

딥러닝을 통한 문서 내 표 항목 분류 및 인식 방법

이동석[†], 권순각^{**}

Methods of Classification and Character Recognition for Table Items through Deep Learning

Dong-Seok Lee[†], Soon-Kak Kwon^{**}

ABSTRACT

In this paper, we propose methods for character recognition and classification for table items through deep learning. First, table areas are detected in a document image through CNN. After that, table areas are separated by separators such as vertical lines. The text in document is recognized through a neural network combined with CNN and RNN. To correct errors in the character recognition, multiple candidates for the recognized result are provided for a sentence which has low recognition accuracy.

Key words: Optical character recognition, Image recognition, Deep learning, Intelligent document processing, Convolutional neural network

1. 서 론

디지털 시대로의 돌입으로 인해 문서에서 필요한 정보를 쉽게 찾고 빠르게 접근할 필요성이 증가하고 있다. 이를 위해서는 종이 문서 형태의 데이터를 디지털 문서로 변환하는 것이 필요하다. 종이 문서를 디지털화함으로써 업무 자동화를 달성할 수 있고, 이는 업무의 효율성과 생산성의 증가를 가져온다. 예를 들어 수출입 서류를 처리하는 관세사 업무[1]는 30%가 수출입선적서류의 단순 입력 작업으로 업무에 피로도가 증가하고 있으며, 반복적이고 비생산적인 업무로 인해 업무의 능률이 떨어지는 문제가 있다. 또한 잘못된 문서 처리로 인해 손해가 발생하여 업무에 가해지는 스트레스가 높다. 이로 인해 적지 않은 퇴사자가 해마다 발생하고 있고, 새로운 인력의 충원도

쉽지 않은 현황이다. 이러한 업무를 자동화한다면 사람의 반복적인 업무를 경감시킬 수 있어 업무 효율이 대폭적으로 증가할 것이다.

업무 자동화에서 제일 중요한 기술은 문자 인식 기술(OCR, Optical character recognition)이다. OCR은 스캐너를 사용해 문서를 이미지로 변환한 후, 이미지에 포함된 문자와 기호를 인식해 텍스트 파일이나 규격화된 형태의 문서 파일로 변환해서 저장하는 것이다. 문서 내 문자를 인식하여 디지털화함으로써 서류의 내용을 인식하여 분류할 수 있다. 현재 OCR은 우편번호 인식, 도면 인식, 금융기관에서의 전표, 수표 자동입력 등의 다양한 분야에서 실용화되어 쓰이고 있다. 하지만 기존 OCR은 숫자 '0'과 영문자 'O', 또는 숫자 '3'과 '8' 등의 비슷한 형태를 구별하지 못한다는 단점이 있었다. 최근 인공지능망을 이용한 영

※ Corresponding Author : Soon-Kak Kwon, Address: (47340) Eomgang-ro 176, Busanjin-gu, Busan, Korea, TEL : +82-51-890-1727, FAX : +82-505-182-6905, E-mail : skkwon@deu.ac.kr

Receipt date : May 24, 2021, Approval date : May 26, 2021

[†] AI Grand ICT Research Center, Dong-eui University (E-mail : ulsan333@gmail.com)

^{**} Dept. of Computer Software Engineering, Dong-eui University

※ This research was supported by the BB21+ Project in 2020, and by the MSIT (Ministry of Science and ICT), Korea, under the Grand Information Technology Research Center support program (IITP-2021-2020-0-01791) supervised by the IITP(Institute for Information & communications Technology Planning & Evaluation).

상 인식 기술은 기존 영상 인식 기술과 달리 능동적으로 검출해야할 중요한 특징을 학습을 통해 인식할 수 있다는 장점을 가진다[1-2]. OCR에 인공신경망의 적용을 통해 OCR의 인식 정확도가 급격히 증가하였다[2-6].

