

## 표의 테두리 유사 라벨을 활용한 문자 영역 검출 방법

한정훈<sup>1</sup> · 박세진<sup>1</sup> · 문영식<sup>2\*</sup>

### Text Region Detection Method Using Table Border Pseudo Label

Jeong Hoon Han<sup>1</sup> · Se Jin Park<sup>1</sup> · Young Shik Moon<sup>2\*</sup>

<sup>1</sup>Ph.D. Student, Hanyang University - Ansan Campus, Department of Computer Science and Engineering, 15588 Korea

<sup>2\*</sup>Professor, Hanyang University - Ansan Campus, Department of Computer Science and Engineering, 15588 Korea

#### 요 약

문자 영역 검출이란 수기 혹은 인쇄된 문서에서 문자의 영역을 검출하는 기술이다. 검출된 문자 영역들은 인식 단계를 거쳐 디지털화되며 이는 활용 목적에 따라 다양한 곳에서 활용된다. 하지만 문자 단위의 검출 결과는 대용량 문서를 인식해야 하는 산업 현장의 문자 인식 단계에는 적합하지 않다. 또한, 문서 내 존재하는 표는 문자 영역 검출 단계에서 오검출을 야기하며 이는 문자 인식 단계에서 악영향을 끼친다. 이를 해결하기 위해 본 논문에서는 표의 테두리 정보를 활용한 문자 영역 검출 방법을 제안한다. 표의 테두리 정보를 활용하기 위하여 제안하는 방법은 2개 디코더를 추가하고 간접적인 학습을 유도하기 위하여 각 디코더의 흐름을 조절하였다. 실험을 통해 표의 테두리 유사 라벨을 이용한 약지도 학습 방법이 성능 향상에 도움이 됨을 보였다.

#### ABSTRACT

Text region detection is a technology that detects text area in handwriting or printed documents. The detected text areas are digitized through a recognition step, which is used in various fields depending on the purpose of use. However, the detection result of the small text unit is not suitable for the industrial field. In addition, the border of tables in the document that it causes miss-detected results, which has an adverse effect on the recognition step. To solve the issues, we propose a method for detecting text region using the border information of the table. In order to utilize the border information of the table, the proposed method adjusts the flow of two decoders. Experimentally, we show improved performance using the table border pseudo label based on weak supervised learning.

**키워드** : 문자 영역 검출, 약 지도 학습, 유사 라벨, 딥러닝

**Keywords** : Text region detection, Weakly supervised learning, Pseudo label, Deep learning

Received 18 August 2020, Revised 20 August 2020, Accepted 2 September 2020

\* Corresponding Author Young Shik Moon (E-mail : ysmoon@hanyang.ac.kr, Tel: +82-31-400-4049)

Professor, Hanyang University - Ansan Campus, Department of Computer Science and Engineering, 15588 Korea

Open Access <http://doi.org/10.6109/jkiice.2020.24.10.1271>

print ISSN: 2234-4772 online ISSN: 2288-4165

©This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.  
Copyright © The Korea Institute of Information and Communication Engineering.

## I. 서 론

광학 문자 인식(Optical Character Recognition, OCR)은 문자 검출(Text Detection), 문자 인식(Text Recognition) 기술로 나눌 수 있으며 인쇄 혹은 사람이 필기한 문자를 디지털(Digital)화된 문자로 변환하는 기술이다. 많은 문서들이 OCR 기술을 통해 디지털화 되고 있으며 목적에 맞게 재 가공되어 많은 분야에 사용된다[1, 2].

OCR 기법을 이용한 대용량 문서 처리 방법에는 많은 양의 문서를 처리하기 위해 문단(Paragraph) 및 표(Table) 검출 방법이 사용된다. 이는 일반적인 문서에 나타날 수 있는 2가지의 큰 구조를 검출하는 것이다. 검출된 각 문단과 표는 문자 인식 알고리즘을 거쳐 디지털화 된다. 문자 인식 단계에서 앞서 검출된 표의 테두리(Border)는 문자 인식에 악영향을 줄 수 있으므로 검출된 문단과 표는 서로 다른 후처리 방법을 필요로 한다.

영상 처리에 특화된 딥러닝(Deep Learning) 기법인 CNN(Convolutional Neural Network)은 많은 컴퓨터 비전 문제 해결에 널리 사용되고 있다. 이 방법은 전통적인 컴퓨터 비전 알고리즘에서 사용되는 특징 추출 단계를 자동으로 수행하며 복잡한 환경에서 제한적인 성능을 보이는 전통적인 방법에 비해 높은 성능을 보이고 있다[3, 4, 5, 6].

