

I.Introduction

- ❖ 한글 OCR 성능을 높이기 위해 딥러닝 모델을 활용하여 문자인식 부분을 개선하고자 함
 - 폰트와 사전데이터를 사용해 딥러닝 모델 학습을 위한 한글 문장 이미지 데이터를 직접 생성, 이를 활용해서 한글 문장의 OCR 성능을 높일 다양한 모델 조합들에 대한 실험
 - 딥러닝 모델은 변환, 추출, 시퀀스, 예측 모듈 각각 $2 \times 3 \times 2 \times 2 = 24$ 가지 모델 조합구성
 - 이전 한글 OCR 연구와 비교해 적용 범위를 글자 단위에서 문장 단위로 확장, 실제 문서 이미지에서 자주 발견되는 유형의 데이터를 사용해 애플리케이션 적용 가능성을 높이고자 함
 - OCR 기술이 어려운 점 3가지
 - 첫째, 문서 이미지 속 글자들은 정형화되어 있지만, 손글씨 및 서명은 비정형화되어 있어 분별이 어려움
 - 둘째, 배경이 복잡한 경우 배경과 문자의 구분이 힘들
 - 셋째, 다양한 간섭요소 -노이즈, 왜곡, 글자 사이 밀도, 저해상도-로 인해 식별 어려움
 - 한글에 대한 OCR이 더 어려운 이유는 영문과 비교해 분류해야 할 글자 수가 실험기준으로 (26 → 900) 35배 정도 차이가 나 정확도를 높이는 데 어려움

