

Online to Offline 상점을 위한 한글 메뉴판 인식 : 어텐션 메커니즘을 적용한 VGG-ResNet 융합 모델

시종욱*, 이상진**, 김성영***

Recognition of Korean Menu for Online to Offline Stores : VGG-ResNet Fusion Model with Attention Mechanism

Jongwook Si*, Sangjin Lee**, Sungyoung Kim***

요약 O2O 상점 모델은 온/오프라인의 경계를 허물어 고객에게 큰 편의성을 제공하는 플랫폼이다. 이러한 플랫폼을 효과적으로 운영하기 위해서는 소상공인들이 필요한 정보를 디지털 형태로 제공해야 한다. 특히, 한글 메뉴판을 디지털화하는 과정이 수동으로 진행될 경우 여러 문제점을 일으킬 수 있으며, OCR 기술 사용 시 한글의 인식 정확도가 낮아 오류 인식의 가능성이 높다. 이에 본 논문에서는 한글 메뉴판의 자동 인식을 위해 대표적인 OCR 모델인 EasyOCR을 기반으로 하되, 한글 문자 인식의 낮은 정확도를 개선하고자 한다. 제안하는 모델은 VGG와 ResNet의 구조적 장점을 통합하고, 어텐션 메커니즘을 도입하여 한글 문자의 인식 성능을 크게 향상시키도록 설계한다. 실험 결과, 제안하는 모델은 EasyOCR에 비해 Accuracy 기준 약 3.5%, Confidence Score와 Normalized Edit Distance 기준 약 1%의 인식 정확도 향상을 보였다. 따라서, 제안한 방법이 기존 문제를 효과적으로 해결할 수 있음을 입증한다.

Abstract The O2O store model dissolves the boundaries between online and offline platforms, providing significant convenience to customers. To effectively operate such platforms, small business owners must provide necessary information in digital format. Specifically, the process of digitizing Korean menus manually can lead to multiple issues, and the use of OCR technology often results in high error rates due to the low accuracy in recognizing Korean. In response, this paper proposes an enhanced OCR model based on the popular EasyOCR framework, aimed at improving the recognition accuracy of Korean. The proposed model integrates the structural advantages of VGG and ResNet, and incorporates an attention mechanism to significantly improve the recognition performance of Korean. Moreover, experimental results indicate that the proposed model achieved approximately a 3.5% improvement in accuracy and around a 1% improvement in both confidence score and normalized edit distance compared to EasyOCR. Therefore, this demonstrates that the proposed method effectively addresses the existing challenges.

Key Words : Online to Offline(O2O), OCR(Optical Character Recognition), Text Recognition, Attention Mechanism, Korean Menu

1. 서론

O2O(Online to Offline) 상점이란, 온라인과 오프

라인 매장을 연결하여 거래를 완료하는 상업 모델을 말한다. 이 모델은 음식점이나 쇼핑몰 등의 다양한 산업에서 활용될 수 있으며, 고객에게 편리함과 추가 가치

This work was supported by the Technology development Program(S3344882) funded by the Ministry of SMEs and Startups(MSS, Korea)

* Dept. of Computer-AI Convergence Engineering, Kumoh National Institute of Technology

** Dept. of Computer Engineering, Kumoh National Institute of Technology

*** Dept. of Computer Engineering, Kumoh National Institute of Technology (Corresponding Author)

Received July 16, 2024

Revised July 29, 2024

Accepted August 12, 2024

를 제공할 수 있다. 이러한 O2O 상점 모델의 성공은 매장과 온라인 플랫폼 간의 원활한 상호작용에 크게 의존하기 때문에 상점의 정확한 정보가 필수적이다.

현재 O2O 상점 운영에 있어서 가장 중요한 절차 중 하나는 상점의 메뉴를 디지털화하여 온라인 상점에 등록하는 것이다[1]. 이 과정은 주로 수동으로 이루어지며, 상점 주인이 직접 메뉴 항목과 가격을 입력하여 가상의 메뉴판을 구성해야 한다. 이러한 수동적 접근 방식은 시간이 많이 소요되며, 효율성이 떨어진다는 단점이 있다. 또한, 이는 자주 발생하는 메뉴 갱신을 어렵게 만들어, 실시간으로 변화하는 시장 요구와 고객의 기대에 신속하게 대응하기 어렵게 만든다.

