

저화질 한글 공공 문서를 위한 OCR 모델 비교 분석 연구

이 혁*

Comparative Analysis of OCR Architectures for Low-Quality
Korean Public Documents

Hyuk Lee*

요약 저화질 공공 문서의 디지털화를 위해 OCR(Optical Character Recognition) 기술이 널리 활용되고 있으나, 이러한 저화질 환경에서 텍스트 탐지 및 인식 모델의 성능은 여전히 개선이 필요하다. 본 연구에서는 저화질 한글 공공 문서를 대상으로 DBNet과 PANet을 탐지 모델로, CRNN과 SAR을 인식 모델로 선정하여 다양한 조합의 성능을 비교 분석하였다. 실험 결과, DBNet과 SAR의 조합이 F1-score 85.30%, Word Accuracy 68.86%, 그리고 End-to-End 성능 58.74%를 기록하며, 정확도가 중요한 환경에서 우수한 성능을 보였다. 반면, DBNet과 CRNN의 조합은 Word Accuracy 66.43% 및 End-to-End 성능 56.66%를 유지하면서도 추론 시간이 SAR 대비 평균 7ms 단축되어, 실시간 처리가 필요한 응용에서 효과적인 대안임을 확인하였다. 본 연구는 공공 문서 디지털화를 위한 OCR 모델 조합의 효율성을 검증함으로써, 다양한 환경에서 활용 가능한 실무적 가이드라인을 제공한다.

Abstract OCR (Optical Character Recognition) technology is widely used for digitizing low-quality public documents. However, text detection and recognition models in such low-quality environments still require performance improvements. This study evaluates and compares various OCR model combinations for low-quality Korean public documents, selecting DBNet and PANet as text detection models and CRNN and SAR as text recognition models. Experimental results show that the combination of DBNet and SAR achieved an F1-score of 85.30%, a Word Accuracy of 68.86%, and an End-to-End performance of 58.74%, demonstrating superior performance in applications where high accuracy is essential. Meanwhile, the combination of DBNet and CRNN maintained a Word Accuracy of 66.43% with an End-to-End performance of 56.66%, while reducing inference time by an average of 7ms compared to SAR, making it an effective alternative for real-time processing applications. This study validates the efficiency of OCR model combinations for digitizing public documents and provides practical guidelines applicable to various environments.

Key Words : OCR, Text Detection, Text Recognition, Korean Public Document Digitization, Low-Quality Images

1. 서론

공공기관에서 보유한 방대한 종이 문서를 디지털화하는 것은 현대 행정의 효율성과 투명성을 높이기 위해 필수적인 과제이다 [1]. 디지털화된 문서는 국민에게

정보 접근성을 제공할 뿐만 아니라, 행정 업무의 효율성을 높이고 데이터 기반 의사결정을 지원할 수 있다. 특히, 1990년대 이전에 생성된 문서들은 저화질 스캔본이나 팩스 형태로 보관된 경우가 많아 물리적 손상이나 정보 소실의 위험이 높다. 이러한 문서의 디지털화

This research received funding from the 'Junior Faculty Research Support Grant' at Changwon National University in 2024. This research used datasets from 'The Open AI Dataset Project (AI-Hub, S. Korea)'. All data information can be accessed through 'AI-Hub (www.aihub.or.kr)'.

*Department of Computer Engineering, Changwon National University (hyuk@cwnu.ac.kr)

