

# 딥러닝 기반 광학 문자 인식 기술 동향

## Recent Trends in Deep Learning-Based Optical Character Recognition

민기현 (G. Min, ghmin@etri.re.kr)	광ICT융합연구실 선임연구원
이아람 (A. Lee, al246@etri.re.kr)	광ICT융합연구실 연구원
김거식 (K.S. Kim, keosikis@etri.re.kr)	광ICT융합연구실 책임연구원
김정은 (J.E. Kim, j.kim@etri.re.kr)	광ICT융합연구실 선임연구원
강현서 (H.S. Kang, hskang87@etri.re.kr)	광ICT융합연구실 책임연구원/실장
이길행 (G.H. Lee, ghlee@etri.re.kr)	호남권연구센터 책임연구원/센터장

### ABSTRACT

Optical character recognition is a primary technology required in different fields, including digitizing archival documents, industrial automation, automatic driving, video analytics, medicine, and financial institution, among others. It was created in 1928 using pattern matching, but with the advent of artificial intelligence, it has since evolved into a high-performance character recognition technology. Recently, methods for detecting curved text and characters existing in a complicated background are being studied. Additionally, deep learning models are being developed in a way to recognize texts in various orientations and resolutions, perspective distortion, illumination reflection and partially occluded text, complex font characters, and special characters and artistic text among others. This report reviews the recent deep learning-based text detection and recognition methods and their various applications.

**KEYWORDS** 광학 문자 인식, 딥러닝 문자 검출, 딥러닝 문자 인식, 인공지능

## 1. 서론

광학 문자 인식(OCR: Optical Character Recognition)은 사람이 쓰거나 인쇄한 문서, 촬영된 사진이나 스캔된 이미지 내의 문자를 인식하여 기계가 읽고 편집할 수 있는 디지털 텍스트로 변환하는 기술을 의미한다. 이것은 법률, 금융, 우편, 병원 등 기록 자료

가 있는 모든 분야에서 아날로그 기록문서의 디지털화를 위해 요구하는 핵심 기술로서 로보틱스, 산업 자동화, 이미지 탐색, 자동 번역, 자율주행, 스포츠 비디오 분석 등 컴퓨터 비전 기반 응용 분야의 주요 기술이다[1].

현재 문자 인식률과 인식속도의 획기적인 증가는 해당 기술 시장의 급속한 성장을 이끌고 있다. 특히

\* DOI: <https://doi.org/10.22648/ETRI.2022.J.370503>

\* 본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 2022년도 문화기술 연구개발 사업으로 수행되었음[과제명: 인공지능 기반 개방형 한글 고서 번역 및 해석 지원 기술 개발, 과제번호: R2021040267, 기여율: 100%].

코로나19의 장기화로 인한 온라인 콘텐츠 및 스트리밍 서비스 이용률이 증가하고 있는 가운데 도서 시장의 디지털 전환이 가속화되고 있으며, 비대면 서비스의 확장으로 문자인식 기술의 활용분야는 더욱 확장될 것이다.

Allied Market Research에 따르면 글로벌 광학 문자 인식 시장은 2021년부터 2028년까지 연평균 16.7% 성장이 전망된다. 그중 북미 시장은 2025년에 8,519백만 달러로 가장 큰 시장을 차지할 것으로 예상되며, 아시아-태평양 시장은 6,023백만 달러로 가장 큰 연평균 성장률(28.8%, '17년~'25년)을 보일 것으로 전망된다[2].

현재 OCR 기술은 인쇄본과 같이 정형화된 서류 문자들에 대해 높은 인식률을 보이고 있지만, 정형화되지 않은 서류, 손글씨, 일상 이미지 내 문자 등의 다양한 크기와 방향성을 가진 문자, 캡처 각도, 위치, 초점, 화질, 음영 등의 외적 요인에 의해 변형된 문자, 부분적으로 훼손된 문자의 텍스트화를 위한 딥러닝 OCR 기술 개발이 진행되고 있다. 본고에서는 딥러닝 기반 OCR 모델, 연구개발 동향 및 응용 분야 등을 살펴보고, 향후 발전 방향을 검토하고자 한다.

## II. 초창기 OCR 모델

초창기 OCR 기술은 머신러닝 기반 문자의 특징 추출 기법과 모델을 결합하여 문자를 검출하고 인식하였다[1]. 문자 검출 방법은 슬라이딩 윈도우[3]를 이용하는 방법과 연결요소 특징[4]을 이용하는 방법으로 나뉜다. 전자는 일정한 크기의 슬라이딩 윈도우를 사용하여 전체적으로 모두 가능한 문자의 위치와 스케일이 스캔될 수 있도록 이미지 피라미드를 구축하여 문자를 검출하는 것으로, 수평 방향으로 기록된 문자 검출에만 적용 가능하며 다양한

방향성을 가진 문자에서는 낮은 검출 성능을 보인다. 후자는 색, 텍스처, 바운더리, 코너 포인트 등의 요소를 사용하여 SVM(Support Vector Machines), Random forest 등의 머신러닝 분류를 통한 이미지 내 문자 영역을 추출한다[5]. 그러나 이러한 방법은 모호한 글자들을 검출하지 못하는 등 낮은 검출률을 갖는다.