OCR은 문서 내 문자를 인식할 수 있지만, 그 내용이 무엇을 의미하는 것인지를 알 수 없다는 한계가 있다. 특히 표 영역의 경우, 표 내부 항목명과 항목 내용을 인식하는 것은 업무 자동화를 위한 문서 처리에 있어서 중요한 부분이다. 하지만 OCR만을 통해서 표 내부의 문자를 인식하는 것에만 국한되어 있고, 해당 내용이 무슨 의미를 가지는지는 인식을 하지 못한다. 따라서 문서 인식을 수행한 후, 사람의 손으로 항목을 분류해야 한다는 불편함이 있다. 또한 표 영역 내 구분자는 문자 인식의 정확도를 떨어뜨리는 요인이 된다.

본 논문에서는 딥러닝 신경망을 통해 표 항목 내의 문자를 인식하고, 이를 통해 문서를 디지털화하는 방법을 제안한다. 먼저 스캔된 문서 이미지 파일에 대해 컨볼루션 신경망(CNN, Convolution Neural Network)을 통해 표 영역을 검출한다. 그 후 표 영역 내 수직선과 수평선의 구분자로 분리된 각 영역을 검출한 후, 각각의 영역에 대해 컨볼루션 신경망과 순환 신경망(RNN, Recurrent Neural Network)이 결합된 신경망을 통해 문자를 인식한다. 이 때 인식된 결과의 후보가 여러 나올 경우나 정확도가 낮은 문장 인식 결과에 대해서는 사용자에게 인식 결과 후보와 원 이미지를 보여줌으로써 잘못된 인식으로 인한 처리 실수를 방지한다. 그 후 인식된 내용을 규격화된 텍스트 파일로 저장하여 다른 응용 프로그램이 이를 활용할 수 있도록 한다.

2. 관련 연구

2.1 인공신경망을 통한 영상 인식

인공신경망은 사람의 신경망을 본뜬 다층의 퍼셉트론 구조를 통해 비선형적인 분류나 회귀 문제를 해결하는 기계학습 방법이다. 퍼셉트론은 입력 값들에 각각의 가중치를 부여하여 더한 합에 대해 시그모이드, Rectified Linear Unit(ReLU) 등의 활성화수를 적용시켜 비선형적인 출력을 다음 계층으로 전달한다. 퍼셉트론을 이용한 인공신경망 구조는 2008년 과

대적합 문제를 해결한 딥러닝 신경망(DNN, Deep Neural Network)의 발견과 컴퓨팅 성능의 발달, 그리고 인터넷의 발달로 인한 학습 데이터양의 급격한 증가로 성능이 급격하게 개선되었다. 인공신경망은 다양한 분류와 회귀 문제를 해결하는데 적용될 뿐만 아니라, 클러스터링 등과 같은 비지도학습과 강화학습 분야에도 적용되고 있다.

영상 기반의 객체 인식을 위해서도 DNN을 사용할 수도 있다. 하지만 DNN의 입력은 1차원의 벡터 형태로, 2차원 구조의 영상을 처리하기에는 적절하지 않다. 모든 화소가 입력 계층의 각각의 퍼셉트론 구조의 노드에 입력되기 때문에 가중치의 개수가 과도하게 많아지기 때문에 신경망의 복잡도가 커진다는 문제점이 있다. 또한 객체의 평행이동, 스케일링, 회전 등의 기하학적 변환이 발생할 경우, 입력 벡터의 많은 요소들이 변하는 문제가 있다. 이는 객체의 정확한 인식을 위해서는 객체의 모든 기하학적인 변환을 학습해야 한다는 의미로, 이는 사실상 불가능하다. 이러한 문제를 해결하기 위해 1998년 Y. LeCun [7]이 합성곱 연산을 통해 영상의 2차원 구조를 유지하는 CNN을 통해 영상을 인식하는 방법을 제안했다. CNN은 기존 DNN에서의 가중치 대신, 정사각형 형태의 커널 또는 필터를 공유한다는 특징이 있다. 입력된 영상에 커널에 대한 합성곱을 계산한 후, 활성화수를 적용하여 출력 영상을 구한다. CNN을 통한 영상 인식 방법은 2012년 AlexNet[8]의 ImageNet Large Scale Visual Recognition Challenge(ILSVRC) 대회 우승 이후로 주목을 받게 되어 현재까지 활발하게 연구되고 있다.