최근에는 이러한 CNN의 장점을 이용한 문자 검출 및 표 검출 방법들이 제안되고 있다 [7, 8]. 하지만 산업 현장의 대용량 문서 디지털화 문제에서 문자 인식은 검출된 결과단위로 문자 인식을 수행하기 때문에 문자 단위의 검출 방법은 적합하지 않다. 또한, 문자 단위로 검출된 결과를 묶어 문단 형태로 바꾸는 후처리 작업에는 검출 결과에 따라 또 다른 소요시간을 발생시킨다. 이를 해결하기 위하여 CNN의 학습 데이터를 문단 형태로 바꾸어 학습할 경우 표 내부에 존재하는 셀(Cell)의 문자

영역을 오검출하거나 표의 테두리를 문자로 검출하는 문제점이 발생한다. 이는 문자 인식 성능에 악영향을 준다. 그림 1은 테두리를 포함하는 영상에서 테서렉트(Tesseract)[9]를 이용한 문자 인식의 예이다.

이와 같은 단점을 극복하기 위해 제안하는 방법은 표의 테두리 정보를 활용한 문자 영역 검출 방법을 제안한다. 그림 2는 본 논문의 목표인 문단 검출 및 셀 내부의 문자 검출의 예시이다.

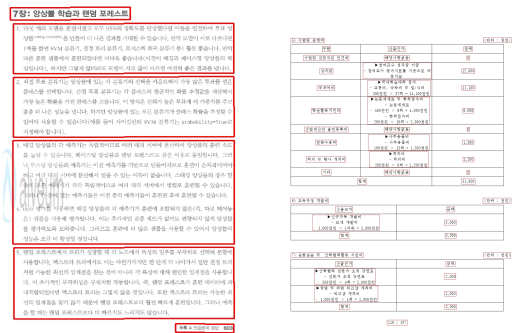


Fig. 2 The example of the text region detection

## II. 관련 연구

### 2.1. U-Net

Ronneberger는 영상 분할을 위한 U-Net[10] 네트워크 구조를 제안하였다. U-Net은 FCN(Fully Convolutional Network) 네트워크를 기반으로 구성된다. 기존 영상 분할 네트워크 대비 간단한 구조로 이루어져 있으며 상대적으로 높은 성능을 보여주었다. 이후 U-Net의 구조는 영상 분할 네트워크에 많은 영향을 주었다. 이러한 구조를 바탕으로 Paliwal은 영상 내에서 표와 표 내부의 셀을 분할하는 TableNet[8]을 제안하였다.

### 2.2. DeepLab V3+

Chen[11]는 기존의 영상 분할 네트워크의 문제점인 작은 수용영역(Receptive Field)을 해결하고 높은 파라미터 수를 줄이기 위해 기존 인코더 디코더(Encoder-Decoder) 형태의 네트워크에 다양한 모듈을 적용한 DeepLab V3+ 네트워크를 제안하였다. Chen은 Xception 모듈 및 ASSPP(Atrous Separable Spatial Pyramid Pooling), U-Net 형태의 디코더를 결합한 구조를 제안하

광역교수 지도학생 수											
석사			박사			석·박사 통합			계		
전체	장여	장여비율(%)	전체	장여	장여비율(%)	전체	장여	장여비율(%)	전체	장여	장여비율(%)
120	100	83.3	150	120	80.0	270	220	81.5	370	320	86.5

(a) Input image

광역교수 지도학생 수											
석사			박사			석·박사 통합			계		
전체	장여	장여비율(%)	전체	장여	장여비율(%)	전체	장여	장여비율(%)	전체	장여	장여비율(%)
120	100	83.3	150	120	80.0	270	220	81.5	370	320	86.5

(b) The text recognition result of input image (a)

Fig. 1 The example of the text recognition

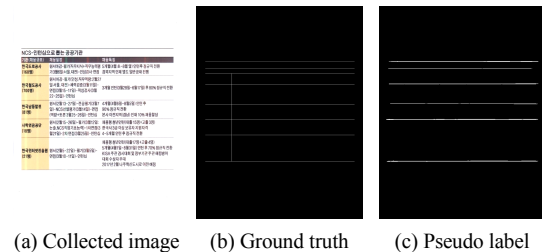
였다. 이 방법은 기존의 방법보다 높은 성능을 보이면서 상대적으로 낮은 연산량을 가지는 이점이 있다.

### 2.3. PSENet

Wang[7]은 기존의 오브젝트 검출 방법의 메커니즘을 활용한 문자 검출 방법에서 나타날 수 있는 문제점(부정확한 좌표 예측, 반복적인 검출)을 해결하기 위해 영상 분할 기반의 문자 검출 방법을 제안하였다. Wang은 문자 검출을 위한 영상 분할 결과에 스케일 확장 알고리즘을 활용하여 문자 검출 성능을 높였다.

