

# A Study on the Deep Learning-Based Textbook Questionnaires Detection Experiment

Kim Tae Jong<sup>†</sup> · Han Tae In<sup>\*\*</sup> · Ji Su Park<sup>\*\*\*</sup>

## ABSTRACT

Recently, research on edutech, which combines education and technology in the e-learning field called learning, education and training, has been actively conducted, but it is still insufficient to collect and utilize data tailored to individual learners based on learning activity data that can be automatically collected from digital devices. Therefore, this study attempts to detect questions in textbooks or problem papers using artificial intelligence computer vision technology that plays the same role as human eyes. The textbook or questionnaire item detection model proposed in this study can help collect, store, and analyze offline learning activity data in connection with intelligent education services without digital conversion of textbooks or questionnaires to help learners provide personalized learning services even in offline learning.

Keywords : Deep Learning, Textbook Questionnaires Detection Model, Detection Model, Computer Vision

## 딥러닝 기반 교재 문항 검출 실험 연구

김 태 종<sup>†</sup> · 한 태 인<sup>\*\*</sup> · 박 지 수<sup>\*\*\*</sup>

## 요 약

최근 학습, 교육 및 훈련으로 일컫는 이러닝 분야에서 교육(education)과 기술(technology)이 접목된 에듀테크(edutech)에 대한 연구가 활발하게 진행되고 있다. 그러나 디지털 기기에서 자동으로 수집이 가능한 학습활동 데이터를 기반으로 학습자 개인에게 맞춤형 학습을 제공하는 연구는 많으나, 오프라인 학습에서 추출하고 활용해야 할 데이터의 수집 연구는 적다. 이에 본 연구는 데이터 수집 연구를 위해 인공지능 컴퓨터 비전 기술을 이용하여 교재 또는 문제지의 문항 검출 방법을 연구한다. 이는 교재 또는 문제지에 대한 디지털로의 변환작업 없이도 오프라인 학습활동 데이터를 수집·저장·분석하여 지능화 교육 서비스와 연계를 통해 오프라인 학습에서도 학습자의 개인 맞춤형 학습 서비스 제공한다.

키워드 : 딥러닝, 문항인식모형, 검출모형, 컴퓨터 비전

## 1. 서 론

지능 정보화 시대에서 지능형 교육 시스템(ITS: Intelligent Tutoring Systems)은 학습자가 교수자의 개입 없이 스스로 학습할 수 있도록 도와주는 컴퓨터 시스템을 말한다[1]. 이런 지능형 교육 서비스가 되려면 지능형 교육을 위한 교수학습 모형이 적용된 동영상, 문제은행, 온라인 수학 도구 등과 같은 디지털 콘텐츠와 학습관리시스템(LMS: Learning Management System) 또는 애플리케이션 구현이 필요로 한다.

최근 인공지능 기술이 전 산업 분야에 적용되어 사람이 처리하던 업무를 대체하고 있다. 특히 사람의 눈과 같은 역할을 하는 컴퓨터 비전(computer vision) 기술은 빅데이터를 통한 딥러닝 학습을 통해 많은 산업 분야에서 사람보다 더 좋은 성능을 나타내기도 한다. 이러한 인공지능 기술은 교육 분야에서도 빅데이터를 통해 학습자 맞춤형 교육, 교수법 및 행정 관리 자동화와 지능화를 가능하게 할 것으로 전망되고 있다. 하지만 교육 부문의 디지털 전환이 다른 부문에 비해 상대적으로 더디게 진행되어 인공지능 기술의 접목도 더디게 접목되고 있다[2].

본 논문에서는 지능형 교육 시스템에 학습자의 학습활동 자료를 전송할 때 인공지능 딥러닝 컴퓨터 비전 기술을 이용하여 교재 또는 문제지에서 다양한 형식의 문항을 검출(detection)한다. 검출된 문항은 디지털로의 변환작업 없이도 오프라인 학습활동 데이터를 수집·저장·분석하여 지능화

※ 본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터 육성지원사업의 연구결과로 수행되었음(IITP-2021-2020-0-01789).

