Transformer 기반 OCR 모델을 이용한 강인한 캡차 영상 인식 알고리즘

서준혁*, 장성진*, 김민준*, 윤현주**

Robust Captcha Image Recognition Algorithm Using Transformer-Based OCR Model

Junhyeok Seo*, Seongjin Jang*, Minjun Kim*, and Hyeonju Yoon**

요 약

본 논문에서는 Transformer 기반 OCR(광학 문자 인식) 모델인 TrOCR 모델을 이용하여 다양한 노이즈가 존재하는 캡차 영상에 강인하게 대응할 수 있는 OCR 알고리즘을 제안한다. 광학 문자 인식의 성능을 높이기 위해 기존 모델들의 복잡한 전, 후처리 과정을 개선한 End-to-End 방식의 TrOCR 모델을 사용하였고 이렇게 구축된 최종 알고리즘의 성능 평가에서 평균 CER(Character Error Rate)을 측정한 결과 약 0.08의 수치를 보였다. 객관적인 비교 분석을 위해 널리 사용되는 Tesseract OCR, EasyOCR 모델을 동일한 환경에서 시험하여 그 성능을 측정한 결과, 시험용 영상 400장에 대해 두 모델 모두 0%의 정확도를 보인 반면, 본 논문에서 구축된 모델은 67.18%의 상대적으로 높은 정확도를 보였다.

Abstract

In this paper, we use the TrOCR model, which is a Transformer-based OCR (optical character recognition) model, to propose an OCR algorithm that can robustly handle captured images with various types of noise. In order to improve the performance of optical character recognition, we used an end-to-end TrOCR model that improved the complicated pre- and post-processing processes of the conventional model, and evaluated the performance of the final algorithm constructed in this way by using the average CER. (Character Error Rate) measurement results showed a value of approximately 0.08. We tested the Tesseract OCR and EasyOCR models, which are widely used for objective comparative analysis, in the same environment and measured their performance, and both models showed an accuracy of 0% for 400 test videos. In contrast, the model constructed in this paper showed a relatively high accuracy of 67.18%.

Key words
Transformer, OCR, Captcha,

* 금오공과대학교 컴퓨터공학과 ssam2s@kumoh.ac.kr, j267175@kumoh.ac.kr, joun2301@kumoh.ac.kr,
** 금오공과대학교 컴퓨터공학과 juyoon@kumoh.ac.kr

Ⅰ. 서 론

오늘날, 인터넷의 성장과 함께 웹 사이트와 온라인에서 제공되는 서비스들은 다양한 형태의 공격으로부터 자신을 보호해야하는 필요성이 더욱 증가하였다. 이러한 공격들을 방어하기 위한 많은 방법 중 하나로, 캡차(CAPTCHA) 시스템이 널리 채택되어왔다.

캡차는 웹 사이트 사용자가 실제 인간인지 자동화된 프로그램인지를 구별하기 위해 설계된 시스템으로, 현재 가장 널리 사용되는 전통적인 캡차 시스템은 특정 이미지에 존재하는 문자를 사용자로부터 입력 받는 방식을 통해 작동된다.

그러나, OCR(Optical Character Recognition; 광학 문자 인식) 기술의 발전과 함께 이러한 전통적인 캡차 시스템이 제 기능을 못할 수 있음이 드러나고 있다.[1]

이에 따라, 본 논문은 최근 딥러닝 기반 영상처리 분야에서 월등히 높은 성능을 보이고 있는 Transformer 구조를 가진 OCR 모델을 사용하여 다양한 종류의 노이즈가 존재하는 캡차 영상을 강인하게 파훼할 수 있는 시스템을 제시하고, 흔히 사용되는 타 OCR 모델들과의 객관적인 성능 비교 및 평가를 통해 최신 기술로부터 기존 캡차 시스템의 취약점을 보완하는데 기여하고자 한다.