### 2.4. 멀티 태스크 러닝(Multi-Task Learning)

멀티 태스크 러닝이란 단일 네트워크에 다수의 목적을 가진 출력층을 사용하여 여러 태스크를 동시에 학습하는 방법이다. 이러한 방법은 사람이 새로운 것을 학습할 때, 관련 지식을 활용 및 학습하는 것이 본래 목적의 학습에 도움을 준다는 것에 영감을 얻어 개발되었다 [12]. 이를 통해 멀티 태스크 러닝으로 학습되는 딥러닝 네트워크는 싱글 태스크로 학습하는 것 보다 풍부한 특징을 학습할 수 있는 장점이 있다.



(a) Collected image (b) Ground truth (c) Pseudo label

Fig. 4 The example of the table border pseudo label

### 2.5. 유사 라벨(Pseudo Label)

유사 라벨이란 사람이 직접 주석(Annotation)을 생성한 정답 데이터를 기반으로 학습된 네트워크로 주석처리가 되지 않은 다수의 데이터에 주석을 생성하는 것을 말한다[13]. 이는 사람이 직접 생성한 주석보다 낮은 정확도를 보이는 단점이 있지만 많은 양의 데이터에 비교적 적은 자원을 이용하여 주석을 생성한다는 장점이 있다.

## III. 데이터 수집

### 3.1. 데이터 수집

네트워크의 학습을 위하여 본 논문에서는 인터넷에서 수집된 총 4117장의 영상을 사용하였다. 수집된 영상은 Xerox D110, HP Color Laserjet Enterprise M553 모델로 출력 후 각각 Xerox C5580, HP Pro-M476dw 모델로 스캐닝 하였다. 수집된 데이터 중 2757장의 문자 영역과 표의 테두리 정답(Ground Truth)은 사람에게 의하여 주석처리 되었으며 2757장의 영상 중 표를 포함하고 있는 영상은 249장이다. 수집된 나머지 데이터 1360장의 표를 포함하고 있는 영상은 유사 라벨 생성에 사용되었다. 그림 3은 수집된 영상의 정답 예시이다. 그림 3에서 각 네모 박스가 한 개의 텍스트 영역 정답이다.

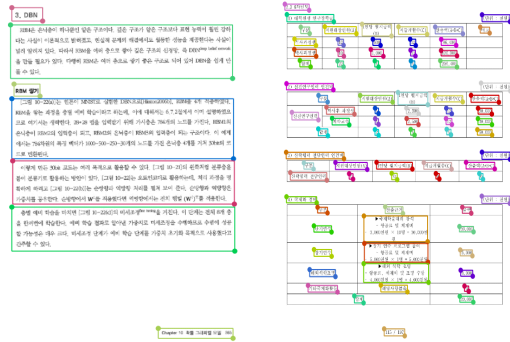


Fig. 3 The example of the generated ground truth

### 3.2. 유사 라벨 생성 방법

수집된 데이터 중 표의 테두리를 사람이 직접 주석 처리한 데이터는 총 249장의 영상이다. 본 논문에서는 사람이 직접 주석 처리한 249장의 테두리 정답을 모두 사용하여 U-Net을 학습한 후 학습된 네트워크를 이용하여 1360장의 표 테두리에 대한 유사 라벨을 생성하였다. 그림 4는 생성된 유사 라벨의 예시이다.

그림 4에서 흰색 선은 표의 테두리이다. 그림 4에서 볼 수 있듯이 사람이 직접 주석 처리한 정답 영상에 비해서 생성된 유사 라벨의 경우 테두리의 굵기가 일정하지 않으며 몇몇 테두리는 검출되지 않은 것을 확인할 수 있다. 표 1은 본 논문에서 사용한 데이터 셋의 정보이다.

**Table. 1** Dataset specifications

Dataset	Size	Annotation		
		Text Region	Table border	Pseudo label
Human Annotation	2757	2757	249	0
Pseudo Annotation	1360	0	0	1360

#### IV. 제안하는 방법

##### 4.1. 네트워크 구조

제안하는 방법의 네트워크 구조는 영상 분할 기법에서 널리 사용되는 U-Net 구조를 변경하여 디자인하였다. 제안하는 방법의 두 개의 디코더는 각각 문자 영역 검출과 표의 테두리 검출에 사용된다. 문자 검출 디코더는 인코더에서 추출된 특징 맵을 확장하기 위한 업 샘플링(Up-Sampling) 레이어와 ASSPP 모듈로 구성된다. ASSPP 모듈은 다양한 크기의 문자를 포함하는 문서에서 스케일에 적응적이며 풍부한 특징을 추출하기 위하여 사용된다.