† 비 회 원 : 월드버텍 주식회사 대표이사

\*\* 비 회 원 : 한국열린사이버대학교 디지털비즈니스학과 석좌교수

\*\*\* 종신회원 : 전주대학교 컴퓨터공학과 조교수

Manuscript Received : October 6, 2021

Accepted : October 12, 2021

\* Corresponding Author : Ji Su Park(jisupark@jj.ac.kr)

교육 서비스와 연계를 통해 오프라인 학습에서도 학습자의 개인 맞춤형 학습 서비스 제공한다.

## 2. 관련 연구

딥러닝 기반 객체 검출 기술은 사람이 가장 많은 정보를 받아들이는 시각 정보를 컴퓨터가 대신하여 분석하고 해석할 수 있도록 하는 연구 분야이다. 이는 기계가 주어진 이미지에 포함된 대상이 무엇인지, 또한 어느 위치에 있는지 등을 파악하도록 하는 것을 주된 목표로 한다. 딥러닝을 이용한 객체를 검출하는 기법에는 구조적으로 1단계 수행 방법과 2단계 수행 방법이 있다. 1단계 수행 방법은 모든 영역에 대해서 지역화와 분류를 순차적으로 수행하여 딥러닝 학습 및 검출 속도가 빠르나, 정확성이 떨어진다는 단점이 있다. 2단계 수행 방법은 입력으로 들어온 영상에서 먼저 지역화를 수행한 후 선출된 후보 영역들에서 분류와 세밀한 지역화를 하므로 정확성은 매우 높으나 속도가 매우 낮은 단점이 있다. 1단계 수행 방법의 대표적 기법으로는 YOLO, SSD, RetinaNet, RestNet, ResNeSt가 있으며, 2단계 수행 방법의 대표적 기법으로는 R-CNN, SPPNet, Fast R-CNN, Faster R-CNN, R-FCN 등이 대표적이다[3,4].

### 2.1 ResNet

ResNet(Residual Network)은 마이크로소프트에서 개발한 네트워크로 지금까지 보다 층을 더 깊게 할 수 있는 특별한 장치를 가진다. 층을 깊게 하는 게 네트워크 성능 향상에 중요하지만 층이 지나치게 깊으면 학습이 잘되지 않고 성능이 떨어지는 경우도 많이 있다. ResNet에서는 그런 문제를 해결하기 위해서 스킵 연결(skip connection)을 사용한다. 스킵 연결은 입력 데이터를 합성곱 계층을 건너뛰어 출력에 바로 더하는 구조를 말한다. 이 구조가 층의 깊이에 비례해 성능을 향상시킬 수 있게 한 핵심이다.

ResNet은 층의 개수에 따라 ResNet-18, ResNet-34, ResNet-50, ResNet-101, ResNet-152 등 5가지 버전으로 나타낼 수 있다. 층 개수를 많이 사용할수록 연산량과 파라미터 개수는 커지지만, 정확도도 좋아지는 효과를 얻을 수 있다. 기존 방식으로는 층을 무조건 깊게 한다고 성능이 개선되는 것은 아니었다. 그러나 ResNet의 핵심인 Residual Block의 출현으로 오차율이 개선되어 shortcut을 연결해서 잔차(residual)를 최소화 되게 학습한 효과가 있다[5][6].

### 2.2 YOLO

YOLO(Real-Time Object Detection)는 입력 영상을 제안된 신경망 합성곱 구조에 전파 과정을 거치는 객체 검출 알고리즘을 제안하였다. 입력 영상을 제안한 신경망 합성곱 구조를 통해 다수의 예측된 검출 영역과 영상을 격자(grid) 7×7로 구성하여 최종 검출 영역을 결정한다[4]. 격자 방식을