문자 인식은 Histogram of Oriented Gradients와 Scale Invariant Feature Transform 등의 이미지 특성을 사용하여 SVM이나 k-nearest neighbors 등의 머신러닝 기술로 개별 글자를 분류하고 통계적 언어 모델(Statistical Language Model) 또는 시각적 구조 예측(Visual Structure Prediction)의 후처리를 통해 잘못된 분류된 문자들을 제거하였다[6]. 그 후 개별 글자를 단어로 연결하는 과정을 거친다. 이 외에도 단어를 먼저 인식하고 개별 글자를 인식하는 방법이 있다.

그러나 이러한 방법들은 다양한 필체의 손글씨와 일상 이미지 내 문자를 인식하는 정확도가 낮거나 사전에 있는 단어 외의 문자들은 인식할 수 없는 등의 단점을 가진다.

## III. 딥러닝 기반 OCR 모델

### 1. 문자 검출 모델

문자 검출(Text Detection)은 페이지나 이미지에 있는 문자의 위치를 찾는 컴퓨터 비전 기술로서 Bounding Box Regression, Part, Segmentation, Fast Scene Text Detection의 4개의 방법으로 분류될 수 있다[7]. Bounding Box Regression은 문자 영역의 위치를 제공하는 바운딩 박스의 좌표를 Regression으로 구하는 방법이고, Part 기반 방법은 Text Instances의 작은 부분을 연결하여 하나의 단어나 문자 라인을 검출하는 방법이다. Segmentation은 바운딩 박스 정보를 획득하기 위해 픽셀 단위로 예측하는 문자

표 1 문자 검출 방법에 따른 딥러닝 모델

Methods	Models
Bounding box regression	TextBoxes, TextBoxes++, DMPNet, SSTD, RRD, EAST, DeRPN
Part-based methods	SegLink, SegLink++
Segmentation-based methods	Mask TextSpotter, PSENet, TextSnake, Pixellink
Fast scene text detection method	TextBoxes, TextBoxes++, SegLink, RRD, EAST, DBNet, DENet++, CentripetalText(CT)

출처 Reproduced from [7].

를 인식하는 것이고, Fast Scene Text Detection은 정확성과 예측 속도에 중점을 두고 개발하는 방법이다. 각 방법에 따른 주요 모델은 표 1에서 보여준다.

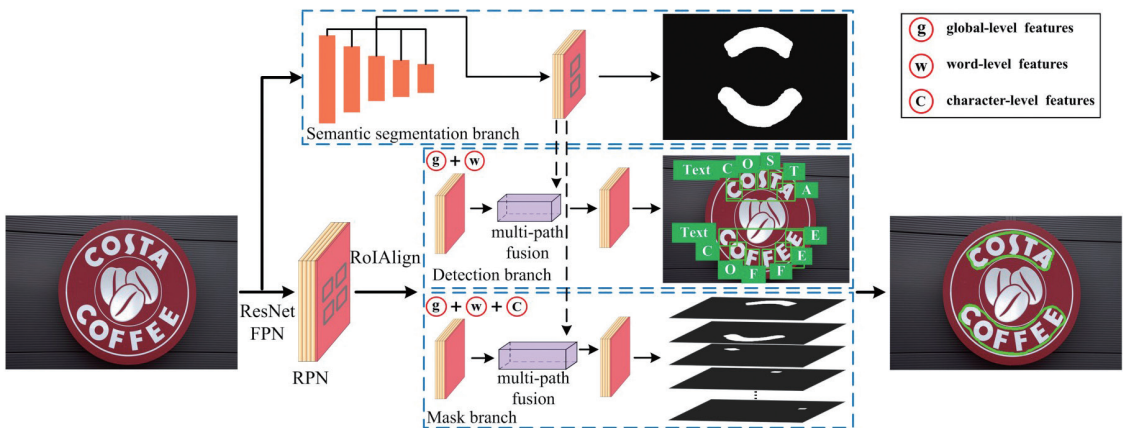
최근 개발된 모델들의 특성은 다음과 같다. TextSnake[8]는 휘어진 문자를 인식하기 위해 FCN (Fully Convolutional Networks)을 기반으로 글자 영역, 중심선, 글자 방향, 후보 반경 등을 예측하였고, Striding 알고리즘과 결합하여 중앙축 점을 추출하

여 개별 글자를 재구성하였다.

Character-Region Awareness For Text detection[9]은 휘어진 문자를 인식하기 위해 FCN 모델을 사용하여 각 개별 글자 영역에 대한 2차원 가우시안 원형의 Score Map을 출력하고 가까운 글자들을 Segmentation으로 그룹화하여 단어들을 검출하였다.

