

온라인 한글 필기 인식 앱 개발

Development of online Hangul-handwritten recognition App

2022 한국정보기술학회 추계종합학술대회

이소연 서경대학교 소프트웨어학과

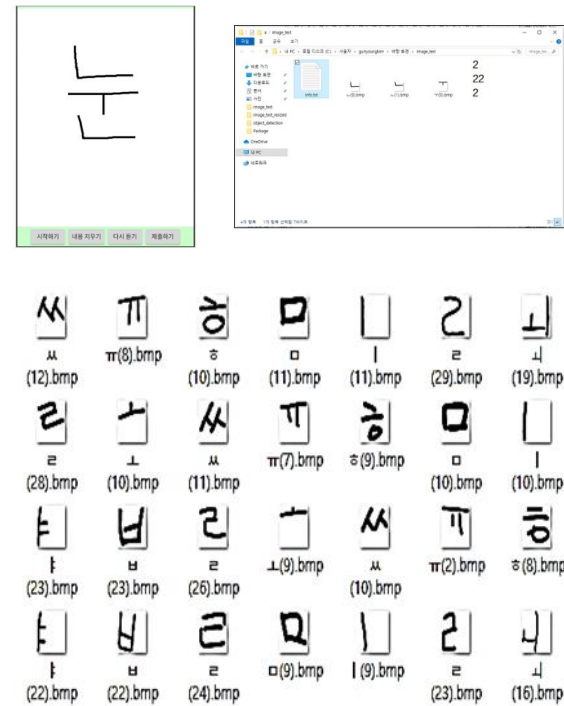


SINCE 1947
서경대학교
SEOKYEONG UNIVERSITY

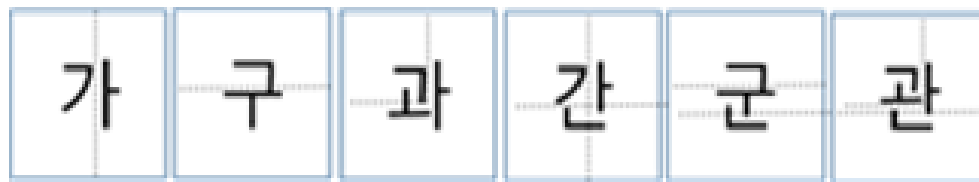
필기 인식의 성능을 높이기 위해서는, 대개 내부 구조를 복잡하게 하거나 계층을 깊게 한다.
 이는 곧 인식 모델이 복잡해지거나 학습 과정에서 많은 수의 인자가 필요하여 많은 부하를 발생시킨다.
 이에 본 연구에서는 모델 단순화를 위해 **전처리 기반 온라인 한글 필기 인식 앱**을 개발하고자 한다.

연구의 개요

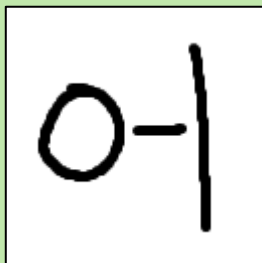
- 필기를 위한 기계학습 기반 모델을 만들기 위해서는 많은 수의 학습 데이터가 필요하다.
- 학습 데이터 획득을 위해 온라인 방식과 오프라인 방식이 병행되었다.
- 온라인 방식이란, 여러 사용자로부터 앱으로 획득한 학습 데이터에 대한 사전 분석을 통해 각 초성/중성/종성에 해당하는 자음/모음을 추출하는 방식이다.
- 오프라인 방식은 한국지능정보사회진흥원에서 제공하는 한글 필기 이미지에서 자음/모음을 이미지 편집 프로그램을 이용하여 구분하여 저장하였다.



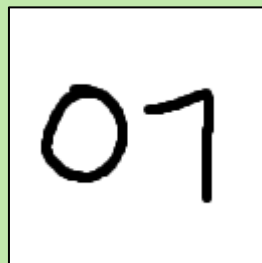
모음의 위치 및 받침(종성)의 유무에 따라 6가지로 구분할 수 있다.