Received January 28, 2025

Revised February 05, 2025

Accepted February 18, 2025

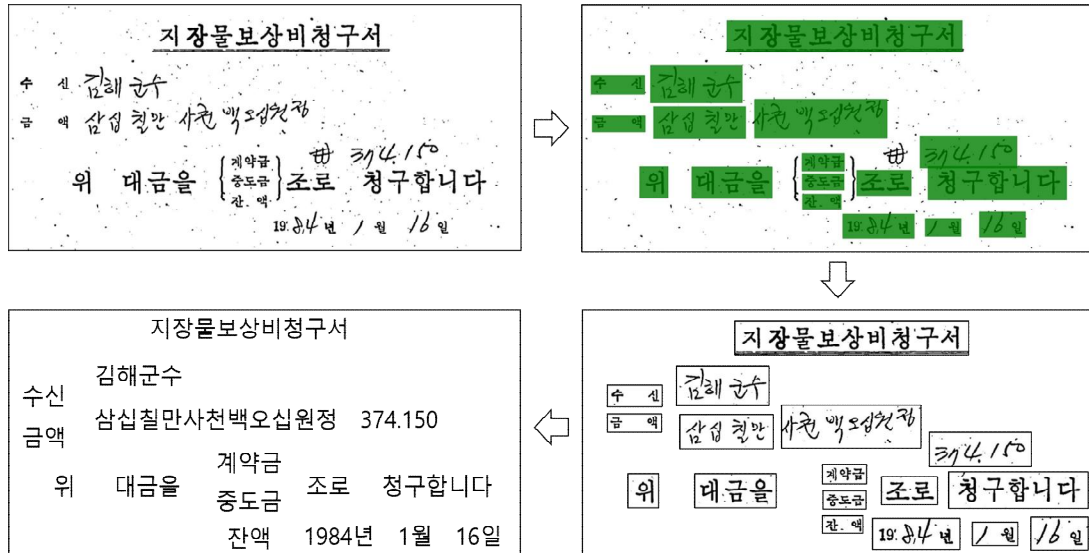


그림 1. 한글 공공문서의 OCR 처리 흐름도: 원본 문서 이미지에서 시작하여 텍스트 영역 탐지(녹색 박스), 개별 텍스트 영역 추출, 최종 한글 텍스트 인식까지의 순차적 과정.

Fig. 1. OCR Processing Flow for Korean Public Documents: Sequential process from original document image, through text region detection (green boxes) and individual text region extraction, to final Korean text recognition.

는 단순한 기록 보존을 넘어 행정 서비스 개선과 역사적 기록의 체계적 관리를 위해 반드시 필요하다.

광학 문자 인식(Optical Character Recognition, OCR) 기술은 이러한 디지털화 과정을 자동화하는 데 핵심적인 역할을 하지만, 저화질 문서나 한글과 같은 복잡한 문자 체계에서 기존 모델의 성능은 여전히 제한적이다 [2, 3]. 한글은 초성, 중성, 종성의 조합으로 이루어진 고유한 구조를 가지고 있으며, 이러한 특성으로 인해 기존 OCR 모델들이 영문 기반으로 최적화된 경우 성능 저하를 야기할 수 있다 [4]. 본 연구에서는 한글 문서의 이러한 고유한 특성을 직접적으로 다루지는 않지만, 저해상도와 노이즈가 포함된 저화질 한글 공공 문서에서의 OCR 성능을 개선하기 위한 다양한 모델 조합과 실험을 수행하여, 한글을 포함한 다양한 문자 환경에서의 가능성을 탐색할 것이다.

본 연구에서는 저화질 한글 공공 문서를 대상으로 OCR 모델의 성능을 비교하고, 텍스트 탐지와 인식 단계에서 최적의 모델 조합을 도출하는 데 초점을 맞춘다. 이를 위해 DBNet [5]과 PANet [6]을 탐지 모델로, CRNN [7]과 SAR [8]을 인식 모델로 선정하여 다

양한 조합의 성능을 평가하고, 저화질 환경에서 각 모델의 강점과 한계를 분석하고자 한다.

본 논문의 구성은 다음과 같다. 2장에서는 OCR 관련 선행 연구와 주요 모델들의 특징을 살펴보고, 3장에서는 실험 방법과 평가 지표를 설명한다. 4장에서는 실험 결과를 분석하여 적합한 모델 조합을 제시하며, 마지막으로 5장에서는 연구의 결론과 향후 연구 방향을 논의할 것이다.

2. 연구 배경 및 관련 연구

2.1 OCR 파이프라인

OCR 시스템은 일반적으로 텍스트 탐지(Text Detection)와 텍스트 인식(Text Recognition)의 두 단계로 구성된다. 텍스트 탐지는 이미지 내에서 텍스트가 포함된 영역을 식별하고, 텍스트의 위치와 크기를 나타내는 경계 상자를 생성한다. 텍스트 인식은 탐지된 텍스트 영역 내의 문자를 판독하여 이를 디지털 텍스트로 변환하는 과정이다 [9].

그림 1은 이러한 두 단계의 파이프라인을 시각적으

로 나타낸다. 텍스트 탐지 단계에서는 문서 내에서 텍스트 영역을 정확히 찾아내는 것이 중요하며, 탐지 결과는 텍스트 인식의 입력으로 사용된다. 텍스트 인식 단계는 탐지된 텍스트 영역을 기반으로 문자를 추출하며, 이미지 품질, 폰트, 언어적 특성 등의 영향을 받는다 [10, 11, 12].