한글 메뉴판의 경우 다양한 글씨체와 복잡한 구성으로 인해 기존 OCR(Optical Character Recognition) 기술로는 정확한 인식이 어렵다는 단점이 있다[2]. 대부분의 상용 OCR 시스템은 영어와 같은 라틴 기반 언어에 최적화되어 있으며, 한글 특유의 문자 구조와 결합 형태를 정확하게 인식하는 데 한계가 있다. 이는 한글 메뉴판 인식에 있어서 낮은 정확도와 높은 오류율을 초래하며, 이로 인해 소상공인들은 자동화된 도구를 통해 효과적으로 메뉴를 디지털화하는 데 어려움을 겪고 있다.

본 논문에서는 O2O 상점의 구축과 운영을 위해 위에서 언급한 소상공인들이 직면하는 문제를 해결하고자 한다. 이러한 문제를 해결하기 위해, EasyOCR[3]을 기반으로 한 향상된 모델을 제안한다. 본 논문에서는 특징 추출을 위한 모델 구조와 Attention 메커니즘을 활용한 OCR 구조를 제안한다. 이 방법을 통해 메뉴판의 한글 문자 인식률을 개선하고, O2O 상점의 운영 효율성을 증진시키는 것을 목표로 한다.

2. 관련 연구

2.1 OCR 기술 연구

OCR(Optical Character Recognition)은 영상에서 글자를 인식하는 기술로, 서류, 차량 번호판 등 다양한 분야에서 글자를 인식하는 연구가 활발히 진행되고 있다. 대표적인 OCR 기술로는 Tesseract-OCR[4], Keras-OCR[5], EasyOCR[3]이 있다.

Tesseract-OCR[4]은 C와 C++ 기반의 OCR 소프

트웨어로, CPU를 사용하여 OCR 과정을 진행하는 특징이 있다. 이는 온오프라인에서의 정보 전달 속도에 영향을 미칠 수 있다. 하지만, 다양한 언어를 지원하며 정확도가 높은 편이지만 한글의 경우 낮은 정확도를 보인다.

Keras-OCR[5]은 고전적인 OCR 구조인 검출(Detection)과 인식(Recognition) 단계를 따르며, 특히 인식 부분에서 글자의 연속적인 성질을 인식하는데 효과적인 CRNN[6]을 이용한다. 그러나 한글 인식을 지원하지 않고 영어만을 지원하기 때문에 국내 O2O 상점에서 활용하기 어렵다.

EasyOCR[3] 역시 검출과 인식 구조를 사용하며, 인식 부분의 모델을 교체할 수 있는 유연성을 제공한다. 이 기술은 다양한 언어에 대한 각각의 모델을 따로 두어, 인식할 언어에 맞는 모델을 선택하여 사용할 수 있다. 이를 통해 한글을 포함한 여러 언어를 인식할 수 있다. 기존의 한글 인식을 위해 학습된 가중치를 사용한다면 효율적이지만, 실제로 활용하기에는 정확도가 낮다. 그러나 EasyOCR은 위치 정보까지 제공할 수 있으며, GPU를 활용하기에 인식 속도가 빠르다는 장점이 있다. 따라서, 본 논문에서는 다양한 장점이 있는 EasyOCR을 채택하여 한글 인식 정확도만을 개선하는 것을 목표로 한다.

2.2 한글 인식 연구

한글은 다른 언어들에 비해 독특한 기원을 가지며, 초성, 중성, 종성을 조합하여 하나의 글자를 구성한다. 이러한 조합으로 인해 한글은 총 11,172개의 글자를 가지며, 이는 영어보다 분류해야 할 글자의 수가 많다는 것을 의미한다.

장혁수 연구팀[7]은 한글의 언어적 특성에 따른 인식을 문제를 개선하기 위해 CNN을 이용한 인식 방법을 제안하였다. 이 연구는 한글의 완성형 글자를 구분하는 것에 중점을 두며, 이를 통해 영수증에서 인식할 때 높은 성능을 보였다. 하지만, 이 방법은 정형화된 한글을 대상으로 한다는 점에서 차이점이 있다.