Ⅱ. 캡차 영상 인식 시스템 개발

2.1 데이터셋 전처리

본 논문에서는 딥러닝 기반 OCR 모델의 학습 및 시험을 위한 데이터로 그림 1과 같이 구글의 캐글(Kaggle)에서 제공하는 4가지 종류의 노이즈가 존재하는 캡차 영상을 각 1,000장씩 총 4,000장을 데이터셋으로 확보하였다.[2][3][4][5]

학습용 데이터셋과 시험용 데이터셋은 각 종류의 캡차 영상이 동일한 비율로 분포하도록 불균형을 고려하면서 원본 데이터셋 4,000장에서 9:1의 비율로 표 1과 같이 분리하였다.

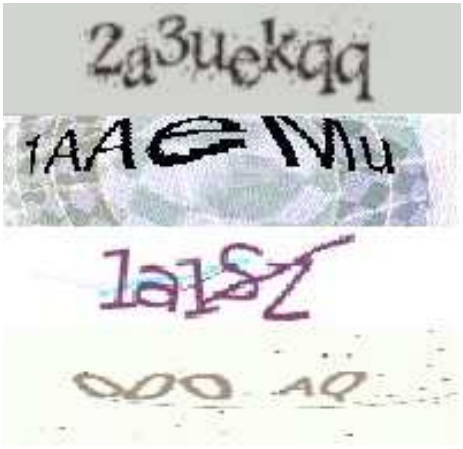

그림 1. 4가지 종류의 노이즈가 존재하는 캡차 영상 데이터셋 예시
Fig. 1. Example of a CAPTCHA image dataset with 4 types of noise

표 1. 실험 데이터셋
Table 1. Experimental datasets

노이즈 타입	학습용	시험용	총 개수
A	900	100	1,000
B	900	100	1,000
C	900	100	1,000
D	900	100	1,000

2.2 TrOCR 모델 학습

기존의 딥러닝 기반 OCR 모델들은 Encoder-Decoder 형태의 구조에 입력 이미지를 처리할 CNN(Convolutional Neural Network)과 출력 텍스트를 처리할 RNN(Recurrent Neural Network)을 접목시킨 것이 일반적이었다. 이러한 모델 구조는 데이터의 전, 후처리 과정이 복잡하여 연산량이 많다는 단점이 있다.

본 논문에서 사용한 모델은 TrOCR(Transformer-based Optical Character Recognition with Pre-trained Models)이다. 본 모델은 기존의 모델들과 동일한 Encoder-Decoder 형태를 띄고 있지만 ViT(Vision Transformer) 모델을 Encoder로, BERT(Bidirectional Encoder Representations from Transformers) 모델을 Decoder로 하여 구성되어있다.[6]

이러한 End-to-End 방식을 통해 기존 모델의 단점이었던 처리 과정의 복잡도를 줄이면서 가장 높은 성능을 보였음을 뜻하는 SOTA(State-of-the-Art)를 달성하였다.