이러한 두 단계 접근 방식은 각각의 과정을 독립적으로 최적화 할 수 있다는 장점이 있다. 그러나 저화질 공공 문서와 같이 노이즈가 많거나 해상도가 낮은 경우, 텍스트 탐지와 인식 모두에서 성능 저하가 발생할 수 있다. 본 연구에서는 이러한 문제를 해결하기 위해, 각 단계에 대한 모델(DBNet, PANet, CRNN, SAR)을 사용하여 성능을 평가하고, 저화질 환경에서의 효율성을 분석하였다.

2.2 텍스트 탐지 모델

본 연구에서는 DBNet과 PANet 두 가지 텍스트 탐지 모델을 선정하였다. DBNet은 텍스트 영역 탐지에서 후처리 과정의 복잡성을 줄이고 경계 탐지의 정확도를 높이기 위해 차별화 가능한 이진화 모듈을 도입한 모델이다. Probability Map과 Threshold Map을 동시에 학습하여 기하학적 변형에도 강건한 성능을 보이며, 간결한 후처리 과정과 빠른 추론 속도로 실시간 응용에 적합하다. PANet은 텍스트 영역의 지역적 특징과 전역적 문맥 정보를 통합적으로 활용하기 위해 Pixel Aggregation 메커니즘을 도입한 모델이다. FPN (Feature Pyramid Network) 구조를 활용하여 다양한 크기의 텍스트를 효과적으로 탐지하며, 복잡한 배경에서도 안정적인 성능을 발휘한다. 특히, 텍스트의 불규칙한 형상과 크기 변화에 강점을 가진다.

DBNet은 빠른 처리 속도와 실시간 응용에 적합하며, PANet은 다양한 크기와 복잡한 배경을 가진 텍스트 탐지에서 강점을 보인다. 두 모델 모두 저화질 공공 문서에서 텍스트 영역 탐지의 성능을 비교 분석하는데 적합하다.

2.3 텍스트 인식 모델

텍스트 인식을 위해 CRNN과 SAR 두 가지 모델을

선정하였다. CRNN(Convolutional Recurrent Neural Network)은 Convolutional Neural Network(CNN)와 Recurrent Neural Network(RNN)를 결합하여 이미지의 시각적 특징과 문자 시퀀스 간의 관계를 효과적으로 모델링하는 구조를 가지고 있다. 이 모델은 CNN을 통해 이미지 특징을 추출하고, RNN을 통해 시퀀스 데이터를 처리하며, CTC(Connectionist Temporal Classification) 손실을 사용하여 가변 길이의 문자 시퀀스를 학습할 수 있다. CRNN은 간결한 구조와 효율적인 학습으로 실시간 처리가 필요한 응용에 적합하다.

SAR(Show, Attend and Read)은 2D Attention 메커니즘을 활용하여 이미지의 지역적 특징과 전역적 문맥 정보를 결합한다. 특히 SAR은 왜곡되거나 복잡한 레이아웃을 가진 텍스트 이미지에서 강점을 가지며, 초성/중성/종성의 결합 구조를 포함한 다양한 문자 유형의 의존성을 모델링할 수 있다. 이는 한글처럼 구조가 복잡한 문자에서도 더 높은 인식 정확도를 제공할 가능성을 보여준다.

3. 실험 설계 및 평가

3.1 데이터셋

본 연구에서는 AI Hub에서 제공하는 '대규모 OCR 데이터(공공) AI 데이터셋'을 활용하였다 [13]. 이 데이터셋은 지방자치단체(창원특례시, 김해시)와 외교사료관의 공공문서를 수집하여 구축되었으며, 특히 저화질 문서의 문자 인식과 다양한 형태의 문자체(인쇄체, 타자체, 수기 등) 인식을 위해 설계되었다.

3.1.1 데이터셋 선정

공공 문서 데이터셋의 특성을 고려하여 단계별 선별 과정을 수행하였다. 초기 데이터는 총 615,053건으로 구성되었으며, 1차 선별 과정에서는 1980년대 이전과 1980년대에 생성된 문서를 기준으로 데이터 183,527건을 추출하였다. 이후, 라벨링 파일과 이미지의 매칭 정확도를 평가하여 133,139건을 2차로 선별하였다.