II.Proposed Method

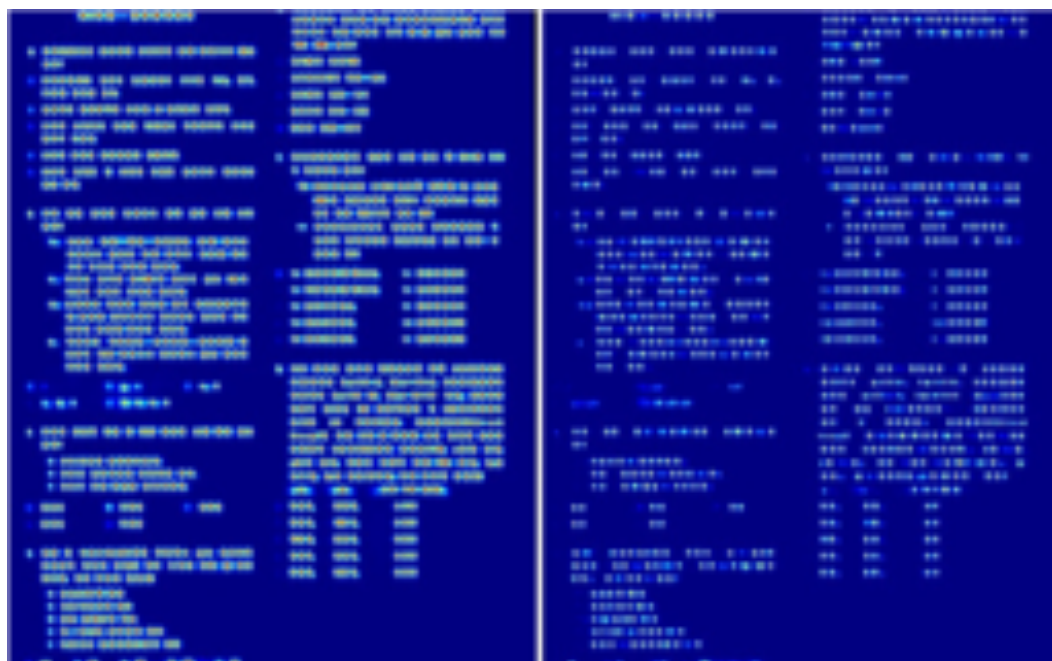
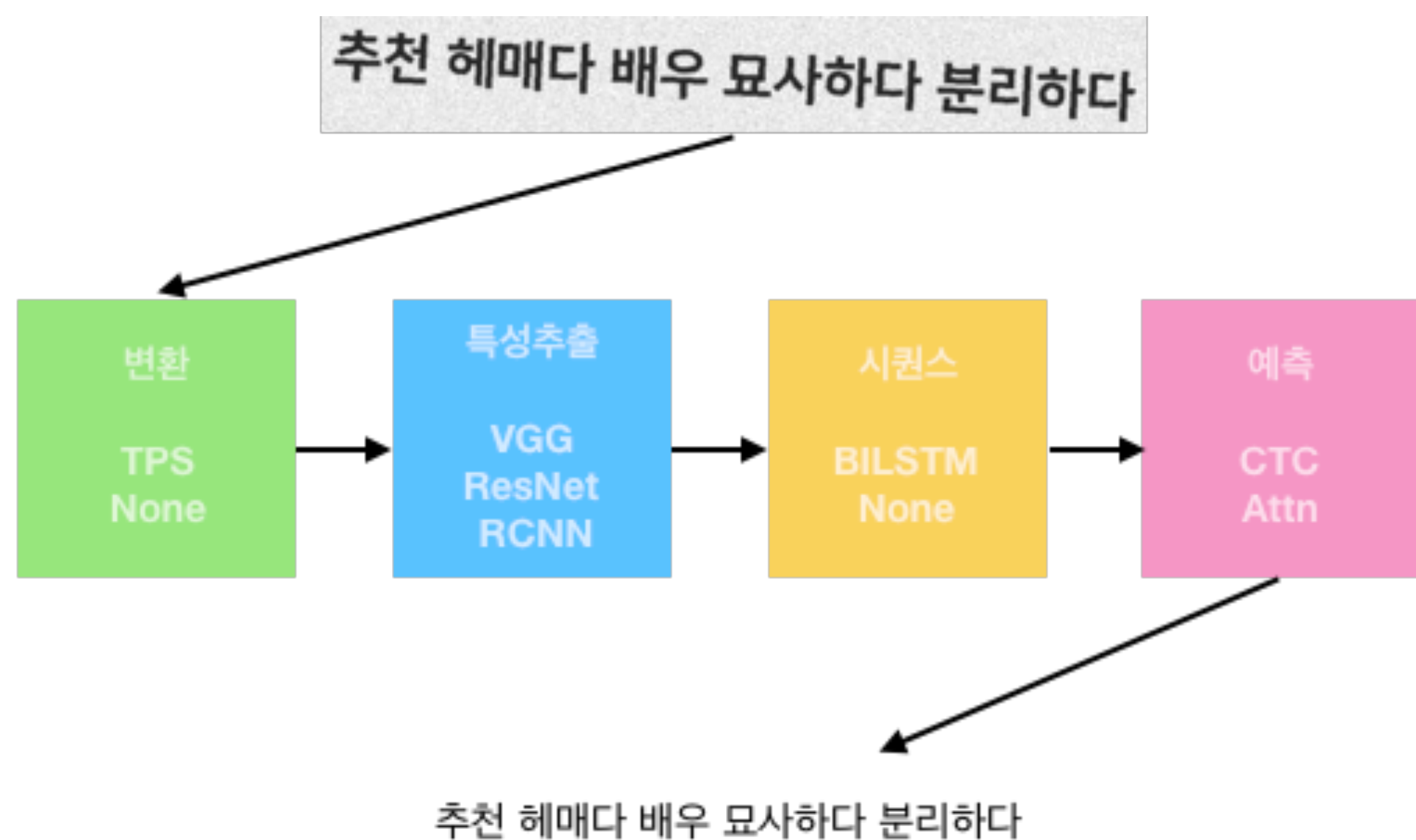