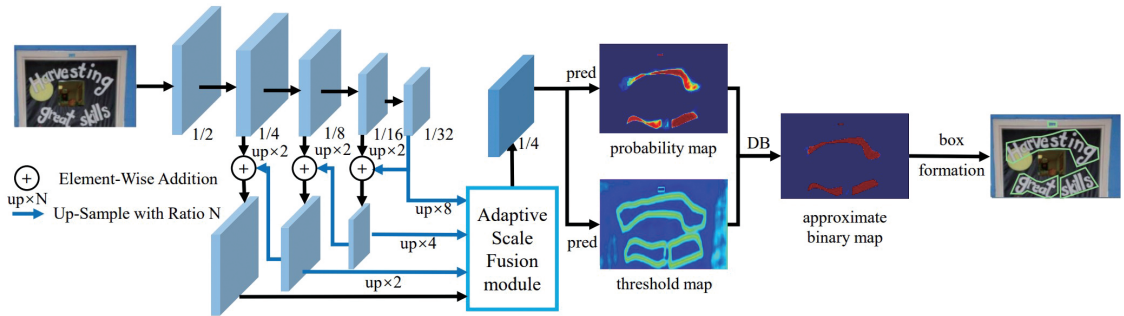
참고문헌[10]에서는 그림 1과 같이 Semantic Segmentation으로 Global-level의 특성을 추출하여 Mask R-CNN 기반으로 구현된 Detection Branch와 Mask Branch를 가이딩하여 Word-level, Character-level의 특징들을 추출하여 문자를 검출하는 TextFuseNet 모델을 제안하였다.

참고문헌[7]에서는 실시간 문자 검출을 위해 DBNet++이 제안되었다. 그림 2에서 보이듯이 다양한 스케일의 글자들이 융합된 이미지에서도 정확한 검출을 위해 Adaptive Scale Fusion Model을 사용하여 Probability Map과 Threshold Map을 추출하였고, Approximate Binary Map이 계산되며 이를 통해 바운딩 박스를 획득한다.



출처 Reprinted with permission from [10], All rights reserved: IJCAI.

그림 1 TextFuseNet 모델 개요도



출처 Reprinted with permission from [7].

그림 2 DBNet++ 모델 개요도

## 2. 문자 인식 모델

문자 인식(Text Recognition) 모델은 검출된 문자가 어떤 글자인지를 판별 후 디지털 텍스트 포맷으로 변환하는 인공지능 모델로서, 개별 글자(Character)를 인식하는 방법과 단어(Word) 단위로 인식하는 방법이 있다. 일반적으로 전자는 다중 클래스 분류 문제에서 클래스 수가 적지만, 단어 단위의 인식 모델에서는 단어 단위로 잘린 이미지와 그 이미지에 정의된 단어의 글자를 텍스트로 전환하기 때문에 분류해야 할 클래스 수가 훨씬 더 많고, 필요한 데이터 수도 많아진다. 그러나 단어 단위의 인식 모델에서는 단어 내 오타자를 포함할 확률이 줄어들게 된다. 예를 들어, “아피트”와 같은 단어에서 “피”를 “파”로 대체하여 인식한 결과를 제공할 수 있다.

이러한 장점들을 활용하여 최근에는 단어 단위로 문자를 인식하는 모델들이 개발되고 있다. CNN(Convolutional Neural Networks)을 이용하여 문자 이미지의 특성을 추출하고 LSTM(Long Short-Term Memory) 또는 Gated Recurrent Unit이 있는 Recurrent 모델을 사용하여 문자 인식에 사용한다. 단어는 개별 글자의 시퀀스로 이루어지기 때문에 시퀀스를 예측할 수 있

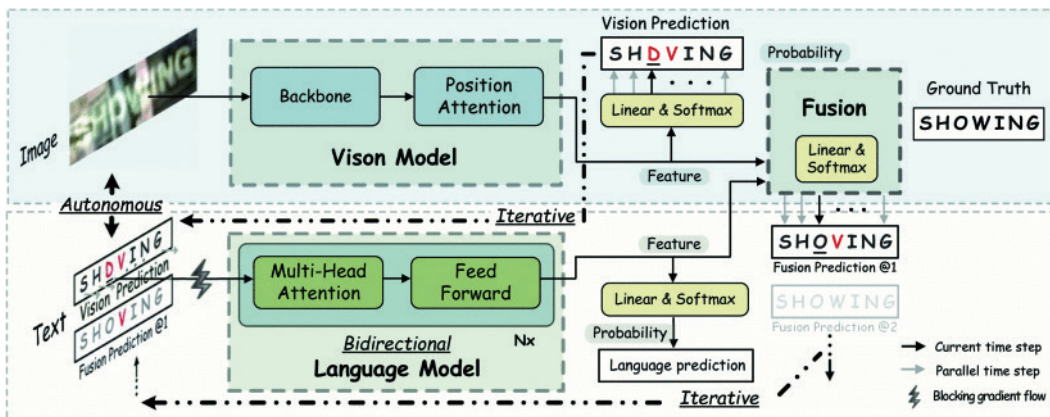
는 CTC(Connectionist Temporal Classification)를 손실 함수로 사용하였다[11]. CTC 기반 문자 인식 방법은 세 종류로 나눌 수 있다[1]. 첫 번째로 글자의 시퀀스를 고려한 문자 인식을 위해 Rosetta[12]와 같이 CNN과 CTC로 연산 후 입력된 문자 이미지의 문자 인식 결과 도출한다. 두 번째로는 참고문헌[13]과 같이 글자의 맥락을 좀 더 정확히 추출하기 위해 첫 번째 방법에서 CNN과 CTC 사이에 RNN 연산을 추가한다. 즉, 입력된 문자 이미지는 CNN에서 Features를 추출하고 RNN, CTC 연산 과정을 통해 예측된 시퀀스와 타겟 시퀀스의 조건부 확률을 식별한다. 세 번째로는 STAR-Net[14]과 같이 휘어지거나 기울어진 왜곡된 문자를 인식하기 위해 두 번째 방법에서 CNN 전 단계에 RN(Rectification Network)을 삽입하여 문자 형태를 직사각형으로 만들고 문자 인식을 진행한다.