마지막으로, 학습에 적합한 문서를 선정하기 위해

표 1. 데이터 필터링 및 선별 과정 요약

Table 1. Summary of Data Filtering and Selection Process

Stage	Number of Data	Ratio (%)	Selection Criteria
Initial Data	615,053	100	Entire public documents
First Selection	183,527	29.8	Documents from the 1980s and earlier
Second Selection	133,139	21.6	Label matching accuracy validation
Final Data	6,583	4.94	Average text length ≥ 6 characters, number of text boxes ≥ 15 , etc.

평균 텍스트 길이, 한글 비율, 텍스트 박스 수와 같은 정량적 기준을 적용한 결과, 최종적으로 6,583건을 확보하였다. 표1은 각 단계별 데이터 수와 선별 기준을 요약하여 나타낸 것이다. 최종 선정된 데이터셋의 특징은 다음과 같다.

- 시기: 1980년대 이전 및 1980년대 공공문서
- 문서 유형: 건설, 경제, 농림수산, 문화, 복지, 환경 등 다양한 행정 분야
- 문자 형태: 인쇄체, 타자체, 수기 등 복합적 구성
- 화질 특성: 저해상도, 노이즈, 변색 등 저화질 특성 포함

이러한 데이터셋 구성은 실제 공공기관에서 직면하는 문서 디지털화 과제의 특성을 잘 반영하고 있으며, 특히 오래된 문서의 OCR 처리에서 발생하는 다양한 문제를 포괄적으로 다룰 수 있게 한다.

3.2 데이터 전처리 및 증강

DBNet과 PANet은 기본적으로 동일한 전처리 과정을 공유하지만, 입력 크기와 데이터 증강 방식에서 차이를 보인다. 두 모델은 mean=[123.675, 116.28, 103.53], std=[58.395, 57.12, 57.375]의 정규화 파라미터를 사용하고, BGR에서 RGB로 색상 변환을 수행하였다.

DBNet은 학습 시 입력 크기를 640×640으로, PANet은 736×736으로 설정하여 처리하였다. 또한, 증강 기법으로는 색상 조정, 좌우 반전, 크기 조정 및 크롭 방식을 적용하였으며, 각 모델의 특징에 따라 설정이 상이하다. DBNet은 Resize와 RandomCrop를 활

용하여 크기를 조정한 반면, PANet은 ShortScaleAspectJitter와 TextDetRandomCrop를 통해 텍스트 탐지에 최적화된 증강을 수행하였다. 테스트 단계에서는 DBNet이 4068×1024의 고해상도로 입력 이미지를 처리하여 탐지 정확도를 높인 반면, PANet은 학습 시와 동일한 크기(736×736)로 테스트를 진행하였다. 이는 각각 높은 해상도 기반의 세밀한 탐지와 일관된 입력 크기에 중점을 둔 설계 차이를 반영한다.

이러한 전처리와 증강 설정은 두 모델의 구조적 특징과 학습 목적에 최적화되어 있으며, 결과적으로 모델 성능 및 활용 가능성에 영향을 미친다.

3.3 실험 설정 및 평가 지표

텍스트 탐지와 인식을 위해 각각 DBNet과 PANet, CRNN과 SAR 모델을 선정하여 동일한 공공 문서 데이터셋으로 학습을 진행하였다. 이들 모델은 PyTorch와 MMDetection을 기반으로 텍스트 탐지와 인식을 지원하는 오픈소스 도구인 MMOCR을 활용하여 구현되었다. MMOCR은 OpenMMLab 프로젝트의 일부로, 다양한 텍스트 관련 작업을 위한 효율적인 학습 및 실험 환경을 제공한다 [14].

탐지 모델(DBNet, PANet)은 100 에포크 동안 학습되었으며, 인식 모델(CRNN, SAR)은 50 에포크 동안 학습되었다. 성능 평가는 텍스트 탐지와 인식 단계별로 적합한 지표를 적용하여 정량적으로 분석하였다.

텍스트 탐지: 탐지된 텍스트 영역과 실제 정답 영역 간의 Intersection over Union (IoU) 기준을 0.5로 설정하여 Precision, Recall, F1-Score를 산출하였다. Precision은 모델이 탐지한 영역 중 실제와 일치하는 비율을, Recall은 실제 정답 영역 중 올바르게 탐

지된 비율을 나타내며, F1-Score는 두 지표의 조화 평균을 계산한다.