X. HoangVan 연구팀[8]은 캐니 엣지 검출을 이용한 전처리와 문자 검출을 위한 여러 딥러닝 모델을 활용하여 영수증에서 적절한 정보를 추출할 수 있는 OCR 기술을 제안하였다. 하지만, 이 연구 역시 정형화된 한글을 대상으로 한다는 특징이 있다.

강태휘 연구팀[9]은 초성, 중성, 종성으로 나누는 한글의 특징에 기반하여 객체 검출 기술을 활용해 초성, 중성, 종성을 따로 인식하는 모델을 제안하였다. 이 방법은 기존의 1만여 개에 가까운 글자를 분류하는 문제를 50여 개의 자음과 모음을 분류하는 문제로 간소화하여 한글의 인식률을 높였다.

김건욱 연구팀[10]도 [9]와 같이 초성, 중성, 종성으로 나누어 한글을 인식하는 방법을 제안하였다. 이 연구는 한글 외래어를 인식하기 위해 트랜스포머 구조를 활용하며, CNN을 이용한 인식 방법보다 더 높은 정확도를 보이는 것을 확인하였다.

3. 한글 메뉴판 인식을 위한 제안 방법

3.1 특징 추출 개선을 위한 구조

EasyOCR에서 처리 과정은 크게 세 단계로 구성된다. 첫 번째는 이미지 전처리 단계로, 이미지에서 노이즈를 제거하고 문자가 더욱 명확하게 인식될 수 있도록 조정한다. 두 번째 단계는 특징 추출과 문자 인식으로, 문자를 추출하고 해석한다. 마지막 단계는 인식된 문자의 후처리로, 인식 결과의 오류를 최소화하고 최종 문자 출력을 정제하는 것이다.

일반적인 EasyOCR의 인식 단계에서는 VGG-LSTM-CTC 과정을 통해 특징을 추출하고 인식한다. 그러나 이러한 프로세스는 한글 문자에 대한 인식률이 낮다는 단점이 있다. 따라서 본 논문에서는 이를 개선하기 위한 새로운 구조를 제안한다. 본 논문에서는 문자 영역이 감지되었다는 가정 하에 인식 정확도를 향상시키는 것에 초점을 맞춘다.

그림 1은 제안된 방법의 전체 과정을 나타낸다. 먼저, 텍스트 영역이 검출되었다는 가정 하에 해당 영역의 이미지를 입력으로 받아 제안된 모델을 통해 특징 추출 단계를 수행한다. 이 과정에서 추출된 특징 맵은 연속된 Bi-LSTM의 입력으로 사용되며, 이를 통해 시퀀스 레이블링 과정을 거쳐 벡터를 생성한다. 생성된 벡터는 Attention 메커니즘을 사용하여 시퀀스 내에서 중요한 부분에 가중치를 부여함으로써 텍스트 인식 과정에서 특정 부분의 중요성을 강조하여 인식 정확도를 개선하도록 한다.

제안하는 모델은 VGG와 ResNet 모델의 장점을 결합하여 특징 추출 능력을 향상시키기 위해 설계된 구조이며 표 1과 같다. 이 구조는 크게 3개의 VGG 블록과 1개의 Residual 블록, 그리고 마지막 부분에 있는 컨볼루션 층들로 구성된다. VGG 블록을 통해 로컬 특징을 추출하고, Residual 블록을 통해 깊은 특징을 효과적으로 학습하는 방식이다. 이러한 접근 방식은 VGG 모델의 컨볼루션 층과 ResNet 구조의 연결을 결합하여 특징 추출의 성능을 향상시킬 수 있다.

VGG 블록은 두 개의 컨볼루션 층과 하나의 맥스 풀링 층으로 구성된다. 각 컨볼루션 층은 3x3 커널을 사용하고, 제로 패딩의 크기는 1로 설정하여 입력과 출력의 공간 차원이 유지되도록 한다. 각 컨볼루션 층 뒤에는 ReLU 활성화 함수를 적용하여 비선형성을 추가한다. 마지막으로, 맥스 풀링 층이 적용되어 특징 맵의 공간 차원을 줄임으로써 연산량을 감소시키고, 모델이 더 높은 차원의 특징을 학습할 수 있도록 한다.