테두리 검출 디코더는 2개의 컨볼루션 레이어와 가이드드 필터(Guided Filter) [14]로 구성되어 있으며 첫 번째 인코더의 특징 맵과 인코더의 마지막 특징 맵을 입력으로 사용한다. 이 구조는 다양한 스케일에서 추출된 특징을 사용하는 동시에 풀링(Pooling) 연산으로 사라

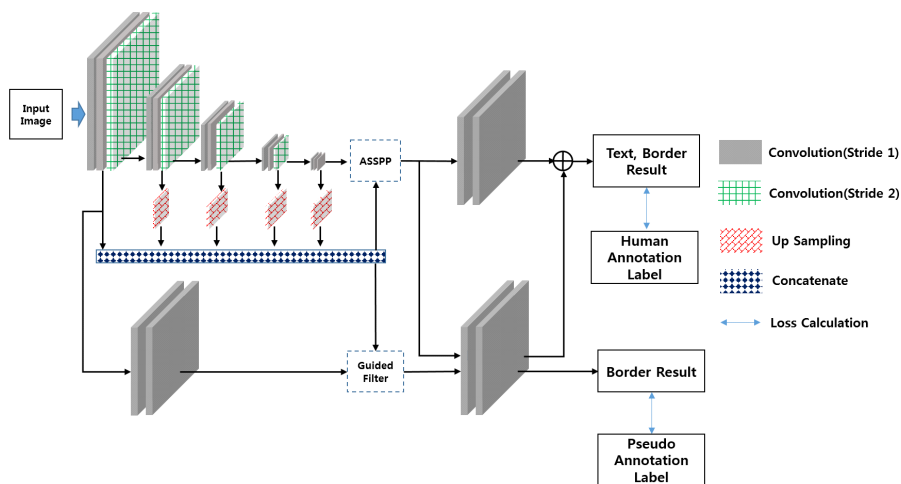
질 수 있는 윤곽선의 정보를 유지하기 위해 사용된다. 가이드드 필터는 2개의 컨볼루션 연산을 통해 보정된 특징 맵과 인코더의 마지막 블록의 결과를 이용하여 영상 내 표 테두리의 확률 맵을 조정한다.

제안하는 방법은 유사 라벨 데이터셋을 이용한 약지도 학습을 위하여 Pseudo Annotation Dataset은 테두리 검출 디코더만을 통과하게 하였다. Human Annotation Dataset의 경우 문자 검출의 흐름이 문자 검출 디코더와 테두리 검출 디코더를 모두 통과하게 하여 테두리 유사 라벨의 학습된 디코더에 영향을 받도록 설계하였다. 그림 5는 제안하는 방법의 네트워크 구조이다.

##### 4.2. 학습 방법

제안하는 방법의 문자 영역 검출 네트워크 및 표의 테두리 검출 네트워크의 학습을 위한 손실함수는 이진 교차 엔트로피(Binary Cross Entropy)를 사용하였다. 모델의 활성화 함수로는 ReLU를 사용하였으며 학습율과 최적화 함수로는 각각  $10^{-4}$ , Adam을 사용하였다.

학습 단계에서 데이터 증강(Data Augmentation)을 위하여 무작위 스케일링, 회전, 좌우 및 상하반전을 사용하였다. 문자 영역 검출 디코더의 학습을 위한 데이터는 0.5, 1, 2, 3배 크기의 무작위 스케일링을 적용하였으며 테두리 디코더 학습을 위한 데이터는 얇은 테두리의 사라짐을 방지하기 위하여 1, 2, 3배의 스케일링을 적용하였다. 학습을 위하여 스케일링 된 데이터는 무작위로  $512 \times 512$  크기로 잘라 학습에 사용되었다.



**Fig. 5** Overview of proposed method

## V. 실험 결과 및 분석

본 논문에서는 Intel i5-8086K 4.00 GHz CPU, 16GB RAM, GeForce RTX 1080Ti GPU 1개 사양 PC에서 Python, OpenCV 라이브러리와 Pytorch 딥러닝 프레임워크를 이용하여 제안 방법을 구현하였다.