텍스트 인식: 탐지된 영역 내의 문자 시퀀스를 기반으로 문자 단위(Character-level) Precision과 Recall, 그리고 F1-Score를 평가하였다. Character Precision은 인식된 문자 중 정답과 일치하는 비율, Character Recall은 정답 문자 중 올바르게 인식된 비율을 의미한다. 이러한 평가 지표를 통해 각 모델의 정확도와 재현성을 비교하고, 저화질 공공 문서에서의 강점과 한계를 체계적으로 파악하였다.

4. 실험 결과 및 분석

4.1 텍스트 탐지 모델 성능 분석

텍스트 탐지 모델의 성능을 평가하기 위해 DBNet과 PANet을 실험하였다. 두 모델의 주요 성능 지표(Precision, Recall, F1-score)는 표 2에 요약되어 있다.

표 2. DBNet과 PANet의 텍스트 탐지 성능 비교
Table 2. Comparison of Text Detection Performance: DBNet and PANet

Model	Precision	Recall	F1-score	Optimal Epoch
DBNet	86.65%	83.98%	85.30%	100
PANet	85.99%	79.32%	82.52%	25

DBNet은 최종 Precision 86.65%, Recall 83.98%, F1-score 85.30%를 기록하였으며, PANet은 Precision 85.99%, Recall 79.32%, F1-score 82.52%를 기록하였다. DBNet은 모든 성능 지표에서 PANet보다 우수한 결과를 보였으며, 특히 Recall에서 4.66%p 높은 성능을 달성하였다.

그림 2와 그림 3은 두 모델의 학습 에포크별 Precision, Recall, F1-score의 변화를 시각적으로 나타낸다. 그림 2는 DBNet의 성능 변화를, 그림 3은 PANet의 성능 변화를 보여준다. DBNet은 높은 Precision과 Recall을 동시에 유지하며 안정적인 학습 과정을 보였다. 반면, PANet은 학습 초기에 급격히 성능이 향상되었으나, 학습 후반부에서는 Gradient Explosion

으로 인해 더 이상 학습이 진행되지 못하였다. 두 모델의 특성을 종합적으로 평가하면, DBNet은 저화질 공공 문서와 같은 까다로운 환경에서 더 높은 성능과 안정성을 제공하는 반면, PANet은 상대적으로 짧은 학습 시간 안에 최적의 성능에 도달할 수 있는 특징이 있다.

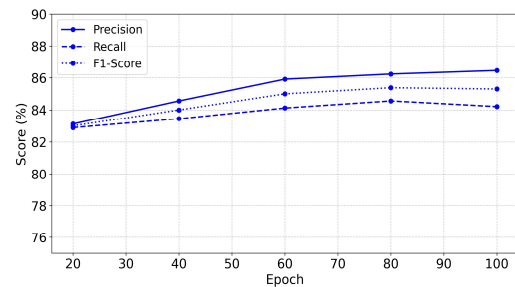


그림 2. DBNet의 에포크별 성능 변화

Fig. 2. Epoch-wise Performance Changes of DBNet

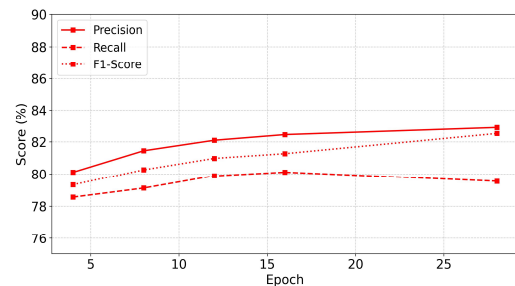


그림 3. PANet의 에포크별 성능 변화

Fig. 3. Epoch-wise Performance Changes of PANet

4.2 텍스트 인식 모델 성능 분석

CRNN과 SAR 모델의 텍스트 인식 성능을 비교 분석하였다. CRNN은 CTC loss 기반의 간결한 구조를 활용하여 효율적으로 학습되었으며, SAR은 2D Attention 메커니즘을 통해 복잡한 텍스트 구조에서 강점을 보였다. 표 3은 두 모델의 최종 성능을 요약한 결과로, SAR이 정확도 측면에서 CRNN보다 1.95%p 높은 87.25%를 기록하였다. 하지만 SAR의 추론 시간은 평균 52ms로, CRNN의 45ms보다 다소 긴 것으로 나타났다. 이는 SAR의 Attention 연산에 따른 추가적인

계산 비용에서 발생한 것으로 보인다.