문자 인식의 Decoding 단계에서 Deep Features를 향상시키기 위해 기계번역에 사용된 Attention Mechanism도 적용되었으며, 다섯 종류의 구조로 나눌 수 있다[1]. 첫 번째 모델 구조는 R<sup>2</sup>AM[15]과 같이 규칙적인 문자를 인식하기 위한 프레임워크이다. Encoder에서는 CNN, RNN을 통해 입력

된 문자 이미지의 1D-Features를 추출하여 Decoder에 입력하고, Decoder에서는 1D-Attention, RNN으로 순차적 연산을 통해 문자 인식 결과를 도출한다. 두 번째 모델 구조는 곡선이나 원근 왜곡처럼 규칙적이지 않은 문자까지 인식할 수 있는 구조로서, ASTER[16] 모델처럼 문자 이미지를 사각형 모양으로 변경하는 RN 모듈이 삽입되었다. 따라서 Encoder에서는 RN, CNN, RNN을 통해 입력된 불규칙한 문자 이미지의 1D-Feature를 추출하여 Decoder로 전달하며, Decoder에서는 1D-Attention, RNN의 순차적 연산을 통해 문자 인식 결과가 도출된다. 왜곡된 문자 이미지를 사각형으로 변경하기 위해 ESIR(End-to-end Trainable Scene Text Recognition Network via Iterative Rectification)[17]에서는 Thin Plate Spline Transformation 기반의 반복적인 RN을 사용하는 Line-Fitting Transformation이 사용되었으며, 이것을 통해 곡선이 있는 문자 인식률을 향상시켰다. 세 번째 모델 구조는 Arbitrary Orientation Network[18]와 같이 Encoder에서는 CNN과 Multi-Orientation RNN을 통해 여러 방향의 이미지 특성 시퀀스들이 통합된 1D-Features를 추출한다. Decoder에서는 1D-

Attention과 RNN을 통해 문자 인식 결과를 도출한다. 네 번째 모델 구조는 참고문헌[19]와 같이 입력 이미지는 2D이지만 1D-Feature를 사용하여 정보를 잃어버리는 단점을 극복하기 위해 2D-Attention을 이용한다. Encoder에서 불규칙한 문자 이미지 입력 후 CNN을 통해 2D-Features 도출하고 Vertical Pooling을 통해 추출된 1D-Features는 RNN을 통해 계산된다. Decoder에서 2D-Features는 2D-Attention으로 계산하고 RNN을 통과한 결과와 함께 Decoder에 있는 RNN을 통해 문자 인식 결과를 도출한다. 다섯 번째 모델 구조는 FACLSTM[20]처럼 2D-Attention 대신 Convolutional Attention Network를 2D CNN Features에 적용하여 연산 속도를 향상시켰다. Encoder에서는 CNN을 통해 불규칙한 문자 이미지의 2D-Features 추출 및 Decoder에 입력하고, Decoder에서는 Convolutional Attention Decoder를 통해 문자 인식 결과를 도출한다.

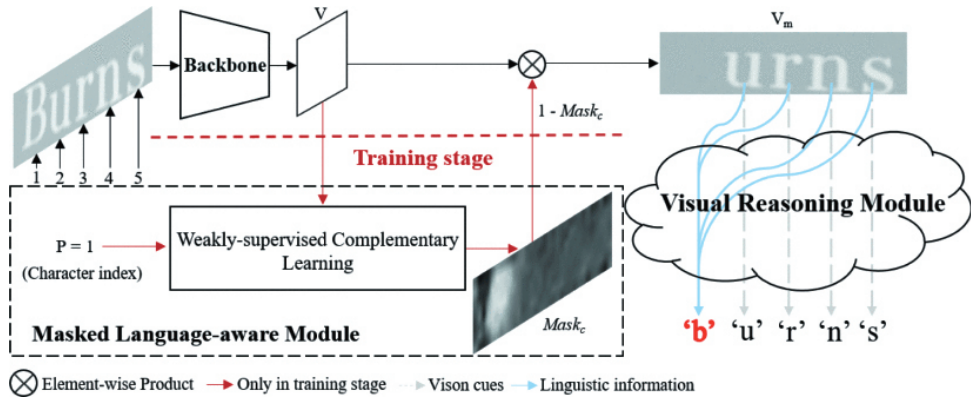
그림 3에서 보이듯이 Transformer 기반으로 개발된 ABINet(Autonomous, Bidirectional and Iterative Language Modeling)[21] 모델에서는 예측 결과를 반복하여 낮은 퀄리티의 이미지라도 문자 인식을 가



출처 Reprinted with permission from [21].

그림 3 ABINet 모델 개요도





출처 Reprinted with permission from [22].

그림 4 VisionLAN 모델 개요도

능하게 한다.

참고문헌[22]에서는 부분 훼손된 문자를 높은 속도와 인식률을 위해 언어 능력을 가진 Vision 모델을 사용하여 비전과 언어 정보를 동시에 제공하는 Visual Language Modeling Network 모델이 제안되었다. Train Stage에서는 MLM(Masked Language-aware Module) 기반으로 Linguistic Information을 캡처할 수 있는 Vision Model을 만들고, Test Stage에서는 Vision Model만 사용한다. MLM은 입력 이미지에 Weakly-supervised Complementary learning으로 Character-wise 마스크 맵을 생성하고 Transformer Unit 기반으로 글자를 예측하는 기능을 한다. 그림 4는 이 모델의 파이프라인을 보여준다.