이용하면 region proposal의 오버헤드를 제거할 수 있다. 이처럼 YOLO는 관심 영역을 추출하기 위한 별도의 네트워크가 필요하지 않으며, 일련의 추론 과정을 통해 객체를 인식한다. PASCAL VOC2007과 2012 데이터 집합으로 실험한 결과, 초당 155장의 영상을 훈련하여 Faster R-CNN보다 월등히 빠른 훈련 속도를 보였다[7]. 하지만 인식 정확도는 정지영상에서는 Faster R-CNN과 비슷한 검출 성능을 가지지만 동영상에서는 다소 떨어지며, 특히 작은 물체를 인식하는 데에 어려움을 보였다[3]. YOLOv4는 MS COCO 데이터 집합에서 43.5%(65.7% AP50)의 평균 정밀도(AP)값을 얻었으며, Tesla V100에서 65FPS의 실시간 속도를 달성하여 속도에서 가장 빠르고 정확한 검출기를 뛰어넘었다[8]. 또한 YOLOv4는 YOLOv3에 비해 평균 정밀도와 임베디드 보드 처리속도(FPS)는 각각 10%와 12% 증가했다[9].

## 3. 딥러닝 기반 교재 문항 검출

본 연구는 학습자들이 교재 또는 문제지 학습을 진행하면서 사용자 단말기의 카메라부를 이용하여 복수의 문항이 있는 문제지에서 문항과 개념 및 고유번호인 페이지 번호를 검출한다. 딥러닝 기반 교재 또는 문제지의 문항을 검출하기 위해 Fig. 1과 같이 1단계, 교재 및 문제지 자료를 수집하고 전처리 과정을 통해 물체 검출을 딥러닝 훈련 데이터를 준비하고, 2단계, 딥러닝 데이터를 훈련한 후, 3단계, 훈련된 모델을 활용한 교재 및 문제지의 문항 검출 실험을 한다.

1단계는 교재 및 문제지 자료수집, 전처리 과정을 통한 학습데이터 준비과정이다. 수집한 교재는 온라인 도서 판매 사이트에서 초등학교 수학 교재 중 판매량이 높은 3종 교재를 구매하여 이미지로 스캔하고, 이미지 데이터를 한 페이지 기준으로 분리하였다. 데이터 훈련을 위한 꼬리표 작업 이미지 꼬리표 작업 전문 도구인 Microsoft VoTT(Visual Object Tagging Tool) 2.1.0을 사용하여 Table 1과 같이 656페이지, 2,371문항, 274개념, 31지문으로 총 3,726개에 꼬리표를 붙인다[10].

2단계에서는 딥러닝 데이터 모형의 데이터 훈련을 위한 데이터 훈련 서버(TITAN xp 12GB GPU, CUDA10.0, CuDDNN7.6.4, Tensorflow2.1 등)를 준비한 후 모형별 문항 검출 데이터 훈련을 진행한다. ResNeSt 모형은 수집 데이터의 부족과 훈련 서버의 GPU 메모리가 부족한 문제로 경량화된 ResNeSt-50을 사용한다.

3단계에서는 다양한 교재 이미지를 입력한 후 정확도를 평가하기 위한 테스트 서버 환경(Ubuntu, Jupyter Notebook, OpenCV)을 구축한 후 2단계에서 훈련된 훈련 데이터를 기반으로 문항 검출 정확도 평가를 진행한다. 더 높은 문항 검출 정확도를 위해 실험 교재 이미지에 대해서 Fig. 2처럼 흑백화(gray scale), 임계 이진화(threshold binarization), 적응적 임계화(adaptive threshold)를 적용하여 실험한다.