CNN을 통한 단일 객체 인식에서는 다층의 CNN 구조를 통해 영상의 특징 맵을 구한다. 그 후 특징 맵을 1차원으로 변환한 후, 기존 DNN 구조를 통해 해당 객체의 종류를 분류하고, 해당 객체의 영역을 계산한다. 영역 검출은 객체를 둘러싸는 경계박스의 모서리 4점을 찾는 과정으로, 이는 회귀 문제로 간주하여 처리할 수 있다. 하지만 이 방법을 통해 다중 객체를 검출하기 위해서는 각 영상마다 출력 계층에서 출력하여야 하는 출력 값의 개수가 달라야 한다는 문제가 있다. 또한 영상 내 다수의 객체에 대해 고려를 하여야 하므로 복잡도가 커진다는 문제점이 발생한다. 따라서 다중 객체를 검출하기 위해서는 먼저 영상 내 객체가 있을만한 영역으로 분할한 뒤, 각각

의 영역들을 객체 인식을 위한 CNN에 입력하여 각각의 객체를 인식한다. 다중 객체를 위한 영상 인식 CNN은 1-stage과 2-stage 검출 방법으로 분류된다 [9].

2-stage 방법은 객체가 있을 확률이 높은 영역을 검출한 후, 검출된 영역에 대해 정확한 영역 검출과, 객체를 분류하는 과정을 순차적으로 수행하는 방법이다. R-CNN(Region-based CNN)[10]은 객체가 존재할 가능성이 높은 경계박스를 찾기 위해 selective search 알고리즘을 적용한다. selective search 알고리즘은 유사한 색상이나 휘도, 질감을 가지는 인접한 화소들을 연결하여 경계박스를 찾는 방법이다. selective search 알고리즘은 인공신경망과 별도로 수행되는 알고리즘이기 때문에 학습 과정에서 학습이 되지 않는다는 문제가 있다. 또한 selective search 알고리즘을 통해 분할된 영역들에 대해 CNN을 통해 특징 맵을 구하기 때문에 검출 속도가 느리다는 단점이 있다. Fast R-CNN[11]은 한 영상 전체에 대한 특징맵을 구하여 영역 분할 후 전체 특징맵에서 해당 영역에 대한 것만을 사용하기 때문에 속도를 개선시켰다. Faster R-CNN[2]은 영역 분할을 위해 selective search 대신 영역제안 신경망(RPN, Region Proposal Network)를 적용하여 영역 분할 과정을 신경망에 통합하였다. 이를 통해 객체 인식의 전체 과정을 학습시킬 수 있게 되어 정확도와 검출 속도를 개선했다.

1-stage 방법은 영상을 특정 단위의 사각형으로 분할하여 처리하는 것이 특징이다. 이를 통해 2-stage 방법에 필요한 영역 제안 과정이 생략되고, 분할된 각 영역에서 객체 영역 검출과 객체 분류가 동시에 이루어진다. 그 후 각 영역에서 검출된 결과들을 통합되어 최종적인 객체 인식 결과를 구한다. YOLO [12,13]는 R-CNN 방법들과 달리 1-stage 검출기이다. YOLO는 객체 검출문제를 회귀 문제로 정의한다. 입력 영상은 특정한 크기의 셀로 분할된다. 각각의 셀에서 검출할 물체의 신뢰도가 계산된다. non-maximal suppression 알고리즘은 바운드박스가 겹치는 영역을 결정하기 위해 적용된다. YOLO는 객체를 신속하게 검출할 수 있다는 장점이 있다. SSD[14]는 다양한 크기의 셀을 허용함으로써 물체 검출의 효율성과 정확성을 높인다. 1-stage 방법은 2-stage 방법보다 객체 검출 속도가 빠르지만, 객체 검출 정

확도는 떨어진다.

2.2 영상 내 문자 인식 기술

OCR은 문서 이미지에서 문자가 있는 영역들을 검출하는 문자 검출 단계와 각 문자 영역에 대해 문자 종류를 분류하는 문자 인식 단계를 거친다. 딥러닝 신경망 등장 전에는 문자 인식 단계에서 기계 학습을 통해 패턴 매칭을 함으로써 문자를 분류했다. 딥러닝 신경망 이후에는 문자 검출 단계와 문자 인식 단계에 대해 각각에 딥러닝을 적용함으로써 기존보다 인식 성능을 크게 개선시켰다.

