 인공지능 기술을 이용하여 텍스트로 설명하면 다양한 스타일의 이미지를 생성해내는 기술이 있지만 아동이 직접 그린 손그림 스타일의 그림을 생성하지는 못한다. 본 논문에서는 아동 손그림 데이터를 통해 Text-to-Image를 학습시켜 새로운 학습 모델을 생성하는 과정에 대해서 기술한다. 이 연구를 통해 생성된 픽셀을 결합하여 텍스트를 기반으로 하나의 아동 손그림을 만들 수 있을 것으로 기대한다.

키워드: Text-to-Image, 아동 손그림(Children's Hand Drawing), 학습(Learning), 생성(Generation)

I. Introduction

인공지능 기술을 이용한 시스템이 실제 생활에 활용할 수 있게 됨에 따라 인공지능에 대한 연구가 우리나라를 포함해 세계 여러 나라에서 경제적으로 이루어지고 있다. 사람 수준을 초월한 인공지능 기술은 창작 산업에 활용되는 단계로 진입하고 있다. 이러한 성과와 함께 이미지에서 텍스트를 인식 또는 추출하는 이미지 인식 기술이 발전하고 있다. 대표적으로 이미지 속 객체와 배경을 식별하거나 비디오를 분석하여 다양한 활동을 인식하는 객체 인식과 이미지와 영상 속 글자를 인식하는 OCR(Optical Character Recognition) 기술이 있다 [1]. 최근에는 Image-to-Text 기술의 반대인 Text-to-Image 연구가 활발히 진행되고 있다. Text-to-Image는 텍스트를 기반으로 해당 텍스트와 유사한 관계를 가지는 이미지를 생성하는 기술이다. 생성되는 그림 스타일은 실사에 가까운 이미지, 유명 화가의 그림체 또는 일러스트, 이모티콘 등 여러 가지 스타일로 사실적인 이미지를 생성한다. 하지만 이러한 Text-to-Image는 생성해내는 그림 스타일에 제한이 있어 정해놓은 그림체로만 이미지를 생성하여 아동의 손그림 스타일로 그려진 이미지를 생성해내지는 못한다. 본 논문에서는 Text-to-Image를 아동 손그림 데이터로 학습시켜 새로운 학습 모델을 생성하는 과정에 대해서 기술한다.

II. Preliminaries

OpenAI에서 개발한 DALL-E는 텍스트 기반으로 이미지를 생성하는 인공지능 프로그램이다. DALL-E는 텍스트와 이미지 쌍에 대해 학습한 뒤 텍스트 설명에 맞추어 이미지를 생성한다. 입력된 텍스트를 통해 생성되는 이미지는 현실에 존재하지 않는 이미지뿐만 아니라 사실적인 이미지를 만들어낸다. 또한, 랜더링 작업을 걸쳐 사진 이미지뿐만 아니라 일러스트, 이모티콘 등 다양한 삽화를 그려낸다 [2]. 그러나 DALL-E는 실사에 가까운 이미지와 일러스트, 스케치 등 다양한 스타일의 그림 이미지를 생성하지만 아동이 직접 그린 손그림으로 표현된 이미지는 생성하지 못하는 한계가 있다.

III. The Proposed Scheme

3.1 데이터 수집

본 연구를 진행하기 위한 데이터셋으로 그림에 대한 설명이 가장 적합한 데이터인 그림일기 이미지 파일과 그에 맞는 텍스트 파일을 같이 수집하였다. 이후 이미지 학습에 영향을 줄 수 있는 그림의 희미한 부분이나 글자가 적힌 부분은 제거 작업을 통하여 양질의

데이터로 정제하는 과정을 진행하였다. Fig. 1은 본 논문에서 사용하는 학습 데이터인 아동 손그림 데이터와 텍스트 설명의 일부이다.

Fig. 1. Preprocessed Learning Data

3.2 Stage One

Stage One 단계에서는 정제된 데이터 중 이미지에 대해서만 학습하여 이미지 토큰으로 압축하는 과정을 거친다. 이미지를 토큰화 시키는 이유는 원본 이미지를 그대로 학습시키면 메모리 사용이 증가되어 모델의 크기를 과도하게 증가시켜 비실용적이기 때문이다. Fig. 2는 원본 이미지를 이미지 토큰으로 압축하는 과정이다.

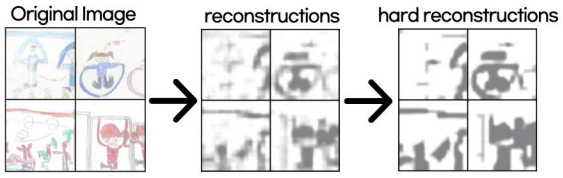

Fig. 2. Stage One Learning Process

3.3 Stage Two

Stage Two 단계에서는 텍스트와 이미지를 동시에 학습을 진행한다. 텍스트와 이미지 데이터가 짝으로 주어지면 주어진 텍스트는 소문자로 인코딩하는 과정을 거치고 이미지는 이미지 토큰으로 인코딩된다. 생성된 이미지 토큰과 텍스트 토큰은 연관 관계를 지어 연결한 후 최종적으로 하나의 토큰 데이터 스트림으로 생성되어 모델에 입력된다. 이후 하나의 텍스트와 앞서 생성된 토큰 데이터 스트림이 함께 Input 값으로 들어가면 텍스트 토큰과 텍스트 토큰, 이미지 토큰과 이미지 토큰, 텍스트 토큰과 이미지 토큰의 모든 가능성을 고려하고 고려한 값을 통해 다음 예측되는 픽셀을 생성한다. Fig. 3은 텍스트와 이미지 그리고 압축한 이미지 토큰이 입력되면 새로운 픽셀이 생성되는 과정이다.

Fig. 3. Stage Two Learning Process

IV. Conclusions

본 논문에서는 아동 손그림 데이터를 통해 Text-to -Image를 학습시켜 새로운 학습 모델을 생성하는 과정에 대해서 제안하였다. 텍스트를 기반으로 이미지를 생성하는 과정에서 이미지를 픽셀 단위로 토큰화 시키는 과정과 결과를 확인하였다. 향후 연구에서는 앞서 만들어진 픽셀들을 결합하여 하나의 이미지를 생성하는 연구를 진행할 것이다.

ACKNOWLEDGMENT

본 연구는 2022년 과학기술정보통신부 및 정보통신기획평가원의 SW중심대학사업의 연구결과로 수행되었음(2019-0-01817)

REFERENCES

[1] 이주열, "인공지능 이미지 인식 기술 동향" TTA Journal, Vol. 187, No. 5, pp. 44-51, JAN/FEB 2020.

[2] DALL-E: Creating Images from Text-OpenAI, *https://openai.com/blog/dall-e/*