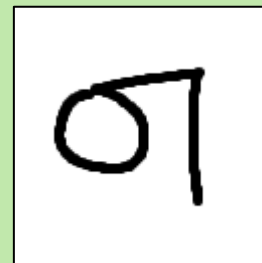
Case 1 ~ 6



흘려 쓰지 않음



초성/중성/종성 각각 흘려 씀

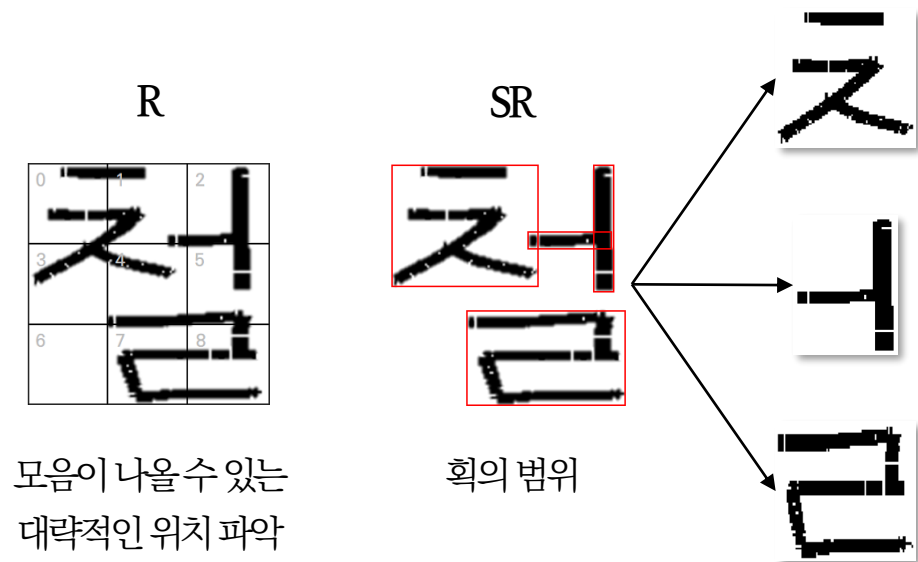


초성/중성/종성 이어 흘려 씀

한글 필기를 위와 같이 크게 세 가지로 구분한다.

흘려 쓰지 않은 경우

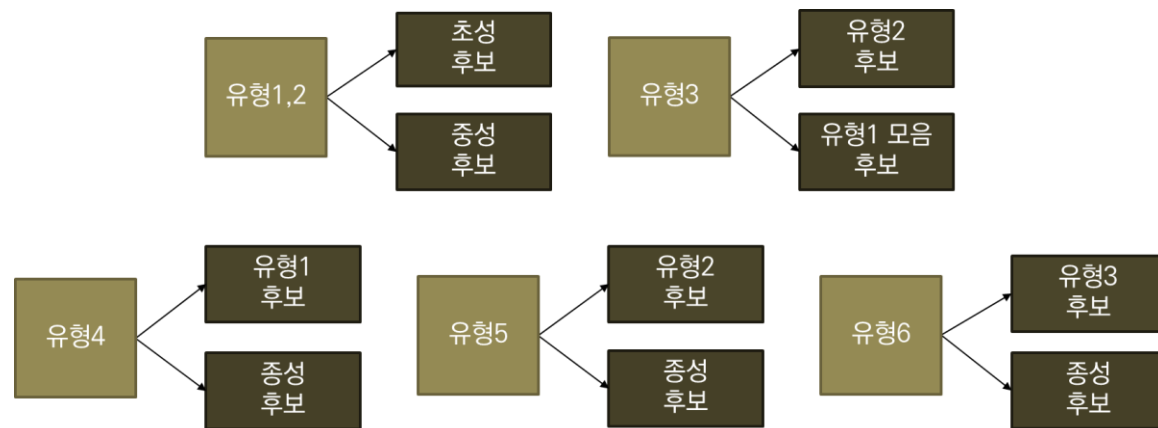
각 유형을 담당하는 프로세스가 전처리를 진행하여,
모음(중성)과 자음(초성 및 종성) 후보군을 결정한 후 인식 모델이 인식하도록 한다.