딥러닝 신경망을 이용한 OCR에서는 문자를 분류하기 위해서 CNN을 통해 문자 영상의 특징을 검출하여야 한다. 또한 문자들로 이루어진 문장의 특성상 특정 문자의 앞 또는 뒤하고 연관성이 크다는 특징이 있다. 따라서 딥러닝을 통해 문자를 인식하기 위해서는 문자 검출 단계에서 검출된 영역의 이미지에서 특징맵을 추출하는 CNN 계층과 시계열 데이터를 처리하는 신경망인 RNN계층을 통해 문자를 인식한다. B. Shi[15]는 CNN을 통해 특징맵을 추출하고, 해당 특징맵을 RNN의 한 종류인 양방향 장단기 메모리 (Bi-LSTM, Bidirectional Long Short-Term Memory) 구조의 신경망에 입력하여 문자열을 예측하는 방법을 제안했고, J. Wang[16]는 RNN 계층에 LSTM보다 개선된 게이트 순환 유닛(GRU, Gated Recurrent Units)을 적용하여 성능을 개선했음을 보였다. Z. Cheng[17]은 RNN 계층에 입력된 특징맵 중 특정 범위에 대해 가중치를 더 부여해서 정확도를 높이는 Attention 기법을 적용하였다. 또한 문자 검출 단계와 문자 인식 단계를 합쳐서 한 번에 처리하는 방법도 제안되었다. 문자 검출 단계와 문자 인식 단계를 합침으로써 선형으로 정렬된 문자뿐만 아니라 곡선 형태의 문장에 대해서도 인식을 수행할 수 있다. Aster[4]는 선형 형태로 정렬되지 않은 문장이 있는 이미지를 변형하여 문장을 선형으로 정렬하는 네트워크와, 해당 네트워크에서 문자를 인식하는 네트워크를 통해 다양한 형태의 문장에 대해서도 인식을 할 수 있다. TextSpotter[5]는 이미지의 스케일 변화에 강인한 다중스케일 특징 피라미드 신경망(FPN, Feature Pyramid Network)을 적용하여 다양한 스케일의 문장 이미지에 대해서 인식 정확도를 개선하였다.

3. 딥러닝을 통한 문서 내 표 항목 분류 및 인식 방법

인공지능을 통해 문서 내 표 항목의 내용을 인식하기 위한 흐름으로는 CNN를 통해 표 영역을 검출하는 단계, 검출된 영역에서 구분자를 검출하여 각각의 항목으로 분할하는 단계, 분할된 항목에서 문자를 인식하는 단계, 인식된 결과를 규격화된 파일 포맷으로 저장하는 단계로 이루어진다. Fig. 1(a)는 제안하는 방법의 흐름도를 보인다.

3.1 딥러닝을 통한 문서 내 표 항목 분류

문서 이미지에서 표 영역을 검출하기 위해 Faster R-CNN을 적용한다. Faster R-CNN은 객체의 크기 변화에 영향을 받지 않는 고유의 특징을 검출할 수 있는 FPN과 객체가 있을 확률이 높은 영역을 제안하는 RPN으로 구성된다. Faster R-CNN의 Backbone

신경망은 ResNet-101[18]을 적용하였다. Fig. 1(b)는 표 검출을 위한 신경망 구조를 보인다.

신경망의 학습을 위해 Fig. 2와 같은 표가 포함된 500개의 문서 이미지를 사용하였다. 학습 데이터에는 표가 포함된 문서 이미지와, 경계박스의 좌표들을 학습했다. 학습에서의 Epoch는 총 100회를 적용하였다. Fig. 3는 이 방법을 통해 학습된 신경망을 통해 검출된 표의 영역을 보인다.