### 5.1. 테두리 정보 추가 실험

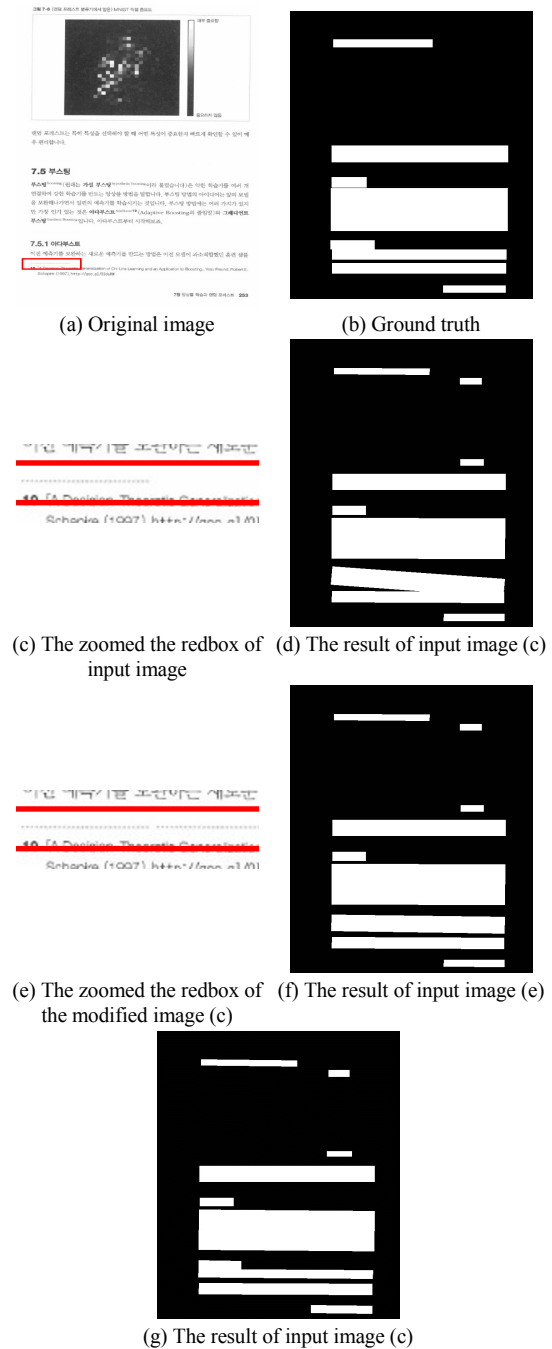
본 실험에서는 테두리 정보를 추가할 경우 네트워크의 성능 변화를 확인하기 위하여 영상 분할 기반 문자 검출 방법인 PSENet[7]을 기반으로 실험을 진행하였다. 네트워크의 학습에는 총 2000장의 Human Annotation Dataset을 사용하였다. 학습에 사용된 영상 중 표를 포함하는 영상은 100장이다. 이외 Human Annotation 데이터는 각각 검증데이터 257장, 테스트 500장으로 사용하였다. 검증과 테스트 데이터 중 표를 포함하는 영상은 각각 49장과 100장을 사용하였다. 정량적 평가를 위하여 제안하는 방법에서는 재현율(Recall), 정확도(Precision), F-척도(F-measure)를 사용하였다. 정량적 평가는 각 영상의 검출된 픽셀과 정답 영상과의 비교를 통해 측정하였다. 표 2는 각 네트워크의 정량적 평가 결과이다.

**Table. 2** Quantitative evaluation of the test dataset by different training dataset

Backbone Network (Training Data)	Metric		
	Recall	Precision	F-measure
PSENet(text data)	0.83	0.86	0.84
PSENet (text, table border data)	0.80	0.91	0.86

표 2에서 볼 수 있듯이 100장의 테두리 정보를 추가하였을 때, 재현율은 약 0.03 감소하였지만 정확도는 0.05 증가한 것을 확인할 수 있다. 이는 멀티 태스크 러닝과 동일하게 표의 테두리 정보를 통해 문자 영역과 테두리 영역의 특징 정보를 풍부하게 학습하여 문자 영역을 보다 정확하게 검출한 것으로 판단할 수 있다.

그림 6은 각 네트워크의 문자 영역 예측 결과이다. 그림 6의 (a), (b)는 원본 영상과 정답 영상이다. (c)는 원본 영상의 각주 표시(빨간 박스)를 확대한 영상이며 (e)는 (c)의 영상 내 각주 표시(빨간 박스 내)를 반복적으로 이어붙이고 확대한 영상이다. (d), (f)는 문자 영역 데이터만을 학습한 PSENet을 이용하여 (c), (e) 영상을 추론한



**Fig. 6** Comparison of results obtained by PSENet trained with different training dataset  
 (d), (f) : The inference result of trained PSENet with text dataset  
 (g) : The inference result of trained PSENet with text, table border dataset