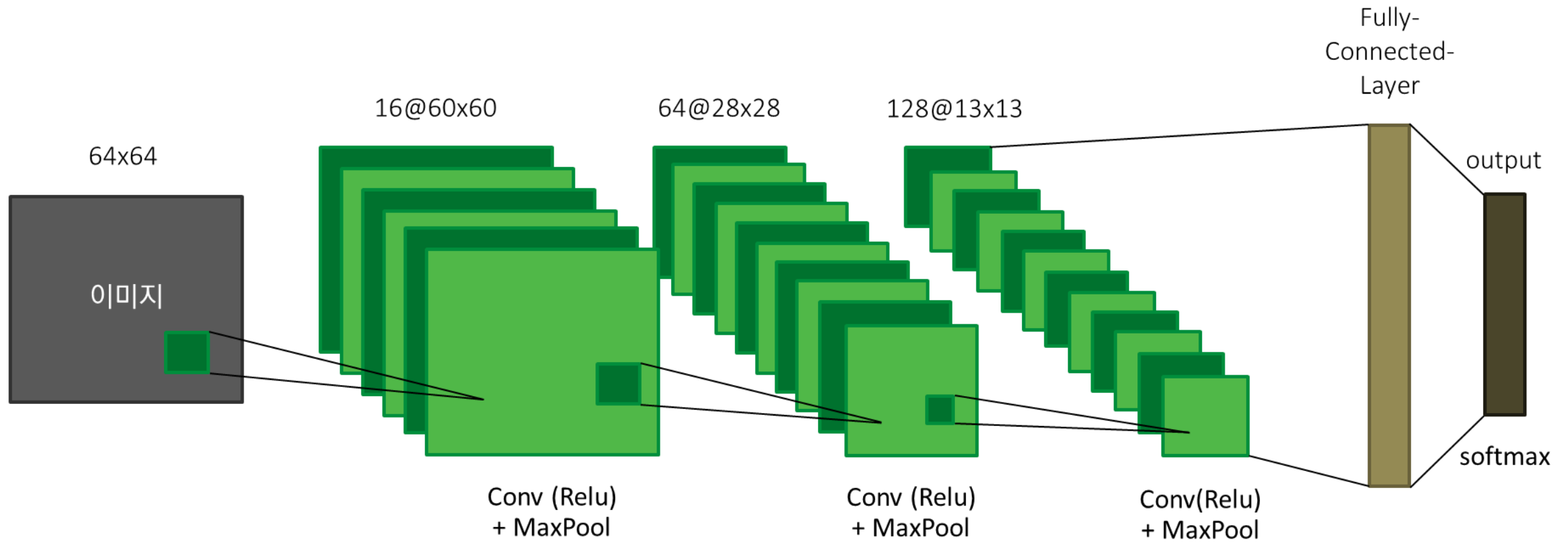
각 프로세스는 획 정보로부터 각 획이 직선인지 판정한 후,
R과 SR을 이용하여 중성 추출 후 초성과 종성을 분리한다.