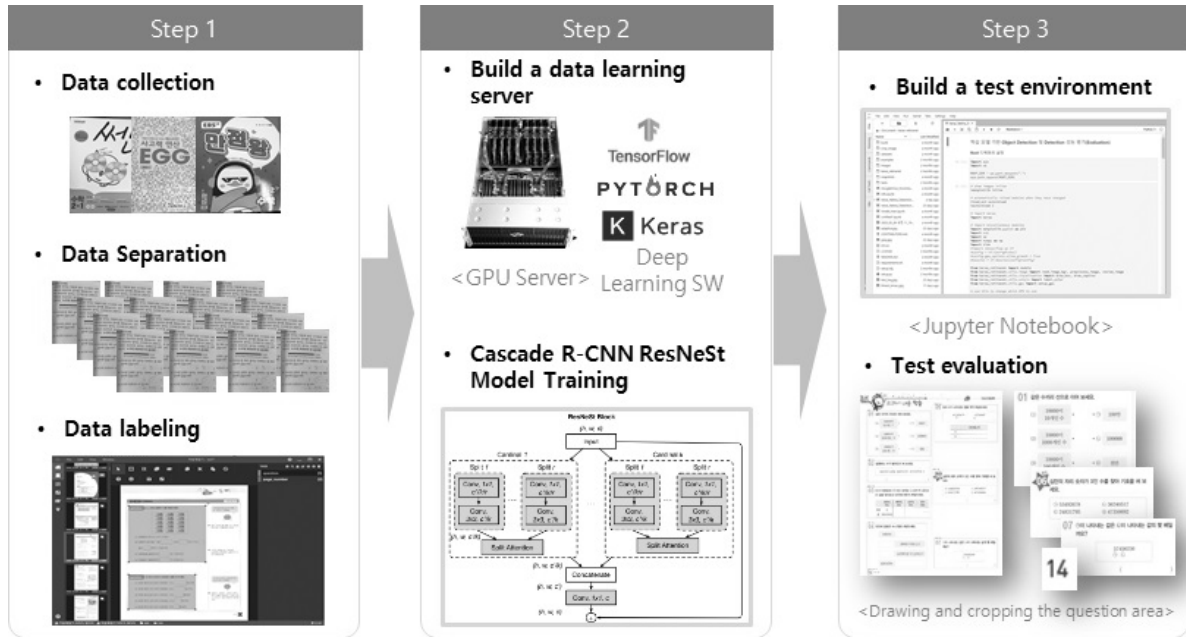


Fig. 1. 3 Steps of Deep Learning Based Textbook Item Detection

Table 1. Learning Data Tagging Status

Division	Page	Questionnaires	Page number	Concept	Text
Textbook1	220	994	210	50	11
Textbook2	184	527	160	81	5
	100	297	100	80	5
Textbook3	152	553	130	63	10
Sum	656	2,371	600	274	31

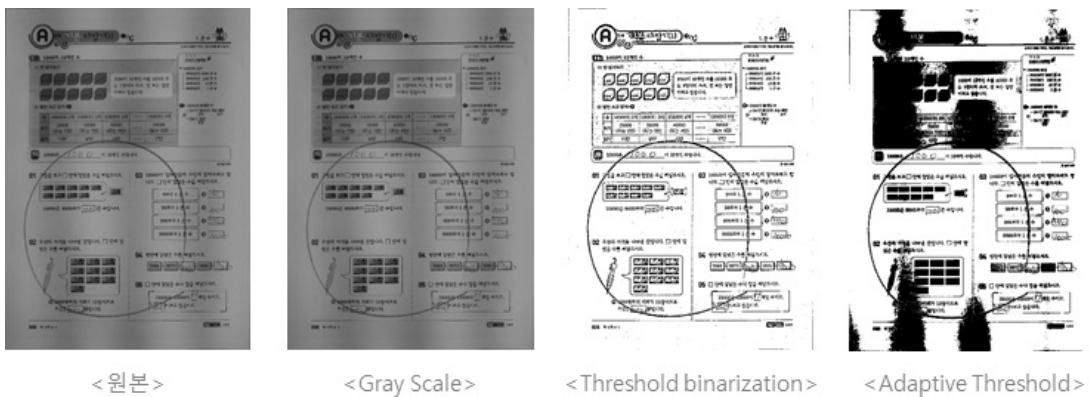


Fig. 2. Image Pre-processing Example Screen

#### 4. 실험 및 결과 분석

##### 4.1 실험 결과 확인 방법

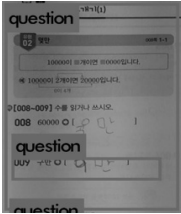
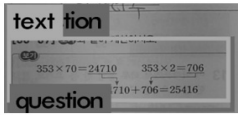
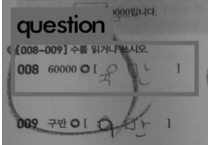
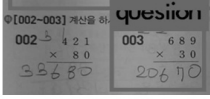
영상 이미지 객체 검출 평가는 육안검사로 Table 2와 같이 긍정 오류(false positive)와 부정 오류(false negative)를 더한 값을 평가 지표로 사용하였다[11]. 긍정 오류는 검출하지 말아야 할 객체가 검출된 것으로, 예를 들면 개념영역이

문항영역으로 검출되는 경우를 말한다. 부정 오류란 검출해야 하지만 검출하지 못한 객체를 의미한다. 예를 들면 문항으로 인식해야 하나 검출하지 못한 것을 의미한다.