표 3. 텍스트 인식 모델의 성능 비교
Table 3. Comparison of Text Recognition Model Performance

Model	Accuracy	Inference Time	Optimal Epoch
CRNN	85.30%	45ms	100
SAR	87.25%	52ms	85

그림 4와 그림 5는 학습 에포크별 Word Accuracy 변화와 최종 Character Precision/Recall 성능을 시각적으로 나타낸다. 그림 4에서 SAR은 초기 에포크부터 CRNN보다 높은 Word Accuracy를 기록하며, 학습 후반부에서도 더 안정적인 수렴 패턴을 보였다. 그림 5는 최종 Character Precision과 Recall을 비교한 결과로, SAR이 각각 10.3%p와 18.1%p 더 높은 성능을 기록하였다.

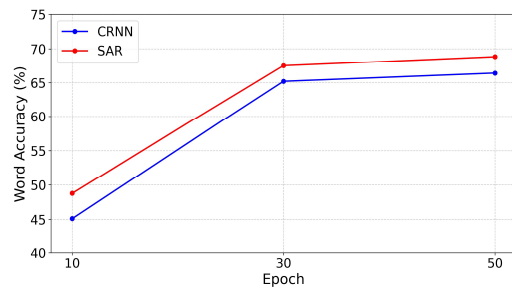


그림 4. 학습 에포크별 Word Accuracy 변화
Fig. 4. Epoch-wise Changes in Word Accuracy

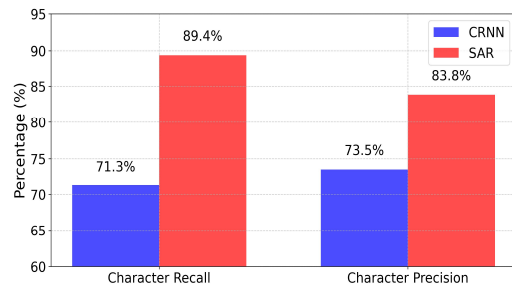


그림 5. 최종 Character Precision 및 Recall 비교
Fig. 5. Final Comparison of Character Precision and Recall

표 4는 에포크 간 Word Accuracy, 기호를 무시한 Word Acc. (Ign.), Character Recall, Precision의 변화를 수치로 나타내어 두 모델의 학습 경향을 구체적으로 보여준다. 특히 SAR은 학습이 진행됨에 따라 Character Recall과 Precision이 꾸준히 향상되어 최종적으로 89.38%와 83.84%를 달성하였다. 반면, CRNN은 30 에포크에서 가장 높은 성능을 기록했으나, 이후 약간의 성능 저하를 보였다.

SAR은 정확도 측면에서 CRNN보다 우수한 성능을 보였으며, 2D Attention 메커니즘이 왜곡된 텍스트나 복잡한 문서 구조를 처리하는 데 강점을 보였다. 반면, CRNN은 더 빠른 추론 속도를 통해 실시간 처리가 필요한 환경에 적합한 것으로 판단된다. 이러한 결과는 한글 문서의 복잡한 구조와 다양한 폰트 변화에서도 두 모델이 효과적으로 작동할 수 있음을 시사한다.

표 4. CRNN과 SAR 모델의 성능 비교 (에포크 별)
Table 4. Performance Comparison of CRNN and SAR Models (by Epoch)

Model	Evaluation Metric	Number of Epochs		
		10	30	50
CRNN	Word Accuracy	45.12%	65.21%	66.43%
	Word Acc. (Ign.)	86.78%	93.81%	88.50%
	Char Recall	56.55%	83.64%	71.29%
	Char Precision	67.44%	90.50%	73.52%
	Char F1-Score	61.52%	86.92%	72.39%
SAR	Word Accuracy	48.78%	67.53%	68.86%
	Word Acc. (Ign.)	90.87%	92.59%	93.09%
	Char Recall	74.65%	89.27%	89.38%
	Char Precision	79.29%	80.86%	83.84%
	Char F1-Score	76.89%	84.86%	86.52%

4.3 모델 조합 성능 분석

텍스트 탐지와 인식 모델의 성능을 종합적으로 비교하여 사용환경에 적합한 최적의 모델 조합을 도출하였다. 탐지 모델의 경우, DBNet은 PANet 대비 모든 평가 지표에서 우수한 성능을 보였으며, F1-score에서 2.78%p, Recall에서 4.66%p 더 높은 결과를 기록하였다. 또한 DBNet은 100 에포크까지 안정적으로 학습된 반면, PANet은 28 에포크에서 학습이 중단되는

인식 모델의 경우, SAR은 CRNN 대비 2.43%p 높은 Word Accuracy(68.86% vs 66.43%)를 기록하였으며, Character Recall과 Precision에서도 각각 89.38%와 83.84%로 우수한 성능을 보였다. 다만, SAR의 평균 추론 시간은 CRNN 대비 7ms 더 소요되었으므로, 실시간 처리가 필요한 환경에서는 제약이 될 수 있다.