Residual 블록은 두 개의 3x3 컨볼루션 층과 배치 정규화 층, ReLU 활성화 함수로 구성된다. 첫 번째 컨볼루션 층은 스트라이드를 2, 제로 패딩의 크기를 1로 설정하여 가로, 세로의 길이를 절반으로 줄임과 동시에 설정한 채널 수로 맞추기 위해 사용된다. 그리고 배치 정규화와 ReLU 활성화 함수가 뒤따른다. 두 번째 컨볼루션 층은 스트라이드를 1로 설정 후 배치 정규화와 ReLU 활성화 함수를 사용한다. 독립적으로 구성된 다음 컨볼루션 층은 Residual 블록의 처음 입력에 대하여 스트라이드는 2로 설정한 것은 동일하지만, 패딩을 제거하고 커널 크기를 1로 한다. Residual 블록에서 첫 번째, 두 번째 컨볼루션 층에 의한 연산의 특징 맵의 결과와 세 번째 컨볼루션 층에 의한 결과의 크기는 같다. 두 특징 맵의 값들을 더하는 Elements-wise Addition 과정을 통해 기울기 소실 문제를 완화하고 효과적으로 학습할 수 있게 한다.

마지막 부분에서는 세 개의 컨볼루션 층과 배치 정규화, ReLU 활성화 함수가 사용된다. 마지막 컨볼루션 층은 스트라이드가 (2, 1)로 설정하여, 출력 특징 맵의 차원을 축소시킨다. 이는 뒤에서 진행할 BiLSTM 모델의 입력을 위해 차원을 조정하는 것이다.

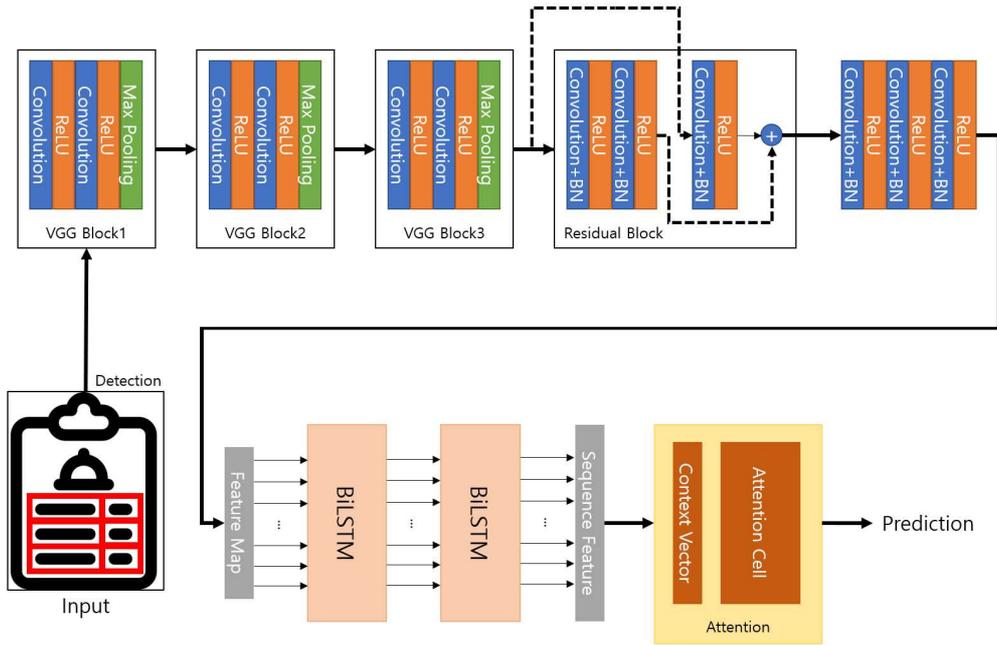


그림 1. 한글 메뉴판 인식을 위한 제안 방법의 전반적인 과정
 Fig. 1. Overall process for recognizing Korean menu