그림 1. 글자기반 문자탐지모델 스코어맵과 어피니티맵

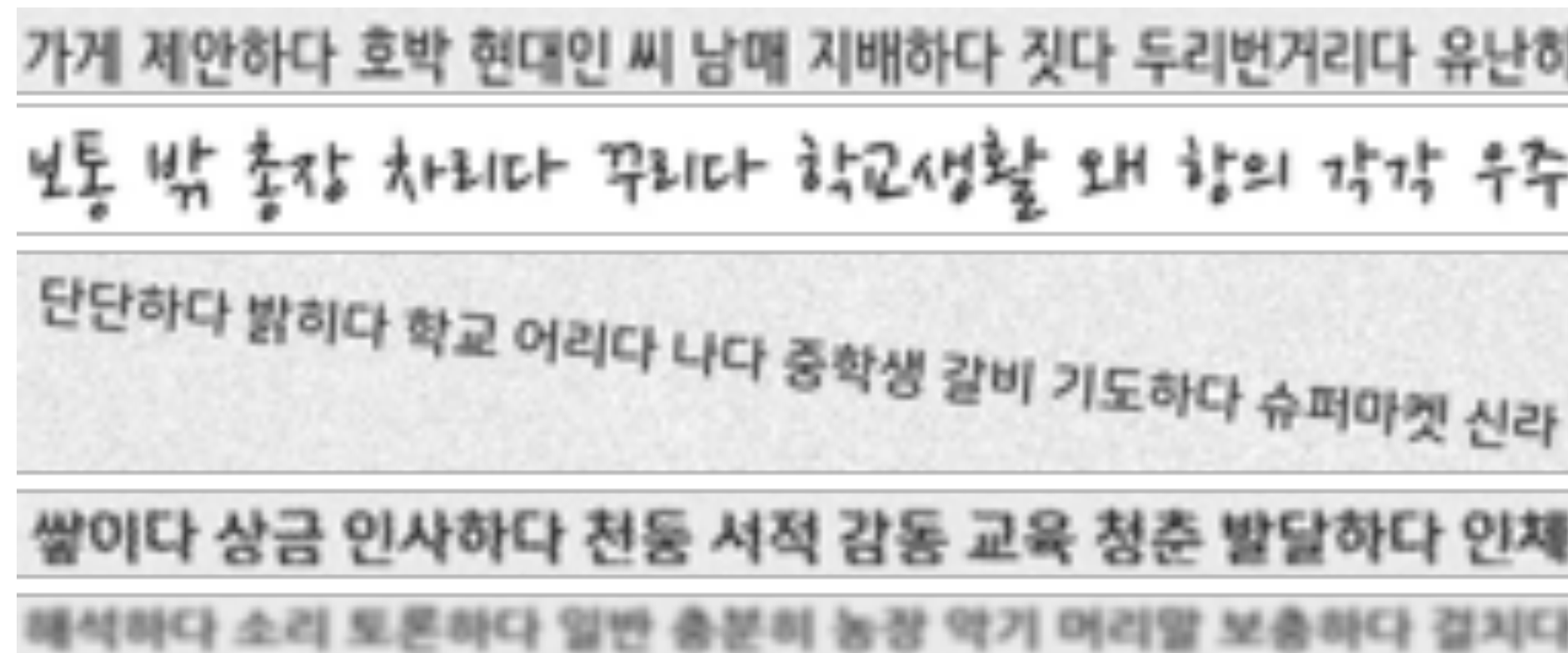


- ❖ 문자탐지 부분과 문자인식 부분의 비교
 - 그림 1은 글자 단위 문자탐지 모델을 사용했을 때 생성되는 글자 스코어 맵과(왼쪽) 글자 사이 여백을 표시하는 어피니티 맵(오른쪽)
 - 그림 2는 하나의 단어가 동일한 바운딩 박스 안에 포함되는 모습
 - OCR 성능을 높이기 위해서 문자인식 부분에 초점
- ❖ 모델구조
 - OCR을 위한 딥러닝 모델은 STR(Scene Text Recognition)의 구조를 가져와서 변환, 특성추출, 시퀀스, 예측 4개 모듈구성