##### 4.2 YOLOv4 실험 결과

YOLOv4 모형의 훈련 데이터를 이용한 실험 결과 Fig. 3의 좌측 최적화된 훈련 데이터 흑백 처리(best gray scale)

Table 2. False Positive and False Negative

Division	Example	
False Positive		
False Negative		

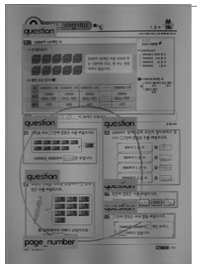
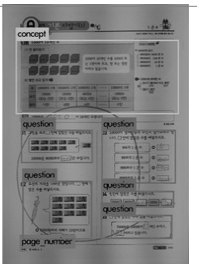
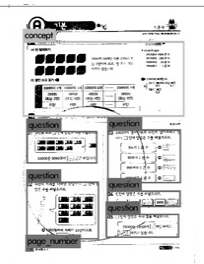
		
<b>Best - Gray Scale</b> question: 64% question: 100% page_number: 95% question: 100% question: 100% question: 100% question: 99%	<b>8000 - Gray Scale</b> concept: 98% page_number: 97% question: 100% question: 100% question: 100% question: 100%	<b>8000 - Threshold_Binarization</b> concept: 33% page_number: 97% question: 100% question: 100% question: 99% question: 100%

Fig. 3. Experimental Results of YOLOv4

이미지에서 인식 객체 6개 중 개념영역을 제외한 5개의 객체를 모두 95% 이상의 인식률로 인식하였다. 그러나 개념영역은 문항영역으로 오인식하였다. Fig. 3 중간의 8,000회 훈련

데이터 흑백 처리(8000-Gray Scale)에서는 인식 객체 6개 중 6개를 97% 이상의 인식률로 모두 인식하였다. 우측의 8,000회 훈련 데이터에서 임계 이진화 처리(8000-threshold binary coded)한 이미지는 인식 객체 6개 중 6개 모두 인식하였으나 개념영역의 인식률이 33%로 낮게 인식되었다. 16,000회 훈련 데이터에서 임계 이진화 처리한 이미지에서는 인식률이 33%에서 50%로 올라간 것을 확인할 수 있었다. 최적화 적응형 임계처리 이미지는 최적화된 훈련 데이터뿐만 아니라 그 외의 훈련 데이터에서도 인식률이 70%를 넘지 않았다.

YOLOv4 모형 훈련 결과 8,000회 훈련 데이터에서 흑백 전처리한 이미지가 높은 인식률을 보였다. Table 3은 8,000회 훈련 데이터로 실험용 데이터 12페이지, 140개 인식 객체로 성능 평가한 결과표이다. 실험용 데이터는 이미지 원본에 문항별 답안을 작성한 이미지와 답안과 채점을 포함한 이미지로 구분하여 실험하였다. 실험에 사용한 이미지는 전처리하지 않은 원본 이미지와 전처리를 한 이미지로 구성했다. 전처리는 흑백 전처리, 임계값을 60, 80, 100으로 적용한 임계 이진화 전처리, 가우시안(Gaussian)과 중앙(median)값을 각각 15, 20, 25 적용한 적응형 임계 전처리로 실험을 실행하였다. YOLOv4 모형에서 검출에 성능값을 보면 원본 133, 흑백 전처리 16, 임계 이진화 전처리 45, 51, 42, 적응형 임계 전처리 57, 62, 61, 55, 57, 76으로 평가되었다. 가장 좋은 성능을 보인 것은 흑백 이미지 전처리로 140 인식 객체 중 16개의 인식 객체를 오인식하였다. 원본 문항에 답안만 작성한 이미지와 답안과 채점을 포함한 이미지 중 답안만 작성한 이미지가 좋은 성능을 예상했으나 흑백 전처리의 경우 답안만 작성시 10개 인식 객체 오인식을 하였으나 답안과 채점이 포함된 경우에는 6개 인식 객체 오인식으로 채점이 포