표 5. 모델 조합별 예상 End-to-End 성능
Table 5. Estimated End-to-End Performance by Model Combination

Det. Model	Recog. Precision	Performance	
		Individual (F1-Score/Word Acc.)	End-to-End
DBNNet	CRNN	85.30% / 66.43%	56.66%
	SAR	85.30% / 68.86%	58.74%
PANet	CRNN	82.52% / 66.43%	54.82%
	SAR	82.52% / 68.86%	56.82%

성능이 제일 우수한 조합인 DBNet과 SAR 모델을 결합한 OCR 파이프라인의 실험 결과를 분석하였다.




분류기준 문서번호		도사 2024년 ~ 2034년		기안 용지		(연하 :)		시행식 문법기준	
보존기간		영구 - 영구		A		A		B	
보존식 보존기간									
서명일자		88. 4.						7-	
비고		부사장 자 결		임				문서통제	
파견		파견		기					
기안제임자		29		관				발행	
결유		대관자적공사		발신		A			
수신		강대사지부		문의		A			
참조									
제목		분할추방의뢰							
송점 ~ 벌매리 (금화 ~ 성림사) 도로 확충공사 평면도면을 다음과 같이									
분할추방의뢰 하되 조판사입내에 결판을 통한 라미 처리 바랍니다									

Fig. 6. Text Detection Results on Public Document

전화 문서통제 보존기간 부서장

(전화 : 문서통제 보존기간 부서장

보존기간 분할출량의리 88 4 조

보존기간 분할출량의리 88. 4. 조

Fig. 7. Examples of Text Recognition Results from Detected Regions

그림 6과 같이 문서 내 텍스트 영역 탐지는 대부분 정확하게 이루어졌으며, 그림 7에서 보듯이 탐지된 영역의 텍스트 인식도 전반적으로 높은 정확도를 보여주었다. 그러나 몇 가지 한계점도 발견되었다. 첫째, 기호나 특수문자에 대한 인식 성능이 상대적으로 낮았다. 둘째, 세로 방향으로 작성된 텍스트나 단어 사이에 큰 간격이 있는 경우, 시스템이 이를 개별 문자로 잘못 인식하는 경향이 있었다. 이러한 결과를 통해 한글 문서 OCR 시스템을 실제로 활용하기 위해서는 다양한 텍스트 레이아웃과 특수문자 처리 능력을 더욱 향상시

켜야 할 것으로 보인다.

5. 결론 및 향후 연구 방향

저화질 공공 문서의 텍스트 탐지와 인식을 위한 최적의 모델 조합을 분석한 결과, 텍스트 탐지 단계에서는 DBNet이 85.30%의 F1-score로 가장 우수한 성능을 보였다. 텍스트 인식 단계에서는 SAR이 87.25%의 정확도를 기록하며 최고의 성능을 나타냈다. 특히 DBNet과 SAR의 조합은 정확도가 중요한 응용 환경에 적합하며, DBNet과 CRNN의 조합은 실시간 처리가 필요한 상황에서 효과적인 대안임을 확인하였다. 이러한 결과는 공공 문서 디지털화를 위한 효율적인 OCR 시스템 구축에 중요한 방향성을 제시한다.

이번 연구에서는 한글 공공 문서 디지털화를 목표로, 저화질 문서에서 발생하는 문제를 체계적으로 분석하고 이를 극복하기 위한 다양한 OCR 모델의 성능을 정량적으로 평가하였다. 이를 통해 공공 문서 디지털화 과정에서 활용할 수 있는 실무적 가이드라인을 제공하였으며, 향후 OCR 기술 개선의 기초 자료로 활용될 수 있을 것이다.