결과이다. (g)는 표의 테두리 정보를 동시에 학습한 네트워크의 결과이다. (c)의 입력영상에 대한 결과 (d)에서 볼 수 있듯이 표의 테두리 정보를 추가하지 않고 학습한 경우 (d)의 하단 부분의 각주 부분은 검출되지 않았으나 이외 문자 영역이 서로 겹쳐 검출되었다. 이는 문자 영역의 범위를 정확히 검출하지 못한 것으로 판단할 수 있다. 결과 (f)에서 볼 수 있듯이 각주 표시가 문자 사이에 존재할 경우 문자 영역의 범위를 정확히 검출하였다. 반면 테두리 정보를 이용하여 학습된 네트워크의 결과인 (g)의 경우 각주 표시가 왼쪽에만 존재함에도 2개의 문자 영역이 서로 분리되어 검출된 것을 확인할 수 있다. 이는 학습단계에서 표의 테두리 정보가 문자 영역 검출에 영향을 준 것으로 판단할 수 있다.

5.2. 네트워크 변경 실험

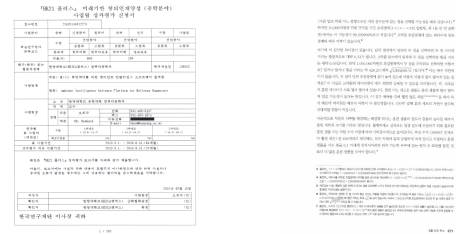
본 실험에서는 제안하는 방법의 문자 영역 검출 성능을 비교를 위하여 PSENet[7], TableNet[8]과 제안하는 방법의 문자 영역 검출 성능을 정량적으로 평가하였다. 네트워크 학습을 위한 데이터 및 검증, 테스트 데이터는 5.1절과 동일하게 진행하였다. 표 3은 각 네트워크 성능의 정량적 평가 결과이다.

Table. 3 Quantitative evaluation of the test dataset by different backbone network based on supervised learning

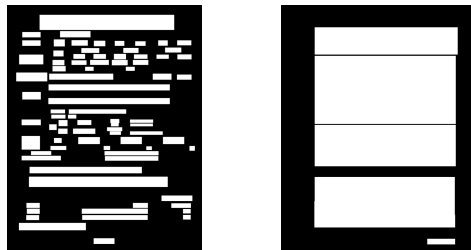
Backbone Network	Metric		
	Recall	Precision	F-measure
TableNet	0.83	0.88	0.86
PSENet	0.80	0.91	0.86
Proposed Method	0.82	0.94	0.88

표 3에서 볼 수 있듯이 제안하는 방법의 네트워크와 기존 문자 검출 및 표 검출 네트워크를 비교하였을 때, 재현율 측면에서 TableNet보다 0.01 낮은 성능을 보였고, PSENet과 비교하였을 때 0.02 성능 향상을 보였다. 반면 정확도 측면에서는 제안하는 방법이 0.94로 기존의 방법과 비교하였을 때 최대 0.06의 성능 향상을 보였다. 이는 제안하는 방법의 네트워크가 기존의 방법과 유사한 재현율 성능을 보이지만 정확도 측면에서는 성능 향상이 있다고 판단할 수 있다.

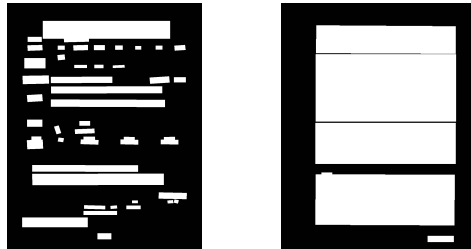
그림 7은 각 네트워크의 문자 영역 검출 결과이다. 각 네트워크의 결과 (c), (d), (e)에서 볼 수 있듯이 테두리가



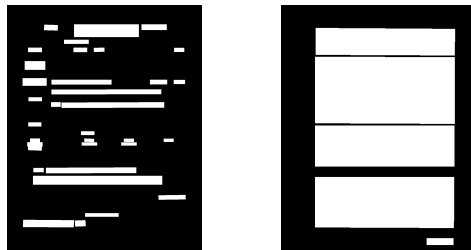
(a) Input image



(b) Ground truth



(c) The result of TableNet



(d) The result of PSENet



(e) The result of proposed method

Fig. 7 Comparison of results obtained by different backbone network trained with supervised learning

포함된 영상을 입력으로 하였을 때, 셀 내부의 문자 영역을 대부분 미검출한 것을 확인 할 수 있다. 또한, 각 셀 내부의 문자 영역이 끊어져 검출되는 것을 확인할 수 있다. 반면 테두리가 없는 영상을 입력으로 하였을 경우 3개의 네트워크 결과 모두 문단 영역을 비교적 정확히 검출한 것을 확인할 수 있다. 그림 7의 실험 결과와 표 3을 비교하였을 때, 재현율과 정확도에 차이를 보이고 있다. 이는 테스트 셋의 표 데이터 수와 문단 데이터의 수와 각 데이터가 포함하고 있는 정답 픽셀의 수가 차이가 나기 때문이다.