### 3. End-to-end 모델

문자 검출과 인식을 동시에 제공하는 End-to-end 방법은 최종 문자 인식의 결과에 영향을 주는 문자 검출과 문자 인식의 기술을 동시에 향상시킬 수 있는 장점이 있다.

TextSpotter[23] 모델에서는 Decoding 과정에서

개별 글자들의 Attention 향상을 통해 LSTM의 정확도를 높였다.

TextDragon[24] 모델은 왜곡된 글자 영역 보정에 특화되고 관심 영역을 구별할 수 있는 Slide Operator RoISlide를 활용하여, 왜곡된 글자 형태를 잘 처리할 수 있는 End-to-end 모델이다.

Mask TextSpotter[25] 모델은 Anchor-free하며, 임의의 글자 형태에 적용할 수 있는 Segmentation Proposal Network로 문자를 검출하고 인식하였다.

## N. 국내외 OCR 기술 현황

국외에서 상용화된 대표적인 문자 인식 서비스로는 Google사(미국)의 Vision API와 ABBYY사(러시아)의 Fine Reader가 있다. Vision API의 경우 LSTM 알고리즘 기반의 딥러닝을 통해 50개 이상의 언어 인식이 가능하며[26,27], Fine Reader의 경우 198개에 이르는 언어를 지원한다[28,29]. 두 서비스 모두 영어와 숫자 인쇄체 및 필기체를 대상으로 95% 이상의 인식률을 제공하지만, 아시아권 언어(한국어/중국어/일본어 등)의 필기체에 대해서는 인식률이 저조

한 실정이다. 반면 IntSig사(중국)의 경우 한-중-일 동양권 언어 인식에 CNN, ANN 알고리즘을 적용하여 인식을 개선에 추가적인 성과를 거두고 있다. 국내에는 명함스캐너 앱(CamCard) 개발로 알려져 있다[30].

서구권의 다언어 사용률에 비해 한국 내의 언어 사용은 한글과 한자에 집중되어 있으며, 인식 기술 개발 또한 언어 종류의 다양화보다는 상기 두 언어에 대한 인식 정확도 증가와 클라우드/모바일 기반의 다양한 서비스 제공에 집중되어 있다. 대표적인 서비스로 네이버클라우드의 ClovaOCR은 자연어 처리 기술(Natural Language Processing)을 기반으로 높은 한글 필기체 인식률을 달성하였으며, 대표적인 OCR 기술 경연인 ICDAR에서 2019년 4개 분야 1위를 기록하였다[31]. 셀바스는 도메인학습 딥러닝 기반의 모바일용 고속 OCR 엔진을 개발하여 0.5초 이내에 신분증, 신용카드, 여권 등에서 문자 정보를 추출하여 자동으로 DB를 구축하는 Selvy OCR 서비스를 제공하였다[32]. 이외에도 사이냅 소프트는 모바일 기반 촬영에 따른 이미지 왜곡/회전/빛반사/음영 보정용 전처리 알고리즘을 개발 중에 있으며, 인식된 문자 중 개인정보를 자동으로 선별/마스킹하는 후처리 기술도 선보이고 있다[33].

## V. OCR 기술의 응용 분야

### 1. 문서/도서 자료 번역

국가 간 산업 교류의 문턱이 낮아짐에 따라 글로벌화된 업무의 특성상 관련 문서들이 다양한 언어로 작성된다. OCR 기술은 이러한 개별 문서들을 디지털화할 수 있고 사용자의 대표 언어로 정보를 제공할 수 있다(그림 5 참고). 이와 유사한 사례로, 한자 기반의 한국 고문서에 대해서도 OCR이 적용되어 번역본 자동 생성을 통한 문화유산의 활용도 상승



출처: 게티이미지뱅크, 무단 전재 및 재배포 금지.

그림 5 문서, 도서, 기록물에 대한 OCR 응용

에 기여하고 있다[34].

그 밖에도 OCR 기술은 인쇄물의 접근성이 부족한 시각장애인들을 위한 전자책 서비스를 지원한다[35]. 점자를 통한 정보전달은 속도가 느릴 뿐만 아니라 실제로 과반수 이상의 시각장애인들이 점자해독 능력을 갖추고 있지 않다. 따라서 OCR 기반의 Text To Speech를 통한 음성지원 전자책 서비스가 각광받고 있다.

### 2. 기록물 데이터베이스화

통신과 DB 기술의 발전에 따라 업무 산출 문서는 전산화를 기본으로 한다. 마찬가지로 기존 오프라인 문서들도 통합적 관리를 위해 전산화가 요구되고 있지만, 수십 년간의 방대한 누적량으로 인해 수작업을 통한 온라인 입력에 많은 시간이 소요된다. 이에 대한 개선책으로 OCR 기술이 도입되었으며

공공/금융기관, 통신사 등 사회 각소에서 가시적인 효과를 보이고 있다[36]. 또한 해당 기술의 적용 대상이 초기 정형화된 공문서에서 영수증, 수표, 도면 등의 사문서까지 확대됨에 다양한 서체, 양식, 또는 훼손 정도에 대한 범용 서비스가 시도되고 있다[37].