표 내부의 각 항목에 구분자가 있을 경우 해당 구분자로 인해 문자 인식 정확도가 저하될 수 있다. 표 내의 구분자를 제거하고, 표 내의 각 항목을 구분하기 위해 Canny 알고리즘을 통해 영역 내 외곽선을 검출한다. 그 후 Douglas-Peucker 알고리즘을 적용하여 외곽선을 이루는 꼭지점을 검출한다. 연결된 하나의 외곽선에서 꼭짓점이 4개일 경우, 해당 외곽선을 표를 이루는 구분자로 검출한다. 표 전체의 외곽을 둘러싸는 사각형이 검출되는 경우를 고려하여 검출

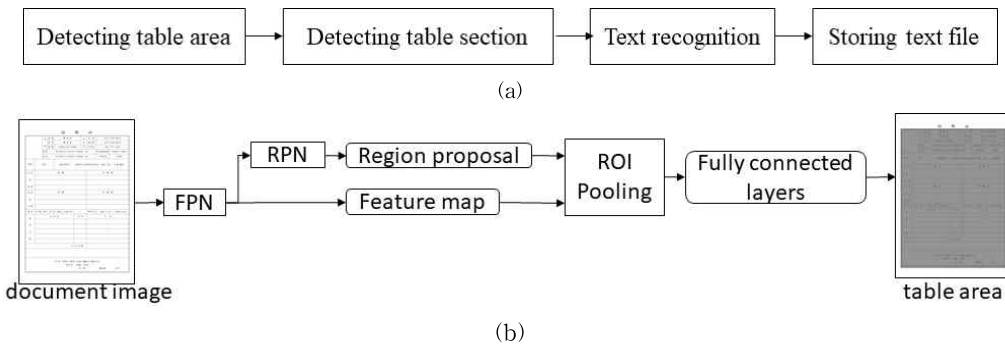


Fig. 1. (a) Flow of proposed method and (b) structure of Faster R-CNN for detecting table area.

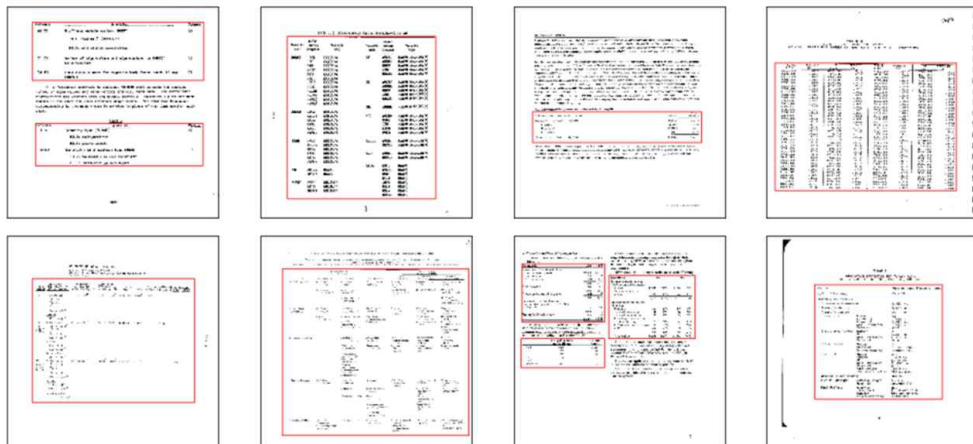


Fig. 2. Example of train images for table detection.

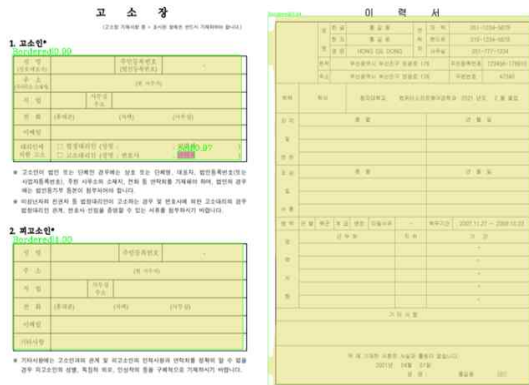


Fig. 3. Table detection results.

된 사각형들 간의 포함관계를 조사하여 2개 이상의 사각형을 포함하는 사각형은 제외한다. Fig. 4는 검출된 표 영역에서 항목 영역을 검출하는 것을 보인다.