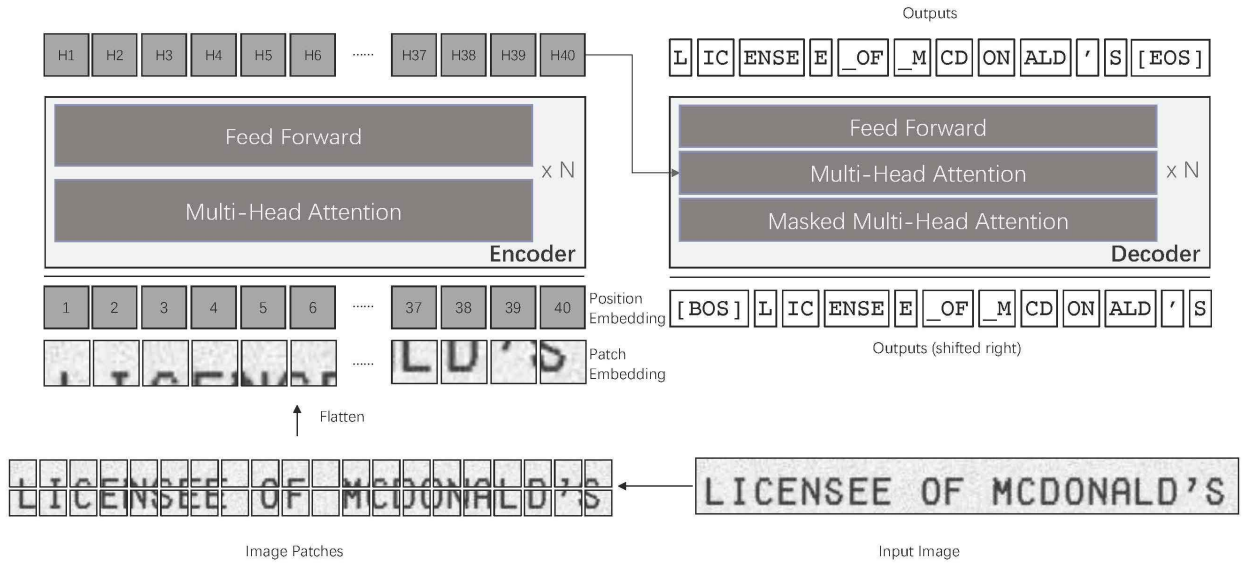

그림 2. TrOCR의 네트워크 구조
Fig. 2. The Architecture of TrOCR

본 논문에서는 TrOCR 모델 중 가장 복잡한 구조를 가짐과 동시에 가장 좋은 성능을 보이는 Large 버전을 사용한다. 표 1과 같이 학습용 데이터는 총 3,600장을 사용하였으며 시험용 데이터는 총 400장을 사용하였다. 실험 환경은 표 2와 같다.

표 2. 실험 환경
Table 2. Experimental environment

CPU	Intel Xeon 4214R
GPU	8 x GeForce RTX A6000
RAM	256GB
Framework	PyTorch 2.0.1
CUDA / CUDNN	11.7 / 8500

2.3 성능 평가

본 논문에서는 성능 평가 지표로 모델이 추론한 문자열과 정답 문자열 사이의 문자 오류 비율을 나타내는 CER(Character Error Rate)을 사용한다. 지표 계산식은 (1)과 같다. 이때 S, D, I, N은 각각 잘못 대체된 음절 수, 잘못 삭제된 음절 수, 잘못 추가된 음절 수, 정답 텍스트의 음절 수를 의미한다.

$$CER = \frac{S + D + I}{N} \quad (1)$$

이러한 지표를 기반으로 모델의 학습 과정에 대한 하이퍼 파라미터 튜닝을 진행하였으며, 본 논문에서 사용한 모델의 학습 결과, 그림 3과 같이 Training loss는 0.07, Validation loss는 0.84에 도달하였으며 CER은 0.08의 값을 보였다.

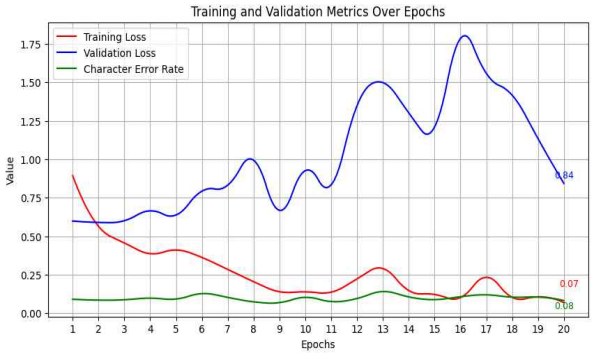

그림 3. 모델 학습 결과
Fig. 3. Training Result of Our Model

학습이 완료된 모델과 기존의 타 모델들과의 객관적인 성능 비교를 위해 동일한 시험용 데이터셋 400장을 널리 사용되는 OCR 모델인 구글의 Tesseract OCR[7]과 JaidedAI가 개발한 EasyOCR[8]에 적용하였다. 각 모델이 읽어낸 캡차 영상의 텍스트를 실제 정답과 비교하여 평균 CER과 정확도를 도출하였고 그 결과는 표 3과 같다.