「개인정보 보호법」의 강화에 따라 한국인터넷진흥원에서는 DB상에 저장된 디지털 문서에 주민등록번호나 계좌번호 등의 개인정보 포함을 지양하고 있다. 이러한 민감정보 관리의 과정에서 OCR 기술이 적용되어 선택적 블라인드 처리 자동화를 지원하고 있다[38].

### 3. 전문 의료 정보 해석

건강 복지 시스템의 발전을 통해 전 국민을 대상으로 고품질의 서비스가 제공되고 있으며 처방전 발급량이 급증하고 있다. 처방전은 전문 의료용으로 구성되어 있어 개인의 복용에 가이드를 제공하지 못하며 다종의 의료서비스를 동시에 이용하는 경우 의약품 오용의 소지도 있다. 이를 개선하기 위해 스마트 헬스케어 시스템인 Medi-U는 OCR을 통해 처방전 사진을 분석하여 중복처방 등의 위험요소를 예방하는 서비스를 제공하고 있다[39].

### 4. 공공서비스 지원

상수도, 도시가스과 같은 공공서비스들은 사용량에 대한 월별 검침이 각 가정에 설치된 아날로그 미터기에 대한 검침원의 나안 관찰로 이루어지고 있다. 무선통신 모듈형의 검침기술이 최근 개발되고 있지만, 비용적 측면에서 실사용에 어려움이 있다. 따라서 공공서비스 사용량은 매번 장시간의 수작업을 통해 수기나 모바일 기기에 입력되고 있다. 해당 작업에 대한 효율성과 정확성의 증가를 위해 계량



출처: 게티이미지뱅크, 무단 전재 및 재배포 금지.

그림 6 공공서비스에 대한 OCR 응용

기를 촬영한 이미지에 대한 OCR 기술의 적용이 제안되었다[40,41].

또 하나의 공공서비스인 우편시스템에 대해서도 OCR 기술의 선두적인 활용이 전망된다. 우편물 표면에 기입된 발신자/수신자 주소 정보는 인쇄체나 필기체의 조합으로 구성되며, 우편물 자동접수기는 이러한 우편물의 상부정보 촬영 영상을 기반으로 공급서비스의 가격 책정 및 우편물 분류를 수행한다. 앞서 언급된 수도/도시가스 검침기와는 다르게 우편물상에 필기체로 표기된 정보에는 각 개인의 필체에 따른 변인을 극복할 수 있는 인식 기술 개발이 선결과제로 주어진다. 현재까지 문자 분리를 통한 인식 기술이 개발되었고 추가적인 연구로 단어(주소)단위 문자열 추출 방안에 대한 연구가 제시되고 있다[42]. 그림 6에서는 공공서비스에 대한 OCT 적용 사례를 보여준다.



## 5. 교통안전 감시, 자율주행 보조

도로상의 위법차량과 범죄자 동선 추적을 위해 CCTV 영상 활용이 활발히 이루어지고 있다(그림 7 참고). 초기 서비스 단계에서는 교통법규 위법상황 순간 촬영된 단일 이미지에 대해서 인적 검사를 통해 번호판 정보를 추출하였고, 범죄자 동선 추적의 경우도 해당 구역 촬영분의 전수 검사를 거쳤다. 이러한 과정은 인적 판단 오류나 영상 자체의 노이즈로 인해 정보의 손실이 발생하며 검출의 과정에 많은 시간이 소요되었다[43]. 추후 딥러닝 기반 OCR의 도입으로 해당 기술의 전자동화를 이루었으며 기후 악조건, 모션블러 등으로 인해 훼손된 영상에 대해서도 필요정보 획득을 위해 딥러닝 기반 패턴 복원을 활용한 인식을 개선 연구가 ETRI를 필두로 수행되고 있다[44].

차세대 교통기술로써 두각을 나타내고 있는 자율주행 개발에서도 OCR은 핵심 개발요소로 자리 잡고 있다[45](그림 7 참고). GPS 기반 내비게이션 정보에



출처 게티이미지뱅크, 무단 전재 및 재배포 금지.

그림 7 교통안전관리에 대한 OCR 응용

서 제공되지 않는 노상 안전 관련 표지판들은 주행 중 실시간으로 검출 및 인식(Traffic-Sign Recognition)되어야 하므로 고속 주행 환경 내에서 높은 정확도와 연산속도를 모두 요구한다. 이러한 trade-off 관계에 속한 두 가지 성능의 균형적 개발을 위해 Faster R-CNN과 MobileNetV2 등의 경량 딥러닝 모델을 사용하는 연구가 활발히 진행되고 있다[46].

