

CNN 기반 한국어 유아 손글씨 OCR 모델 연구

곽수민(응용통계학과), 김우엽(기계공학부)

마아이

2023 CUA이 중앙대학교 인공지능 학회 컨퍼런스
Proceeding of 2023 Chung-Ang University Artificial Intelligence Conference

Abstract (나눔스퀘어 48pt)

본 연구는 차세대 에듀테크 서비스의 핵심 기술로 자리 잡은 OCR 기술을 유아 손글씨 인식에 특화하여 연구하였다. 구현한 CNN 기반 모델은 Test Accuracy 85.55%의 성능을 기록하였으며 차후 교육 업계와 연동한다면 아이들의 효율적인 학습을 위한 교두보의 역할을 수행할 수 있을 것이다.

Introduction

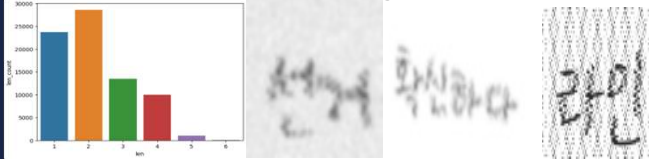
본 연구는 OCR 기술이 유아 교육에 활용될 수 있도록 유아의 손글씨 데이터 인식에 최적화된 OCR 모델을 개발하고자 시작하였다. 교원그룹에서 제공한 유아 손글씨 데이터를 활용하였다.

Materials and Methods

1) 데이터 소개 및 EDA

데이터는 폰트 손글씨 학습 이미지 데이터 76,888개, 폰트 손글씨 평가 이미지 데이터 74,121개로 이루어져 있다.

글자수의 불균형과 흐릿하고 불규칙한 데이터들을 고려하여 Cut-Mix와 STRaug 기법을 적용하였다.



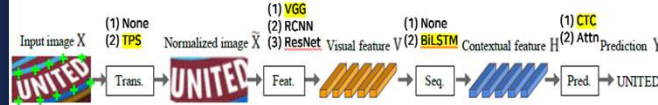
2) 데이터 Augmentation

Cut-Mix 기법은 두가지 이미지를 합쳐 글자 길이가 긴 데이터를 생성하는 기법이다. STRaug는 STR 특화 augmentation 모음이다.



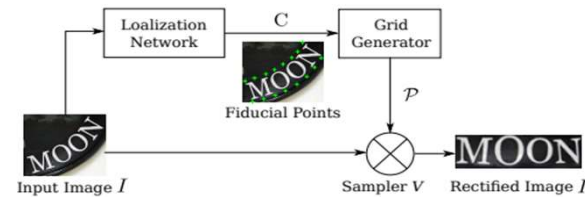
3) 모델링

본 연구에서 사용한 OCR 모델은 Transformation, Feature Extraction, Sequence Modeling, Prediction 모듈로 구성하였다. 위 모듈들은 네이버 Clova AI 팀에서 벤치마크하기 위해 사용했던 모듈 구조를 바탕으로 하였다.



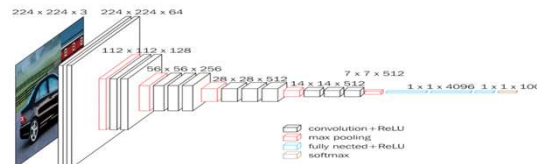
3-1) Transformation : TPS

불규칙한 텍스트 이미지를 모델에게 보다 더 잘 인식할 수 있도록 하기 위한 과정이다.



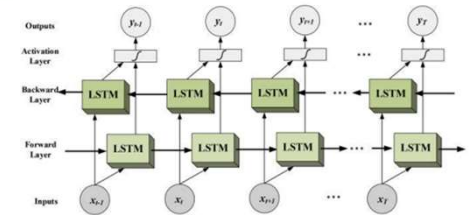
3-2) Feature Extraction : VGG

컴퓨터비전 분야의 대표적인 모델로서 3x3 사이즈의 아주 작은 필터를 사용하여 신경망 깊이에 초점을 두었다.



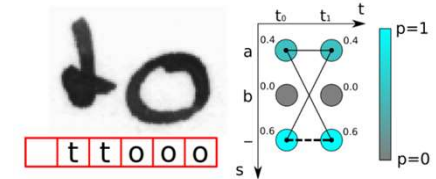
3-3) Sequence Modeling : BiLSTM

BiLSTM 모델은 기존 LSTM의 한계인 앞의 문장의 정보만 참고하는 것을 보완하여 이후의 정보 또한 이용할 수 있게 해준다.



3-4) Prediction : CTC

모델 출력의 Sequence가 이미지에 걸쳐지는 문제로 기대한 것과 다른 결과가 나오는 문제를 해결하기 위해 사용된다.



Conclusion

	Valid accuracy	Valid Loss	Test Accuracy
TPS-VGG-BiLSTM-CTC	0.9697	0.0988	0.8186
TPS-RCNN-BiLSTM-CTC	0.9775	0.0741	0.8198
TPS-ResNet-BiLSTM-CTC	0.9804	0.0713	0.7962
TPS-ResNet-None-CTC	0.9799	0.0783	0.6159
None-VGG-BiLSTM-CTC	0.9239	0.2732	0.7601

기본 데이터로 여러 가지 모듈을 테스트한 후 최적의 성능을 나타낸 2가지 모델을 선정 하였고 두 모델에 3배 가량 증식한 데이터셋을 적용하였다. 결과적으로 TPS-VGG-BiLSTM-CTC 모델에서 0.8555의 최종 스코어를 얻을 수 있었다. 제한적인 시간과 학습 여건 등의 문제로 조합 가능한 24개의 모듈 조합을 전부 비교하지 못했다는 점과 실제 최종 학습에 사용한 데이터와 비교시에 사용한 기본 데이터가 달라서 최적의 모델을 찾는 방법에 한계가 있었다는 점이 아쉬움으로 남는다.