3.2 딥러닝을 통한 표 항목 내 문자 인식

본 논문에서는 표 내의 항목을 인식하기 위해 CNN과 RNN을 결합한 신경망을 적용한다. 먼저 CNN층에서는 표 내 각 항목 영역의 이미지를 입력받아 해당 이미지에 대한 특징 맵을 검출한다. RNN층은

CNN층에서 추출된 이미지의 특성을 Bi-LSTM구조를 활용하여 시계열 데이터의 특징을 추출한다. 그 후 CTC 층을 통해 정렬된 형태로 인식된 텍스트를 출력한다. 이 때 다양한 유형의 글꼴에 대한 글자가 있는 문장들이 있는 데이터 셋 이미지를 생성하여 학습에 활용한다. 이 때 단어들을 무작위로 배치하여 한 줄로 구성된 단어들이 있는 이미지를 Fig. 5와 같이 생성하고, 해당 이미지에 라벨을 붙인다. 한글의 경우에는 국립국어연구원에서 제공하는 단어 목록 [19]을 활용하여 문장 데이터 셋 10000장을 생성한다. 그 후 전체 문자에 대해 Epoch를 1000회 수행하여 신경망을 학습시킨다.

그 후 문자 인식 결과를 csv나 엑셀 파일 등 규격화된 파일 포맷으로 저장한다. 이를 통해 표 내부의 항목 별 내용을 디지털화하여 자동으로 문서 처리를 수행하도록 한다. Fig. 6은 해당 과정을 통해 규격화된 문서로 저장된 결과를 보인다.

문자 인식 결과에 대해 제일 좋은 인식 결과뿐만 아니라 차 순위 인식 결과들도 나올 수 있다. 또한 인식정확도가 현저히 낮아 사용자 입장에서 신뢰를 할 수 없는 경우도 있다. 이러한 경우에는 각각의 인식 정확도를 판단하여 상위 3개의 인식 결과와 해당

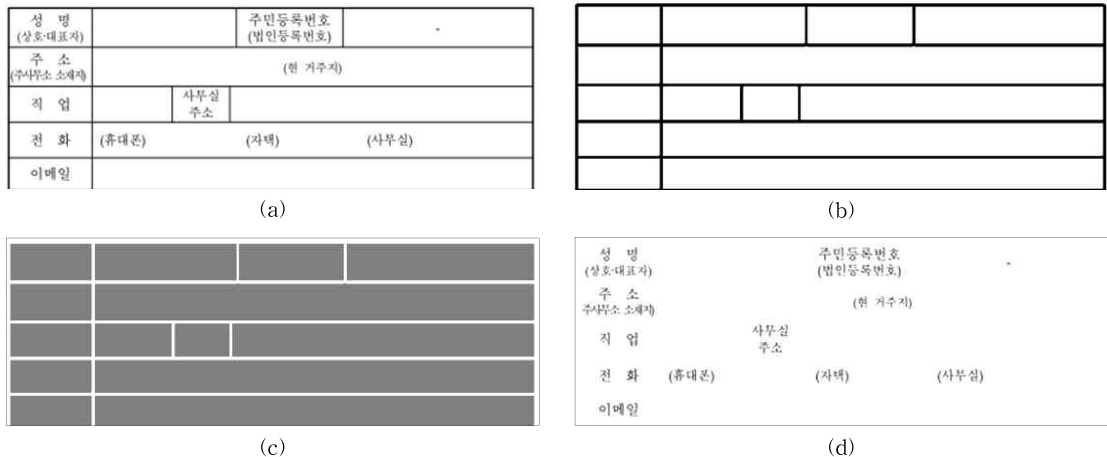


Fig. 4. Table section detection. (a) original document image, (b) detecting table contour, (c) extracting table section areas, and (d) document image with table outline removed.

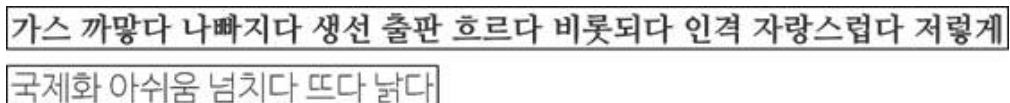


Fig. 5. Generated trained images for text recognition.