## VI. 결론

본고에서는 문서/고서의 번역, 사회적 약자를 위한 문헌의 자동 음성/점자 제공 서비스, 산업 자동화, 문서의 텍스트화, 개인 의료정보의 디지털화, 공공서비스의 자동화 등 다양한 서비스 실현을 위한 핵심 요소 기술인 딥러닝 기반 OCR 연구 동향에 대해 살펴보았다. 현재 OCR 기술은 정형화된 문자에 한해서 높은 문자 인식률 결과를 제공하지만, 다양한 시각으로 촬영된 문자, 손글씨, 아트문자, 부분 훼손된 문자, 흐릿한 문자 등을 인식하기 위한 기술이 연구되고 있다. CNN 기반의 이미지 특성을 추출하고, 개별 글자의 순서까지 고려한 RNN 기반의 OCR 기술이 개발되고 있다. 최근에는 번역에 사용되고 있는 Transformer를 기반으로 다양한 OCR 기술이 개발되고 있다. 또한 실시간 영상 인식 서비스 제공을 위한 OCR 기술이 개발되고 있다. 전자문서 및 전자거래 기본법 개정안이 2020년 12월부터 국내에서 시행되면서 디지털 문서도 종이 문서와 동일한 수준으로 법적 효력이 인정된다[47]. 따라서 사전에 없는 신생 단어 인식, 모바일/태블릿에서 쉽게 적용할 수 있는 경량 딥러닝 모델 개발을 위한 다양한 연구가 진행되고 있으며, 더욱 열악한 환경에서 촬영된 문서, 훼손되거나 흘려 쓴 글씨, 일부 가려진 문자, 영상 내 문자 등을 실시간으로 정확하게 문자 텍스트로 전환할 수 있는 기술이

개발되고 있다.

ETRI에서는 한문 고서 번역을 대상으로 하는 문자 인식 기술을 개발하고 있다. 해당 기술은 기존 번역 전문가들이 스캔된 한문 고서 이미지의 한자를 직접 타이핑하여 텍스트로 전환하던 노동 집약적 업무를 간소화하는 데 도움이 될 수 있을 것으로 기대된다.

### 약어 정리

ABINet	Autonomous, Bidirectional and Iterative Language Modeling
ANN	Artificial Neural Network
ASTER	Attention Scene Text Recognizer
CNN	Convolutional Neural Network
CTC	Connectionist Temporal Classification
DeRPN	Dimension-decomposition Region Proposal Network
DMPNet	Deep Matching Prior Network
EAST	Efficient and Accurate Scene Text Detector
FACLSTM	ConvLSTM with Focused Attention for Scene Text Recognition
FCN	Fully Convolutional Network
LSTM	Long Short-Term Memory
OCR	Optical Character Recognition
R <sup>2</sup> AM	Recursive Recurrent Neural Networks with attention modeling
RNN	Recurrent Neural Network
RRD	Rotation-sensitive Regression Detector
SegLink	Segment Linking
SSTD	Single Shot Text Detector
STAR-Net	SpaTial Attention Residue Network
SVM	Support Vector Machines

### 참고문헌

- [1] Z. Raisi et al., "Text detection and recognition in the wild: A review," arXiv preprint, CoRR, 2020, arXiv: 2006.04305.
- [2] R. Rake, "Image recognition market," Allied Market Research, 2018.
- [3] A. Bissacco et al., "PhotoOCR: Reading text in uncontrolled conditions," in Proc. IEEE Int. Conf. Comput. Vis., (Sydney, Australia), Dec. 2013, pp. 785–792.
- [4] L. Neumann et al., "A method for text localization and recognition in real-world images," in Proc. Asian Conf. Comput. Vis., (Queenstown, New Zealand), Nov. 2010, pp. 770–783.
- [5] Y. Zhu, C. Yao, and X. Bai, "Scene text detection and recognition: Recent advances and future trends," Front. Comput. Sci., vol. 10, no. 1, 2016.
- [6] Q. Ye and D. Doermann, "Text detection and recognition in imagery: A survey," IEEE Trans. Pattern Anal. Mach. Intell., vol. 37, no. 7, 2015.
- [7] M. Liao et al., "Real-time scene text detection with differentiable binarization and adaptive scale fusion," IEEE Trans. Pattern Anal. Mach. Intell., 2022, p. 1.
- [8] S. Long et al., "Textsnake: A flexible representation for detecting text of arbitrary shapes," in Proc. Eur. Conf. Comput. Vis., (Munich, Germany), Sept. 2018, pp. 20–36.
- [9] Y. Baek et al., "Character region awareness for text detection," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., (Long Beach, CA, USA), June 2019, pp. 9365–9374.
- [10] Y. Ye et al., "TextFuseNet: Scene text detection with richer fused features," in Proc. Int. Joint Conf. Artif. Intell. (IJCAI-20), (Yokohama, Japan), Jan. 2021, pp. 516–522, <https://www.ijcai.org/proceedings/2020/0072.pdf>
- [11] N. Subramani et al., "A survey of deep learning approaches for ocr and document understanding," arXiv preprint, CoRR, 2020, arXiv: 2011.13534.
- [12] F. Borisyuk, A. Gordo, and V. Sivakumar, "Rosetta: Large scale system for text detection and recognition in images," in Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., (London, United Kingdom), July 2018, pp. 71–79.
- [13] 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 Trans. Pattern Anal. Mach. Intell., vol. 39, no. 11, 2016, pp. 2298–2304.
- [14] W. Liu et al., STAR-Net: A SpaTial Attention Residue Network for Scene Text Recognition, Proceedings of the British Machine Vision Conference (BMVC), BMVA Press, 2016, pp. 43.1–43.13.
- [15] C.-Y. Lee et al., "Recursive recurrent nets with attention modeling for OCR in the wild," in Proc. IEEE Conf.