### 5.3. 유사 라벨 테두리 정보 추가 실험

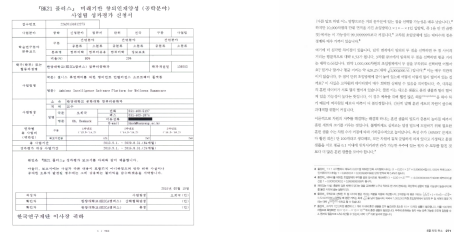
앞선 5.1, 5.2절의 실험에서는 표의 테두리 정보 추가 및 네트워크의 구조 변경에 대한 실험을 진행하였다. 본 실험에서는 수집된 2개의 데이터 셋을 모두 활용한 약지도 학습 실험을 진행하였다. Human Annotation 데이터는 앞선 5.2 실험과 동일한 데이터를 사용하였고 Pseudo Annotation은 1360장의 데이터를 모두 사용하였다. 표 4는 두 개 데이터셋을 활용한 문자 영역 검출 결과의 정량적 평가이다.

**Table. 4** Quantitative evaluation of the test dataset by different backbone network based on weakly supervised learning

Backbone Network	Metric		
	Recall	Precision	F-measure
TableNet	0.85	0.91	0.88
PSENet	0.89	0.89	0.89
Proposed Method	0.88	0.92	0.90

앞선 5.2 실험의 결과인 표 4와 비교하였을 때 재현율은 약 0.06의 평균적인 향상을 보였으나 정확도 측면에서는 0.003의 성능 하락을 보였다. 하지만 재현율 향상 대비 정확도 하락은 낮은 수치이며 2개의 수치를 기준으로 계산된 F척도 기준 평균 0.02의 성능 향상을 보였다. 이는 약지도 학습 기반의 테두리 유사 라벨이 문자 검출 성능 향상에 도움을 준 것으로 판단할 수 있다.

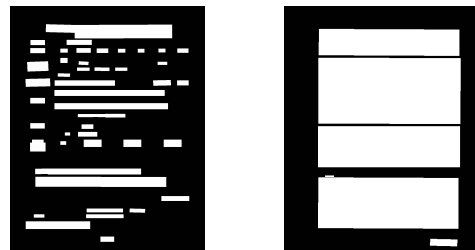
그림 8은 유사 라벨을 이용한 약지도 학습 결과이다. 그림 7과 비교하였을 때, 많은 테두리를 포함하는 입력 영상을 사용한 경우 (c), (d), (e)에서 볼 수 있듯이 표의 셀 내부 문자 검출 결과가 큰 차이를 보인다. 이는 유사 라벨을 이용한 테두리 정보의 약지도 학습이 네트워크



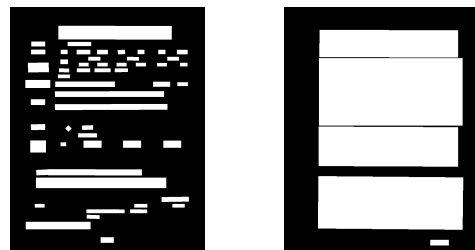
(a) Input image



(b) Ground truth



(c) The result of TableNet



(d) The result of PSENet



(e) The result of proposed method

**Fig. 8** Comparison of results obtained by different backbone network trained weakly supervised learning

성능에 크게 영향을 준 것으로 판단할 수 있다. 표가 없는 영상을 입력으로 사용하였을 경우 각 문단이 영역별로 정확히 나누어져 검출된 것을 볼 수 있다. 특히 그림 7과 비교하였을 때, 각 문단 사이가 벌어져 검출되어 문단 사이의 공간이 떨어져 검출된 것을 확인할 수 있다. 따라서 테두리 유사 라벨을 이용한 약지도 학습 방법이 문단 검출에 대한 성능을 낮추지 않으면서 셀 내부의 문자 검출의 성능을 향상시킨다고 판단할 수 있다.