표 3. 실험 결과
Table 3. Experimental results

Model	Mean CER	Accuracy
Ours	0.08	0.6718
Tesseract OCR	0.89	0
EasyOCR	13.44	0

3가지 모델의 비교에서, 본 논문에서 학습시켜 사용한 모델이 0.08의 평균 CER로 현저히 낮은 오차율을 보였으며, 67.18%의 가장 높은 정확도를 보였다. 이에 반해, 나머지 두 모델은 단 하나의 영상도 정확하게 식별하지 못하는 결과를 보였다.

그림 4와 같이 모델의 추론 결과를 시각화한 영상에서, 상대적으로 길고 노이즈가 심해 식별하기 어려운 캡차 영상도 각 종류별로 잘 식별해내는 것을 확인할 수 있었다.

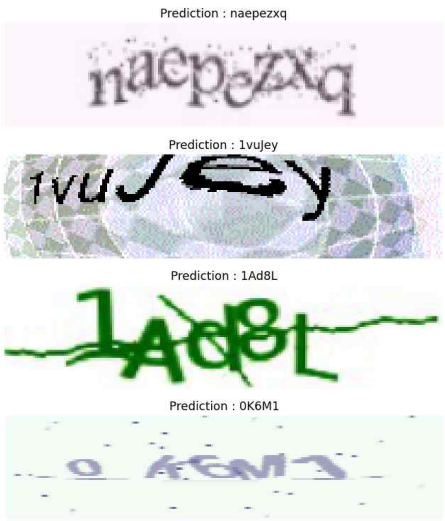

그림 4. 추론 결과 시각화 영상
Fig. 4. Visualization Image of Inference Results

III. 결 론

본 논문에서는 다양한 노이즈가 존재하는 캡차 영상을 강인하게 파훼하기 위해 Transformer 기반 OCR 모델인 TrOCR 모델에 4가지 종류의 노이즈가 존재하는 캡차 영상 3,600장을 학습시켰다. 학습이 끝난 모델의 성능 평가에서 시험용 데이터셋 400장에 대해 평균 CER은 0.08의 수치를 보였고, 오차 없이 정확히 식별한 결과에 대한 정확도를 측정한 결과 67.18%의 수치를 보였으며, 시각화한 결과에서도 양호한 식별 성능을 보였다.

또한, 널리 사용되는 OCR 모델인 Tesseract OCR과 EasyOCR을 이용하여 동일한 조건에서 시험을 진행하여 그 결과를 객관적으로 비교하였다. 평균 CER과 정확도를 비교한 결과 TrOCR 모델이 월등히 좋은 식별 성능을 보였다.

다만 숫자 '1'과 알파벳 'l', 숫자 '0'과 알파벳 'O'과 같이 유사한 모습의 문자에 대해 오식별을 하는 경우가 발생해 향후 연구에서 학습용 데이터셋의 크기를 늘리거나, 영상처리 기술을 전/후처리 과정에 적용한다면 문자 식별 성능을 향상시킬 수 있을 것으로 기대된다.

참 고 문 헌

[1] 권예슬, "보안기술 '캡차' 뚫는 인공지능 나왔다", 동아일보, 2017.10.27., https://www.donga.com/news/article/all/20171027/86977937/1

[2] Kaggle, CAPTCHA Images, (2018). https://www.kaggle.com/datasets/fournierp/captcha-version-2-images

[3] Kaggle, 66k Captchas Dataset, (2023). https://www.kaggle.com/datasets/jassoncarvalho/comprasnet-captchas

[4] Kaggle, CAPTCHA Dataset, (2020). https://www.kaggle.com/datasets/parsasam/captcha-dataset/

[5] Kaggle, Large Captcha Dataset, (2021). https://www.kaggle.com/datasets/akashguna/large-captcha-dataset

[6] Minghao Li and Tengchao Lv and Lei Cui and Yijuan Lu and Dinei Florencio and Cha Zhang and Zhoujun Li and Furu Wei. (2021). TrOCR: Transformer-based Optical Character Recognition with Pre-trained Models. arXiv:2109.10282 [cs.CL]

[7] Tesseract OCR (2021) Repository [Source Code]. https://github.com/tesseract-ocr/tesseract

[8] EasyOCR (2022) Repository [Source Code]. https://github.com/JaidedAI/EasyOCR