표 1. 특징 추출 부분의 제안하는 모델 구조

Table 1. Proposed model structure of feature extraction part

Block	Layer	Stride	Zero Padding Size	Kernel Size	Activation Function	Batch Normalization	Input Size	Output Size
VGG Block1	Convolution	(1, 1)	(1, 1)	(3, 3)	ReLU	×	[3, 224, 224]	[64, 224, 224]
	Convolution	(1, 1)	(1, 1)	(3, 3)	ReLU	×	[64, 224, 224]	[64, 224, 224]
	MaxPooling	(2, 2)	0	(2, 2)	×	×	[64, 224, 224]	[64, 112, 112]
VGG Block2	Convolution	(1, 1)	(1, 1)	(3, 3)	ReLU	×	[64, 112, 112]	[128, 112, 112]
	Convolution	(1, 1)	(1, 1)	(3, 3)	ReLU	×	[128, 112, 112]	[128, 112, 112]
	MaxPooling	(2, 2)	0	(2, 2)	×	×	[128, 112, 112]	[128, 56, 56]
VGG Block3	Convolution	(1, 1)	(1, 1)	(3, 3)	ReLU	×	[128, 56, 56]	[256, 56, 56]
	Convolution	(1, 1)	(1, 1)	(3, 3)	ReLU	×	[256, 56, 56]	[256, 56, 56]
	MaxPooling	(2, 2)	0	(2, 2)	×	×	[256, 56, 56]	[256, 28, 28]
Residual Block	Convolution	(2, 2)	(1, 1)	(3, 3)	ReLU	○	[256, 28, 28]	[512, 14, 14]
	Convolution	(1, 1)	(1, 1)	(3, 3)	ReLU	○	[512, 14, 14]	[512, 14, 14]
	Convolution	(2, 2)	0	(1, 1)	ReLU	○	[256, 28, 28]	[512, 14, 14]
	Element-wise Addition	×	×	×	ReLU	×	[512, 14, 14]	[512, 14, 14]
-	Convolution	(1, 1)	(1, 1)	(3, 3)	ReLU	○	[512, 14, 14]	[512, 14, 14]
	Convolution	(1, 1)	(1, 1)	(3, 3)	ReLU	○	[512, 14, 14]	[512, 14, 14]
	Convolution	(2, 1)	(1, 1)	(3, 3)	ReLU	○	[512, 14, 14]	[512, 7, 14]

위에서 언급한 바와 같이 3개의 VGG 블록과 1개의 Residual 블록, 그리고 컨볼루션 연산을 선형적으로 연결하여 사용한다. VGG 모델의 경우 지역적 특징을 추출하며 여러 층을 거치면서 더 복잡한 패턴을 학습하는 장점이 있다. 그리고 ResNet의 경우 Residual 연결 구조에서 기울기 소실 문제를 해결하고 더 정교한 특징을 학습하는데 도움을 준다. 따라서, 이러한 장점을 가진 구조를 병합한 제안하는 모델은 VGG와 ResNet의 장점을 결합하여 이미지의 지역 및 전역 특징을 효과적으로 추출할 수 있다.

3.2 문자 인식의 개선을 위한 방안

이렇게 추출된 특징 맵은 BiLSTM을 사용하여 시퀀스 모델링하는 과정을 거친다. Final Block을 통해 생성된 결과를 BiLSTM의 입력으로 하기 위해 차원을 [b, c, h, w]에서 [b, w, c, h]로 변경하고 Adaptive Average Pooling을 적용한다. 사용한 BiLSTM은 2번 연속 진행되며, 입력 크기와 출력 크기는 모두 동일하게 256으로 설정하며 히든 층의 크기는 512로 고정한다. 이를 통해 연속된 문자의 문맥을 이해하여 정확도를 향상시킬 수 있다.

BiLSTM을 통해 추출된 특징을 이용하여 문자를 인식하기 위해 기존의 CTC 방식이 아닌 어텐션 메커니즘을 적용한다. 먼저, 정의한 어텐션 셀은 이전 은닉 상태와 인코더의 출력, 그리고 현재 타임 스텝의 one-hot 인코딩된 문자를 입력으로 받는다. 그리고 입력과 이전 은닉 상태를 변환하고, 이를 합쳐 score를 계산 후 소프트맥스 함수를 통해 정규화한다. 이어서 어텐션 가중치를 생성하며, 이 가중치는 입력 텐서와 결합되어 문맥 벡터를 형성한다. 이 문맥 벡터는 최종적으로 LSTM 셀에 입력되어 새로운 은닉 상태를 생성하게 된다. 이렇게 생성된 어텐션 셀을 활용하여 디코더는 이전의 특징 벡터를 사용하여 최종 출력을 생성한다. 각 타임 스텝마다 어텐션 메커니즘을 사용하여 문맥 벡터를 계산하고, 이를 통해 다음 문자의 확률 분포를 예측한다. 예측된 분포에서 가장 높은 확률을 가진 하나를 최종 예측값으로 표현한다.