## VI. 결론

본 논문에서는 표의 테두리 정보를 추가한 약지도 학습을 활용하여 영상 내 문자 영역 검출 방법을 제안하였다. 문단 위주의 학습 데이터 사용으로 인해 셀 내부의 문자 영역이 미검출 및 오검출되는 것을 방지하기 위하여 표의 테두리 정보를 활용하였으며 부족한 학습 데이터의 양을 늘림과 동시에 효율적인 학습을 위하여 유사 라벨을 활용하였다. 제안하는 방법의 성능 평가를 위하여 문자 검출 방법인 PSENet과 표 검출 방법인 TableNet을 비교하였으며 최대 재현율 0.03, 정확도 0.01, F-척도 0.02의 성능 향상을 실험을 통해 확인하였다. 차후 연구에서는 데이터 셋의 증량 및 본 논문의 학습 방법을 활용하여 표를 포함하는 서식화된 문서에서 개인 정보를 추출하는 연구를 진행할 예정이다.

### ACKNOWLEDGEMENT

This research was supported by the MISP (Ministry of Science, ICT), Korea, under the National Program for Excellence in SW(2018-0-00192) supervised by the IITP(Institute of Information & communications Technology Planning & Evaluation)" (2018-0-00192)

## REFERENCES

[ 1 ] Y. Xiao, M. Xue, T. Lu, Y. Wu, and S. Palaiahnakote, "A Text-Context-Aware CNN Network for Multi-oriented and Multi-language Scene Text Detection," in *Proceedings of*

*International Conference on Document Analysis and Recognition (ICDAR)*, pp. 695-700, 2019.

[ 2 ] O. Tursun, R. Zeng, S. Denman, S. Sivapalan, S. Sridharan, and C. Fookes, "MTRNet: A Generic Scene Text Eraser," in *Proceedings of International Conference on Document Analysis and Recognition (ICDAR)*, pp. 39-44, 2019.

[ 3 ] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770-778, 2016.

[ 4 ] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated Residual Transformations for Deep Neural Network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1492-1500, 2017.

[ 5 ] G. Pavlakos, L. Zhu, X. Zhou, and K. Daniilidis, "Learning to Estimate 3D Human Pose and Shape from a Single Color Image," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 459-468, 2018.

[ 6 ] J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation Networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7132-7141, 2018.

[ 7 ] W. Wang, E. Xie, X. Li, W. Hou, T. Lu, G. Yu, and S. Shao, "Shape Robust Text Detection With Progressive Scale Expansion Network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 9336-9345, 2019.

[ 8 ] S. S. Paliwal, R. Rahul, M. Sharma, and L. Vig, "TableNet: Deep Learning Model for End-to-end Table Detection and Tabular Data Extraction from Scanned Document Images," in *Proceedings of International Conference on Document Analysis and Recognition (ICDAR)*, pp. 128-133, 2019.

[ 9 ] R. Smith, "An Overview of the Tesseract OCR Engine," in *Proceedings of International Conference on Document Analysis and Recognition (ICDAR)*, 2007.

[ 10 ] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *Medical Image Computing and Computer-Assisted Intervention*, vol. 9351, pp. 234-241, 2015.

[ 11 ] L. C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adan, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European Conference on Computer Vision*, pp. 801-818, 2018.

[ 12 ] R. Caruana, "Multi-task learning," *Machine Learning*, vol. 28, no. 1, pp. 41 - 75, 1997.



- [13] Z. Chen, R. Zhang, G. Zhang, Z. Ma, and T. Lei, "Digging Into Pseudo Label: A Low-Budget Approach for Semi-Supervised Semantic Segmentation," in *IEEE Access*, vol. 8, pp. 41830-41837, 2020.
- [14] H. Wu, S. Zheng, J. Zhang, and K. Huang, "Fast End-to-End Trainable Guided Filter," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1838-1847, 2018.



**한정훈(Jeong Hoon Han)**

2008년 2월 한림대학교 컴퓨터공학 학사  
2008년 3월~현재 한양대학교 컴퓨터공학과 석박사과정  
※관심분야: 컴퓨터비전, 패턴인식, 딥러닝



**박세진(Se Jin Park)**

2007년 8월 아주대학교 컴퓨터공학 석사  
2012년 9월~현재 한양대학교 컴퓨터공학과 박사과정  
2015년~2016년 LG 전자 책임연구원  
2016년~현재 Vuno Inc, 선임연구원  
※관심분야: 컴퓨터비전, 패턴인식, 딥러닝



**문영식(Young Shik Moon)**

1980년 2월 서울대학교 전자공학과 공학사  
1982년 2월 한국과학기술원 전기 및 전자공학과 공학석사  
1990년 Ph. D. in Department of Electrical and computer Engineering University of California at Irvine  
1982년~1985년 한국 전자통신연구소 연구원  
1989년~1990년 Inno Vision Medical 선임연구원  
1990년~1992년 생산기술연구소 선임연구원  
1992년~현재 한양대학교 컴퓨터공학과 교수  
※관심분야: 영상처리, 컴퓨터비전, 패턴인식