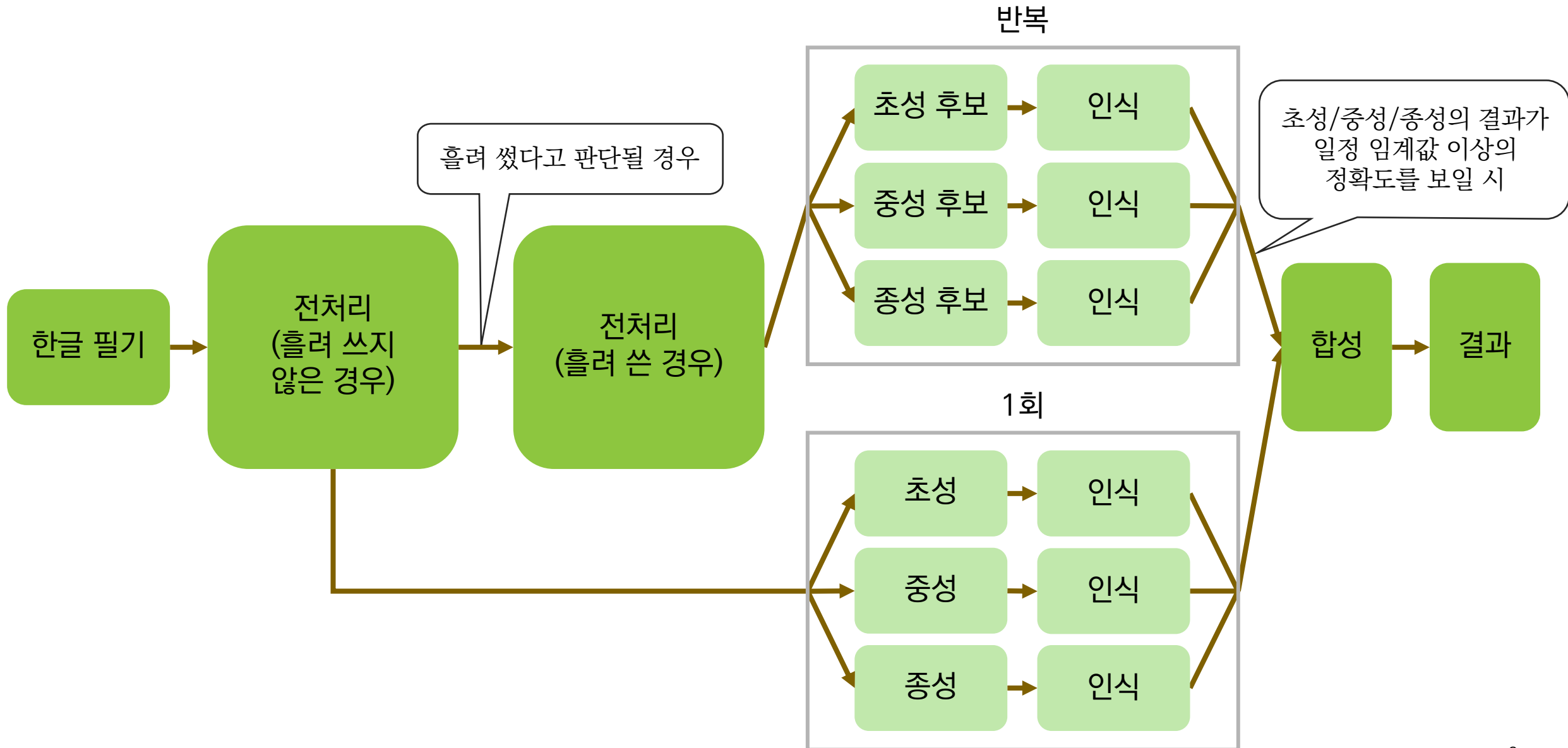
흘려 쓴 경우

각 유형을 담당하는 프로세스는 전처리를 진행하여,
유형에 따른 초성/중성/종성 후보군을 결정한 후 인식 모델이 인식하도록 한다.



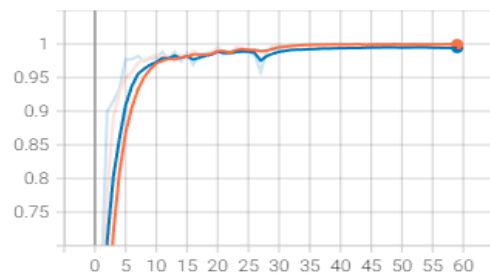
인식 진행 중 조합한 획 데이터가 자음 또는 모음이 아니거나 해당 유형이 아닐 시
자음, 모음 및 유형 인식 모델에서 조합 없음이라 판별 후
해당 조합을 후보군에서 탈락시키는 과정을 반복해 초성/중성/종성을 분리한다.





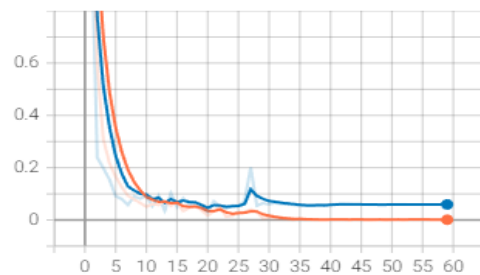
epoch_accuracy

epoch_accuracy



epoch_loss

epoch_loss

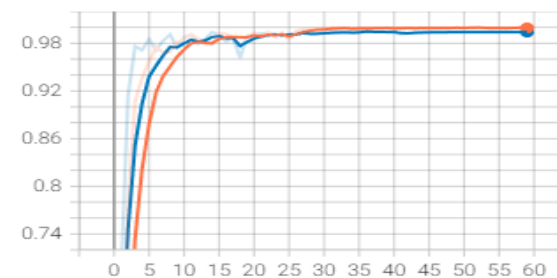


[전처리 정확도]

학습 데이터 셋: 97.21% 테스트 데이터 셋: 97.02%

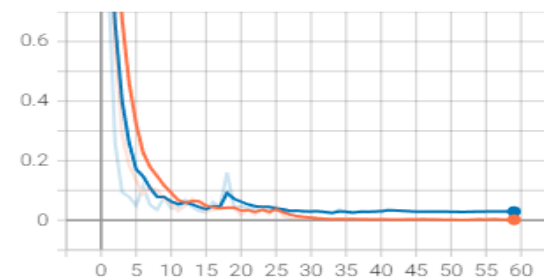
epoch_accuracy

epoch_accuracy



epoch_loss

epoch_loss



[인식 정확도]

학습 데이터 셋: 98.04% 테스트 데이터 셋: 97.94%

연구의 결론

- 온라인 한글 필기 인식을 위한 전처리 알고리즘과 일반적인 인식 모델을 이용한 인식을 담당하는 앱 개발을 하였다.
- 한글의 창제 원리에 기반한 전처리 알고리즘을 개발하여, 단순하고 경량 인식 모델임에도 불구하고 상당히 높은 인식률을 얻을 수 있었다.
- 앞으로 이번에 개발된 앱 기능을 확장하여, 더욱 난해한 이어쓰기 된 글자 인식뿐만 아니라 오프라인 필기 인식을 위한 전처리 알고리즘을 개발하고자 한다.