4. 실험 및 결과

4.1 데이터 세트

본 논문에서는 학습을 위해 AI-Hub에서 제공하는 한국어 글자체 이미지[11]에서 한글 실사 데이터인 "04.Text in the wild_230209_add"를 사용하였다. 이 중 4개의 카테고리(book, goods, signboard, traffic)를 선정하여 학습 및 검증에 활용하였다. 각 카테고리에 대한 데이터는 동일한 비율로 학습 과정에 포함하였다. 모든 실험은 메뉴가 검증되었다는 가정하게 진행되므로, 메뉴판에서 인스턴스만을 추출하여 데이터로 활용하였다. 테스트를 위해서는 카메라를 이용하여 수집한 메뉴판 데이터를 사용하며, 평가를 위해 수동으로 구축하였다. 표 2는 실험에서 Train, Validation, Test를 위한 인스턴스의 개수를 나타낸다. 구축한 메뉴판의 인스턴스는 총 930개로, 학습과 검증에 사용된 인스턴스가 전체의 99% 이상을 차지한다. 이를 통해 다양한 한글 문자 데이터를 효과적으로 구분할 수 있도록 한다.

표 2. 메뉴 인스턴스의 개수

Table 2. Number of menu instances

	Datasets	Instances
Train	Text in the wild	1,596,283
Validation	Text in the wild	82,479
Test	Self-built	930

4.2 실험 환경 및 세부 사항

본 논문에서는 Ubuntu 18.04 운영 체제를 사용하는 환경에서 진행되었으며, 빠른 연산을 위해 NVIDIA RTX 3090 그래픽 카드 2대를 사용하였다. Pytorch는 1.10.1 버전, Python은 3.9 버전으로 프로젝트를 구축하였다. 총 클래스 수는 412개로, 다양한 한글 문자와 기호를 포함하고 있으며, 메뉴판의 문자 중 빈도가 높은 글자들을 선정한 것이다.

모델의 입력으로는 검출된 문자 영역이며, 입력 영상의 크기는 높이 32픽셀, 너비 100픽셀로 설정하며, 그레이스케일 영상으로 처리한다. 배치 크기는 2048로 설정하고 학습은 총 50,000번의 반복에 걸쳐 수행

하였다. 그리고 1,000번의 반복마다 검증 데이터셋을 사용하여 모델의 성능을 평가하였다. 최적화 알고리즘으로는 Adam Optimizer를 채택하며, 초기 학습률은 1, 그리고 β_1 과 β_2 는 일반적으로 많이 사용되는 0.9, 0.999로 설정하였다.

4.3 성능 평가 및 결과

본 논문에서는 성능 평가를 위해 Normalized Edit Distance, Accuracy, Confidence Score의 3가지 지표를 통해 제안 방법의 타당성을 검증한다.

Normalized Edit Distance는 예측된 문자와 실제 문자 간의 최소 수정 횟수를 의미한다. 이는 문자를 삽입, 삭제, 교체하는 데 필요한 연산 수를 의미하는 지표이다. 이렇게 계산된 지표를 정규화하여, 예측된 문자와 실제 문자 길이 중 더 긴 것에 대한 편집 거리의 비율로 나타낸다. Accuracy는 실제 문자 레이블과 비교하여 정확도를 계산한다. 이 정확도는 실제와 예측 문자가 완전히 일치하는 비율로 정의한다. 마지막으로 Confidence Score는 모델이 처리한 출력에 대한 확률값을 의미하며, 이 값이 높을수록 신뢰도가 높다고 평가할 수 있다. 평가할 3가지 지표의 특성에 따라 Normalized Edit Distance는 값이 낮을수록, Accuracy와 Confidence Score는 높을수록 우수하다는 의미이다. 위에서 언급한 3가지의 지표를 전체 실험 데이터를 통해 평균한 값으로 분석하며 성능 평가의 결과는 표 3과 같다.