❖ 데이터세트

- 데이터 세트 생성과정
 - 1) 단어 사전에서 랜덤으로 단어를 선정
 - 2) 한글 폰트 파일과 배경 3종 준비 (가우시안 노이즈, 순백색, Quasi crystal)
 - 3) 기본, 배경, 기울기, 왜곡, 흐리게 한 5가지 종류의 한글 문장 데이터 세트 생성
- 문서 파일에서 가능한 변형으로 (기울기, 왜곡, 흐리게하기, 배경) 선택, 한 개 문장 데이터가 포함하는 단어의 수는 10개로 설정



• 그림 3. 문자인식을 위한 생성 데이터. 위에서부터 차례로 기본,배경,기울기,왜곡,흐리게 한 문장 데이터

III.Experimental Results

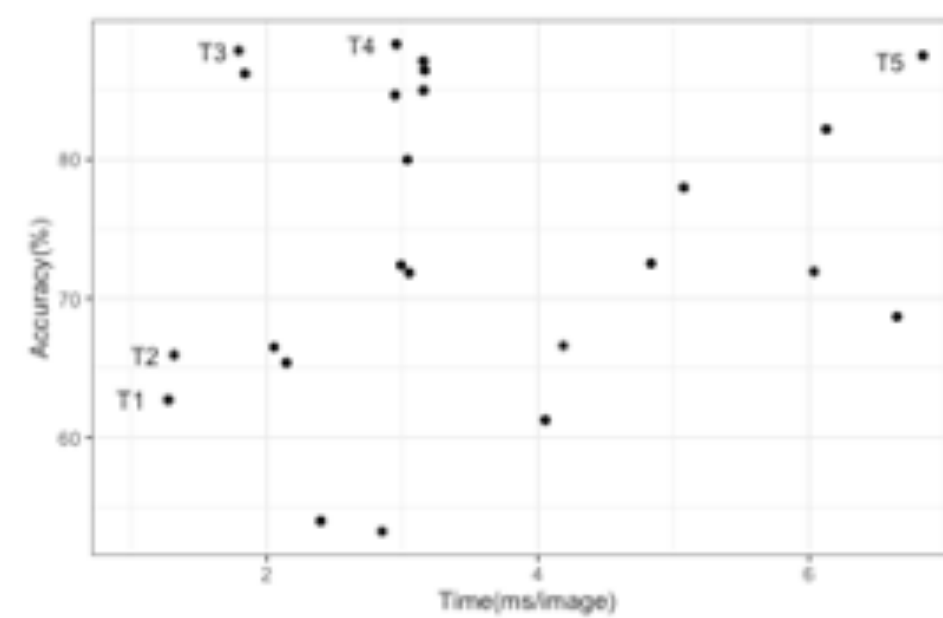


그림 4. 모델조합별 시간과 정확도 관계

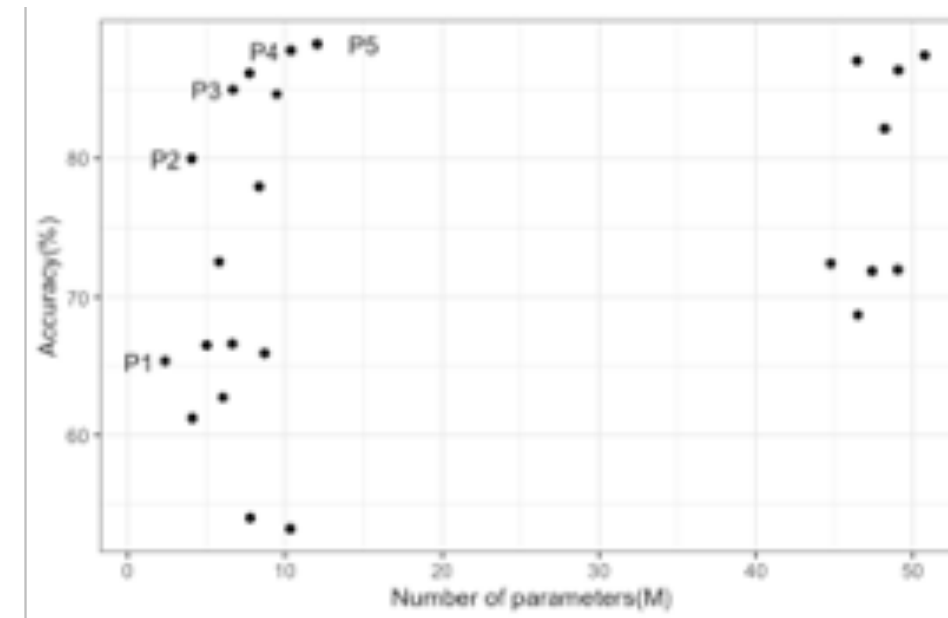


그림 5. 모델조합별 파라미터 수와 정확도 관계

#	변환	추출	시퀀스	예측	정확도%	시간 ms	파라미터
T1	None	VGG	None	CTC	62.72	1.27	6.04
T2	None	VGG	BiLSTM	CTC	65.91	1.31	8.69
T3	TPS	VGG	BiLSTM	CTC	87.79	1.79	10.38
T4	TPS	VGG	BiLSTM	Attn	88.24	2.95	12.04
T5	TPS	ResNet	BiLSTM	Attn	87.43	6.84	50.75

#	변환	추출	시퀀스	예측	정확도%	시간 ms	파라미터
P1	None	RCNN	None	CTC	65.36	2.14	2.35
P2	TPS	RCNN	None	CTC	79.96	3.03	4.05
P3	TPS	RCNN	BiLSTM	CTC	84.95	3.15	6.69