성 명	홍길동	주민등록번호	970126 - 1849571
주 소	부산광역시 부산진구 가야3동 엄광로 176		
직 업	학생	사무실 주소	부산광역시 부산진구 가야3동 엄광로 176
전 화	010 - 1234 - 5678		
이메일	software@deu.ac.kr		

성명, 홍길동
 주민등록번호, 970126-1849571
 주소, 부산광역시 부산진구 가야3동 엄광로 176
 직업, 학생
 사무실 주소, 부산광역시 부산진구 가야3동 엄광로 176
 전화, 010-1234-5678
 이메일, software@deu.ac.kr

Fig. 6. Result of text recognition for table.

문자가 있는 영역을 같이 제시하여 사용자가 직접 판단하게 함으로써 오인식으로 인한 서류 자동화의 오류를 사전에 차단한다.

4. 실험결과

본 논문에서 제안하는 방법의 정확도를 측정하기 위해 한글 문서 10장, 영어 문서 5장을 스캔하여 실험을 수행하였다. Fig. 7은 실험에 쓰인 문서 이미지의 일부이다.

본 논문에서 제안한 표 검출 방법의 정확도를 측정 한 결과는 Table 1과 같다. 이 때 총 19개의 표 중에 18개의 표를 검출하여 94.7%의 검출 정확도를 보여 CNN 신경망을 통해서 표가 거의 정확하게 검출이 된 것을 보인다.

검출된 표 영역에서 문자 인식의 정확도를 측정하였다. 이 때 문자 인식 정확도는 식 (1)과 같이 전체 문자와 정확하게 인식된 문자의 개수의 비율로 측정

하였다.

$$\text{accuracy} = \frac{\text{number of correctly recognized characters}}{\text{number of total characters}} \times 100 \quad (1)$$

제안된 방법을 통해 표 영역에서 문자 인식의 정확도는 Table 2와 같다. 이 때 표 항목 분류를 수행하지 않고 문자 인식을 했을 때는 정확도가 통상 문서를 인식할 때보다 떨어졌는데, 표의 항목을 구분하는 세로 줄과 같은 구분자가 별도의 문자로 인식되어 전체적인 문장의 인식률이 저하되었다. 반면 표 항목 분류를 수행한 후 문자 인식을 수행한 결과는 문자 인식 정확도가 약 7% 증가하였다. 이는 표 항목 분류를 통해 의미가 다른 표 항목이 같은 문장으로 인식되지 않고, 별개로 인식이 수행되어 시계열을 처리하는 신경망인 RNN의 성능이 좋아졌기 때문이다.

Table 3은 전체 문서에 대한 문자 인식 정확도를 보인다. 한글 문서의 경우 93.2%, 영어 문서의 경우 98.1%의 정확도를 보인다.

Table 1. Accuracy of table detection.

Total tables in documents	Detected tables	Accuracy of table detection
19	18	94.7%

5. 결 론

본 논문에서는 문서 처리를 자동화하기 위해 표가 포함된 문서에 대해 딥러닝 신경망을 통해 항목 별 내용을 분류하고 항목 내 문자를 인식하는 방법을



Fig. 7. Document images for simulation.

Table 2. Accuracy of character recognition in table.

Document language	Character recognition accuracy		
	Without table item classification	After table item classification	Increase rate of character recognition accuracy
Korean	86.2%	93.1%	6.9%
English	90.3%	97.4%	7.1%

Table 3. Accuracy of character recognition in document.

Document language	Character recognition accuracy
Korean	93.2%
English	98.1%

제안하였다. 먼저 문서 이미지에서 CNN을 통해 표 영역을 검출하고, 외곽선 검출을 통해 항목에 대한 영역을 검출했다. 그 후 CNN과 RNN이 결합된 신경망을 통해 표 내부의 항목 별 문자를 인식하였다. 그 후 규격화된 텍스트 파일로 결과를 저장하였다. 이때 인식 결과로 여러 결과가 나온다면 인식을 상위 3개의 후보를 제시하여 사용자가 판단할 수 있도록 했다. 제안된 방법을 통해 OCR로 인식한 문서의 내용을 항목 별로 처리할 수 있게 되어 문서 처리의 자동화가 원활해질 수 있을 것이다. 또한 문서 처리 업무의 자동화를 달성할 수 있게 되어 업무 효율의 증가와 문서 처리 작업의 오류를 줄일 수 있을 것으로 기대한다.