그럼에도 불구하고, 몇 가지 한계점이 존재한다. 본 연구는 제한된 공공 문서 데이터셋을 기반으로 진행되었기에 결과를 일반화하는 데에는 신중한 접근이 필요하다. 또한, 한글의 구조적 특성을 충분히 반영하지 못한 모델 아키텍처와, 조명이나 촬영 각도와 같은 실제 환경 요인에 대한 추가 검증이 부족하다는 점도 보완해야 할 과제로 남아 있다.

향후 연구는 이번 연구에서 제시한 모델 조합 분석을 기반으로 OCR 시스템의 실용성과 범용성을 더욱 확장하는 데 초점을 맞출 계획이다. 한글 문서에서 자주 발생하는 왜곡과 번짐 현상을 처리할 수 있는 데이터 증강 기법을 개발하여 저화질 문서 환경에서도 강건한 성능을 확보하고자 한다. 또한, 모델 경량화 및 추론 속도 최적화를 통해 실시간 처리가 필요한 응용 환경에서의 활용 가능성을 높일 예정이다. 이러한 확장은 공공 문서뿐만 아니라 다양한 실제 환경에서 OCR 기술의 실질적 활용성을 한층 높이는 데 기여할 것이다.

REFERENCES

- [1] A. Risnumawan, P. Shivakumara, C. S. Chan, and C. L. Tan, "A robust arbitrary text detection system for natural scene images," *Expert Systems with Applications*, vol. 41, no. 18, pp. 8027-8048, 2014.
- [2] A. Gupta, R. Gutierrez-Osuna, M. Christy, B. Capitanu, L. Auvil, L. Grumbach, R. Furuta, and L. Mandell, "Automatic assessment of ocr quality in historical documents," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 29, 2015.
- [3] G. Kang, J. H. Ko, Y. J. Kwon, N. Y. Kwon, and S. J. Koh, "A study on improvement of korean ocr accuracy using deep learning," in *Proceedings of the Korean Institute of Information and Commucation Sciences Conference*, pp. 693-695, The Korea Institute of Information and Commucation Engineering, 2018.
- [4] H. Kim, "Considerations for applying korean natural language processing technology in records management," *Journal of Korean Society of Archives and Records Management*, vol. 22, no. 4, pp. 129-149, 2022.
- [5] M. Liao, Z. Wan, C. Yao, K. Chen, and X. Bai, "Real-time scene text detection with differentiable binarization," *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 11474-11481, 2020.
- [6] W. Wang, E. Xie, X. Song, Y. Zang, W. Wang, T. Lu, G. Yu, and C. Shen, "Efficient and accurate arbitrary-shaped text detection with pixel aggregation network," in *ICCV*, pp. 8439-8448, 2019.
- [7] B. Shi, X. Bai, and C. Yao, "An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition," *IEEE transactions on pattern analysis and machine intelligence*, 2016.
- [8] H. Li, P. Wang, C. Shen, and G. Zhang, "Show, attend and read: A simple and strong baseline for irregular text

- recognition," in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, pp. 8610-8617, 2019.
- [9] S. Long, X. He, and C. Yao, "Scene text detection and recognition: The deep learning era," International Journal of Computer Vision, vol. 129, no. 1, pp. 161-184, 2021.
- [10] Z. Raisi, M. A. Naiel, P. Fieguth, S. Wardell, and J. Zelek, "Text detection and recognition in the wild: A review," arXiv preprint arXiv:2006.04305, 2020.
- [11] S. Ahn, H. Hwang, and J. H. Yim, "A case study on the application of ai-ocr for data transformation of paper records," Journal of the Korean Society for Information Management, vol. 39, no. 3, pp. 165-193, 2022.
- [12] J. Lee and M. Yoo, "A research on digitization of non-electronic records through ocr technology - focused on deep learning -," Journal of D-Culture Archives, vol. 7, no. 1, pp. 259-276, 2024.
- [13] "OCR data (public documents)," 2023. An OCR dataset provided by AI-Hub, including various character types from public documents.
- [14] M. D. Team, "MMOCR: A comprehensive toolbox for text detection, recognition and understanding," 2022.

저자약력

이 혁 (Hyuk, Lee)

[정회원]



- 2006년 9월: 시드니공과대학 IT 학부 학사
- 2009년 2월: 고려대학교 컴퓨터 학과 석사
- 2019년 2월: 고려대학교 컴퓨터 학과 박사
- 2022년~현재: 국립창원대학교 컴퓨터공학과 조교수

〈관심분야〉 정형기법, 제약만족 문제, 시큐어 소프트웨어 공학, 인공지능