P4	TPS	VGG	BiLSTM	CTC	87.79	1.79	10.38
P5	TPS	VGG	BiLSTM	Attn	88.24	2.95	12.04

- 가장 높은 정확도를 보인 모델조합은 TPS-VGG-BiLSTM -Attention
- 변환 모듈에서 TPS(Thin Plate Spline)를 사용할 때에 정확도는 최대 34%까지 향상, 파라미터 수의 증가량은 1.6(M)으로 일정하지만, 추론시간은 모델 대부분이 증가
- 추출 모듈에서 VGG와 ResNet이 RCNN보다 다소 정확도가 높게 나왔고 모듈 간 평균 파라미터 수를 비교했을 때 ResNet(47.7M), VGG(9.0M), RCNN(5.3M) 순으로 큰 값
- 시퀀스 모듈에서 BiLSTM 사용 여부는 다른 모듈조합과의 관계에 따라 정확도와 추론시간이 바뀜
- 시퀀스-예측 모듈에서 BiLSTM-어텐션의 조합이 BiLSTM 미사용-어텐션 조합보다 정확도를 높이는 효과를 보여 BiLSTM과 어텐션을 함께 사용했을 때 전체적인 정확도에 영향

IV.Conclusions and Future works

한글에 대한 OCR 연구는 그 중요성에 비해 다양한 연구가 진행되지 않았다. OCR 모듈별로 주요 특징을 정리해보고 한글 문자인식에 적합한 딥러닝 모델 조합을 찾는 실험을 진행했다. 실제 애플리케이션에 적용 가능성을 높이고자 한글 문장 OCR 데이터를 문서에서 발견되는 형태로 직접 생성하였다. 한글 문장에 적합한 모델 조합은 TPS- VGG-BiLSTM-Attenion이 다른 모델 조합에 비해 높은 정확도를 보였다. 정확도 향상으로 실제 서비스에 사용하려면 최근 주목받는 트랜스포머, 버트 모델을 고려해볼 만하다.