- Comput. Vis. Pattern Recognit., (Las Vegas, NV, USA), June 2016, pp. 2231-2239.
- [16] B. Shi et al., "Aster: An attentional scene text recognizer with flexible rectification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 9, 2018.
- [17] F. Zhan and S. Lu, "Esir: End-to-end scene text recognition via iterative image rectification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, (Long Beach, CA, USA), June 2019, pp. 2059-2068.
- [18] Z. Cheng et al., "Aon: Towards arbitrarily-oriented text recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, (Salt Lake City, UT, USA), June 2018, pp. 5571-5579.
- [19] H. Li et al., "Show, attend and read: A simple and strong baseline for irregular text recognition," in *Proc. AAAI Conf. Artif. Intel.*, vol. 33, no. 1, 2019, pp. 8610-8617.
- [20] Q. Wang et al., "FacIstm: ConvIstm with focused attention for scene text recognition," *arXiv preprint, CoRR*, 2019, arXiv: 1904.09405.
- [21] S. Fang et al., "Read like humans: Autonomous, bidirectional and iterative language modeling for scene text recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, (Virtual), June 2021, pp. 7098-7107.
- [22] Y. Wang et al., "From two to one: A new scene text recognizer with visual language modeling network," in *Proc. IEEE Conf. Comput. Vis.*, (Virtual), Oct. 2021, pp. 14194-14203.
- [23] T. He et al., "An end-to-end textspotter with explicit alignment and attention," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, (Salt Lake City, UT, USA), June 2018, pp. 5020-5029.
- [24] W. Feng et al., "Textdragon: An end-to-end framework for arbitrary shaped text spotting," in *Proc. IEEE Conf. Comput. Vis.*, (Seoul, Rep. Korea), Oct. 2019, pp. 9076-9085.
- [25] M. Liao et al., "Mask textspotter v3: Segmentation proposal network for robust scene text spotting," in *Euro. Conf. Comput. Vis.*, (Glasgow, United Kingdom), Aug. 2020, pp. 706-722.
- [26] <https://cloud.google.com/vision/>
- [27] 황선명, 염희균, "윈도우 기반의 광학문자인식을 이용한 영상 번역 시스템 구현," *사물인터넷융복합논문지*, 제5권 제2호, 2019, pp. 15-20.
- [28] <https://pdf.abbyy.com/>
- [29] A.P. Tafti et al., "OCR as a service: An experimental evaluation of Google Docs OCR, Tesseract, ABBYY FineReader, and Transym," in *Proc. Int. Symp. Vis. Comput. (ISVC)*, (Las Vegas, NV, USA), Dec. 2016, pp. 735-746.
- [30] <https://www.camcard.com/>
- [31] <https://clova.ai/ocr>
- [32] <http://ocr.selvasai.com/>
- [33] <https://www.synapsoft.co.kr/ocr/>
- [34] 민기현, 이아람, 강현서, "인공지능 기반 한문 고서의 한자 검출을 위한 전처리 알고리즘에 관한 연구," *한국통신학회 추계종합학술발표회*, 2021, pp. 597-598.
- [35] 류은주, 문미경, "광학문자인식(OCR)기반 시각장애인용 셀프 E-book," *한국컴퓨터종합학술대회*, 2016, pp. 1801-1803.
- [36] 김인택, 안대진, 이해영, "인공지능을 활용한 지능형 기록관리 방안," *한국기록관리학회지*, 제17권 제4호, 2017, pp. 225-250.
- [37] 임윤지 외, "OCR 기반의 자동 문자인식," *한국소프트웨어종합 학술대회*, 2019, pp. 1318-1320.
- [38] 백종경 외, "전자문서에서 서식인식과 광학문자인식을 이용한 개인정보 탐지 및 보호 시스템," *한국산학기술학회논문지*, 제21권 제5호, 2020, pp. 451-457.
- [39] 이승훈 외, "OCR 기술을 이용한 한글 처방전 문자 인식 시스템," *한국정보과학회 학술발표논문집*, 2017, pp. 362-364.
- [40] 차영화 외, "객체 감지와 광학 문자 인식을 이용한 아날로그 전력 계량기 이미지에서의 숫자 영역 인식," *한국통신학회 학술대회논문집*, 2020, pp. 334-335.
- [41] 이교혁 외, "광심 문자열 인식 기술을 이용한 가스계량기 자동 검침 시스템," *지능정보연구*, 제26권 제2호, 2020, pp. 1-25.
- [42] 김재철 외, "필기체 우편영상 주소인식을 위한 문자 추출 알고리즘," *한국정보과학회 학술발표논문집*, 2017, pp. 1414-1416.
- [43] 장일식 외, "지능형 감시 카메라 동향 및 시나리오 연구," *한국ITS학회논문지*, 제8권 제4호, 2009, pp. 93-101.
- [44] 김건우, "딥러닝 기반 열악 자동차 번호 이미지 복원 및 인식 기술," *주간기술동향*, 2020, pp. 27-32.
- [45] Y. Zhu et al., "Cascaded segmentation-detection networks for text-based traffic sign detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 1, 2018, pp. 209-219.
- [46] R. Ravindran et al., "Traffic Sign Identification Using Deep Learning," in *Proc. Int. Conf. Comput. Sci. Comput. Intell.*, (Las Vegas, NV, USA), Dec. 2019, pp. 318-323.
- [47] 노대경, "광학문자인식," *ASTI Market Insight* 2021-024, 2021.