먼저, Normalized Edit Distance로 분석한 결과를 살펴보면, EasyOCR의 경우 Normalized Edit Distance 값이 87.104로 나타났다. 반면, CTC를 사용한 제안한 모델은 77.176으로 이 값을 크게 낮추었으며, 어텐션 메커니즘을 도입한 경우에는 86.016으

로 EasyOCR보다 높은 값으로 나타났다. 이는 제안한 모델이 기존 EasyOCR 모델보다 특징을 추출하는 구조가 우수하다는 것을 의미한다. 제안 방법에서는 VGG-ResNet 구조와 어텐션 메커니즘을 조합한 구조의 중요성을 강조하고 있지만, 제안 모델과 CTC의 조합이 더 우수한 것으로 나타났다. 이는 예측한 문자와 실제 문자 간의 모든 수정 횟수를 포함하기 때문에, 일부 부분적 오류가 전체 수정 횟수에 영향을 미쳐 어텐션 메커니즘을 사용한 경우의 Normalized Edit Distance 값이 높아진 것으로 분석된다. 기존 모델과 비교하였을 때 제안 모델과 어텐션 메커니즘의 조합은 약 1% 향상된 수치로 나타난다.

Accuracy 측면에서, EasyOCR 모델은 67.742의 정확도를 보였으며, 제안 모델과 CTC의 조합은 56.022로 더 낮은 정확도를 보였다. 그러나, 어텐션 메커니즘을 통해 예측한 제안 모델은 71.290으로 기존 모델과 비교했을 때 약 3.5% 더 높은 수치를 기록하였다.

Confidence Score를 통해 살펴보면, EasyOCR 모델의 점수는 68.457이었으나, CTC를 사용한 제안 모델은 68.513으로 약간 높은 수치를 기록하였다. 특히, 어텐션 메커니즘을 사용한 제안 모델은 69.414로 가장 높은 신뢰 점수를 기록하였다. 이는 기존 모델과 비교했을 때 약 1% 향상된 수치이다.

종합적으로 볼 때, 제안한 모델은 EasyOCR 모델보다 우수한 성능을 나타낸다. 특히, 특징 추출의 성능을 높이기 위해 제안한 VGG-ResNet 구조와 어텐션 메커니즘을 도입한 모델은 Accuracy와 Confidence Score에서 크게 향상된 결과를 보였다. 이는 제안한 모델이 한글 문자의 인식 성능을 효과적으로 개선하였음을 나타낸다. 그림 4는 제안 방법을 통해 한글 메뉴판을 인식한 예시를 나타내며, 높은 정확도로 인식하고 있음을 확인할 수 있다.

표 3. 구축한 데이터셋에 대한 성능 평가

Table 3. Performance evaluation with constructed datasets

Method	Feature Extraction	Prediction	Normalized Edit Distance ↓	Accuracy ↑	Confidence Score ↑
EasyOCR[3]	VGG	CTC	87.104	67.742	68.457
Ours		CTC	77.176	56.022	68.513
		Attention	86.016	71.290	69.414



그림 4. 제안 방법을 이용한 구축한 한글 메뉴판의 인식 예시
 Fig. 4. Recognition example of constructed Korean menu using proposed method

5. 결론 및 향후 과제

본 연구에서는 VGG와 ResNet의 구조적 장점을 결합하고 어텐션 메커니즘을 도입하여 메뉴판의 한글 문자 인식의 정확도를 향상시키는 모델을 제안하였다. 실제 메뉴판 데이터를 수집하여 사용한 성능 평가에서, 제안된 모델은 기존 EasyOCR 모델에 비해 전반적으로 우수한 인식 능력을 보여주었다.

Normalized Edit Distance, Accuracy, Confidence Score의 3가지 지표를 통해 분석한 결과, 제안된 모델은 EasyOCR보다 전반적으로 향상된 성능을 기록하였다. 구체적으로, Normalized Edit Distance에서는 제안된 모델이 86.016으로 EasyOCR의 87.104에 비해 개선된 결과를 보였다. 또한, Accuracy에서는 71.290으로 EasyOCR의 67.742보다 약 3.5% 높은 성능을 나타냈다. Confidence Score에서도 제안된 모델이 69.414로 EasyOCR의 68.457보다 약 1% 높은 점수를 기록하였다.