REFERENCE

[1] Korean Import and Export Logistics Process Vol 5(2016), <https://www.nlic.go.kr/nlic/WhsBoardPdfSch.action> (accessed May 28, 2021).
 [2] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-time Object Detection with Region Proposal Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 39, No. 6, pp. 1137-1149, 2017.
 [3] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 42, No. 2, pp. 386-397, 2020.

[4] B. Shi, M. Yang, X. Wang, P. Lyu, C. Yao, and X. Bai, "ASTER: An Attentional Scene Text Recognizer with Flexible Rectification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 41, No. 9, pp. 2035-2048, 2019.
 [5] T. He, Z. Tian, W. Huang, C. Shen, Y. Qiao, and C. Sun, "An End-to-End TextSpotter with Explicit Alignment and Attention," *Proceeding of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5020-5029, 2018.
 [6] S. M. Gang and J. J. Lee, "Coreset Construction for Character Recognition of PCB Components Based on Deep Learning," *Journal of Korea Multimedia Society*, Vol. 24, No. 3, pp. 382-395, 2021.
 [7] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based Learning Applied to Document Recognition," *Proceedings of the IEEE*, Vol. 86, No. 11, pp. 2278-2324, 1998.
 [8] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Communications of the ACM*, Vol. 60, No. 6, pp. 84-90, 2017.
 [9] P. Soviany and R. T. Ionescu, "Optimizing the Trade-off between Single-Stage and Two-Stage Deep Object Detectors using Image Difficulty Prediction," *Proceeding of the International Symposium on Symbolic and Numeric Algorithms for Scientific Computing*, pp. 209-214, 2018.
 [10] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmenta-

tion,” *Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 580-587, 2014.

[11] R. Girshick, “Fast R-CNN,” *Proceeding of the IEEE International Conference on Computer Vision*, pp. 1440-1448, 2015.

[12] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You Only Look Once: Unified, Real-Time Object Detection,” *Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 779-788, 2016.

[13] J. Redmon and A. Farhadi, “YOLO9000: Better, Faster, Stronger,” *Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6517-6525, 2017.

[14] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, and A. C. Berg, “SSD: Single Shot MultiBox Detector,” *Proceeding of the European Conference on Computer Vision*, pp. 21-37, 2016.

[15] 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, 2017.

[16] J. Wang and X. Hu, “Gated Recurrent Convolution Neural Network for OCR,” *Proceeding of the International Conference on Neural Information Processing Systems*, pp. 334-343, 2017.

[17] Z. Cheng, P. Bai, Y. Xu, G. Zheng, S. Pu, and S. Zhou, “Focusing Attention: Towards Accurate Text Recognition in Natural Images,” *Proceeding of the IEEE International Conference on Computer Vision*, pp. 5076-5084, 2017.

[18] K. He, X. Zhang, S. Ren and J. Sun, “Deep Residual Learning for Image Recognition,” *Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770-778, 2016.

[19] Vocabulary List for Learning Korean(2003), https://www.korean.go.kr/front/etcData/etcDataView.do?mn_id=46&etc_seq=71 (accessed May 28, 2021).



이 동 석

동의대학교 컴퓨터소프트웨어공학과 공학사
 동의대학교 컴퓨터소프트웨어공학과 공학석사
 동의대학교 컴퓨터소프트웨어공학과 공학박사

동의대학교 인공지능그랜드ICT연구센터 연구교수
 관심분야 : 멀티미디어 신호처리, 영상 인식



권 순 각

경북대학교 전자공학과 공학사
 KAIST 전기및전자공학과 공학석사
 KAIST 전기및전자공학과 공학박사
 동의대학교 컴퓨터소프트웨어공학과 교수

관심분야 : 멀티미디어처리, IoT