이러한 결과는 O2O 상점에서 메뉴판의 자동 인식 가능성을 입증하며, 디지털 메뉴 관리 시스템에 효과적으로 통합될 수 있음을 시사한다. 향후 연구에서는 제

안된 모델의 경량화와 다양한 언어 데이터셋에 대한 확장성을 평가하는 것이 필요하다.

그러나 O2O 상점의 실제 환경 적용을 위해서는 단순한 문자 인식을 넘어서 메뉴와 가격 정보를 연계하여 인식하는 기능이 필요하다. 현재 제안 방법은 일상 사용에 적합한 높은 수준의 정확도를 달성하지 못했기 때문에, 정확도를 더 향상시켜야 한다. 이는 특히 복잡한 배경이나 다양한 폰트의 스타일을 포함하는 환경에서 정확하고 신뢰할 수 있는 인식 결과를 보장하기 위함이다.

REFERENCES

- [1] J. Si, D. Kim and S. Kim, "Automation of Online to Offline Stores: Extremely Small Yolov8 and Feature-Based Product Recognition", Journal of Korea Institute of Information, Electronics, and Communication Technology, Vol 17, No. 3, pp.121-129, 2024.
- [2] S. Lee, J. Si, and S. Kim, "Fine-tuning EasyOCR for Unstructured Korean Menu Recognition in O2O Stores", Proceedings of

KIIT Conference, pp. 476-477, 2024.

[3] EasyOCR, <https://github.com/jaidedAI/EasyOCR>

[4] Tesseract-OCR, <https://github.com/tesseract-ocr/tesseract>

[5] Keras-OCR, <https://github.com/faustomorales/keras-ocr>

[6] 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, Vol. 39, No. 11, pp. 2298-2304, 2016.

[7] H. Jang, S. Goh, J. Lee, and S. Park, "The Deep Learning-Based OCR System for Korean Word with Web Search Engine", The Journal of Korean Institute of Communications and Information Sciences, Vol. 48, No. 9, pp. 1169-1174, 2023.

[8] X. HoangVan, P. TranQuang, M. DinhBao, and T. VuHuu, "Developing an OCR Model for Extracting Information from Invoices with Korean Language", International Conference on Advanced Technologies for Communications (ATC), pp. 84-89, 2023.

[9] T. Kang, E. Kim, M. Park, Y. Yoon, Y. Jo, and H. Min, "Development of an OCR recognition model using the structure of initial, middle and final consonants in Hangeul", Proceedings of ICROS Conference, pp. 68-69, 2023.

[10] G. Kim, J. Son, and J. Min, "Character decomposition to resolve class imbalance problem in Hangul OCR", Proceedings of IEE Conference, pp. 1087-1090, 2022.

[11] <https://www.aihub.or.kr/aihubdata/data/view.do?currMenu=&topMenu=&aihubDataSe=data&dataSetSn=81>

저자약력

시 종 욱 (Jongwook Si)



- 2020년 8월 : 국립금오공과대학교 컴퓨터공학과(공학사)
- 2022년 2월 : 국립금오공과대학교 컴퓨터공학과(공학석사)
- 2022년 3월~현재 : 국립금오공과대학교 컴퓨터·AI융합 공학과 대학원(박사과정정수료)
- 2023년 9월~현재 : 국립금오공과대학교 인공지능공학과 강사

〈관심분야〉 : 컴퓨터비전, 영상처리, 이상감지, 생성형 AI, 딥러닝

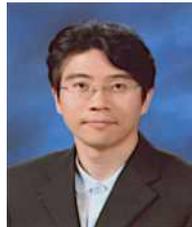
이 상 진 (Sangjin Lee)



- 2020년 3월~현재 : 국립금오공과대학교 컴퓨터공학과 재학

〈관심분야〉 : 문자 인식, 딥러닝, 인공지능

김 성 영 (Sungyoung Kim)



- 1994년 2월 : 부산대학교 컴퓨터공학과(공학사)
- 1996년 2월 : 부산대학교 컴퓨터공학과(공학석사)
- 2003년 8월 : 부산대학교 컴퓨터공학과(공학박사)
- 2004년~현재 : 국립금오공과대학교 컴퓨터공학과 교수

〈관심분야〉 : 영상처리, 컴퓨터비전, 기계학습, 딥러닝, 메타버스