Table 3. Performance Results of YOLOv4

	page	Original	Gray	Threshold			Gaussian			Mean		
				60	80	100	15	20	25	15	20	25
question	8	3	0	2	2	1	1	1	1	1	1	3
	10	10	0	2	3	2	2	2	2	2	3	3
	11	10	2	3	3	3	4	4	1	3	3	4
	12	10	4	4	1	1	2	1	2	2	2	1
	21	7	0	0	0	0	0	0	1	0	0	1
	38	5	0	0	0	2	2	2	1	2	3	3
	39	1	0	0	1	1	1	1	0	0	1	2
	41	6	0	0	3	3	7	6	6	6	7	7
	79	6	1	4	1	1	3	3	2	3	2	3
	81	11	3	4	4	1	4	4	4	1	3	3
answer	90	7	0	1	2	1	1	2	3	1	1	3
	8	5	0	2	0	0	1	1	2	2	1	2
	10	9	0	5	5	3	4	3	1	4	4	3
	11	1	2	3	3	3	3	4	4	5	4	5
	12	2	3	1	3	3	3	5	5	3	4	5
	21	4	0	0	0	2	0	1	3	0	0	5
	38	1	0	0	0	2	2	2	2	1	1	0
	39	7	0	1	1	0	0	1	0	0	0	1
	41	7	0	4	6	6	6	8	8	6	7	7
	79	5	0	3	2	1	3	1	3	2	1	4
grading	81	9	1	2	4	2	2	3	3	6	4	3
	89	3	0	2	3	4	4	5	4	4	4	6
	90	4	0	2	4	0	2	2	3	1	1	2
	sum	133	16	45	51	42	57	62	61	55	57	76

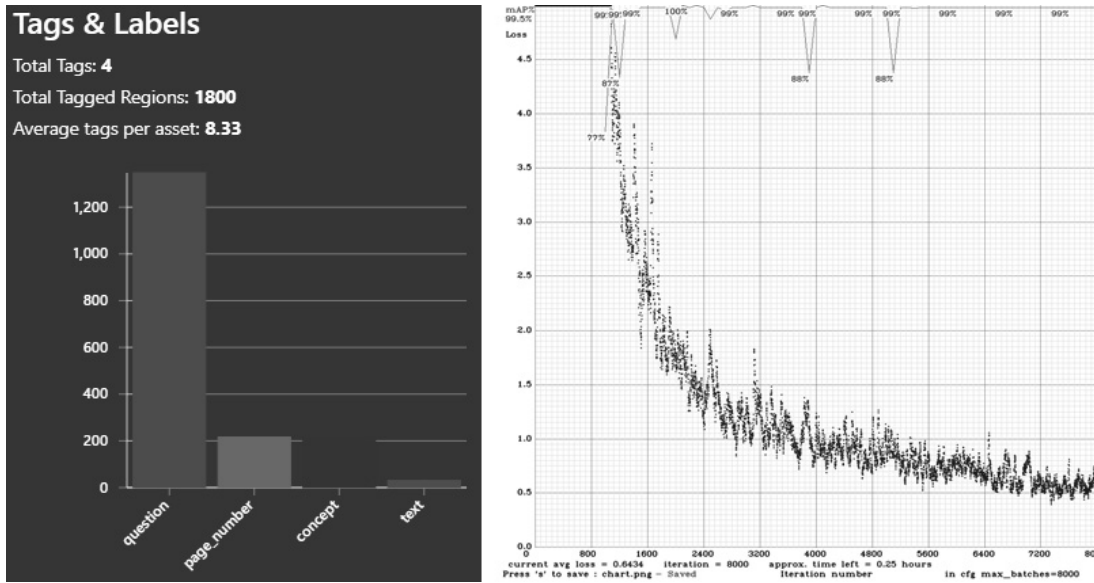


Fig. 4. Label Work Status and Learning Graph

함된 경우 성능이 더 좋음을 알 수 있었다.

흑백 전처리한 이미지의 오인식에 대해 분석해본 결과 Fig. 4와 같이 좌측의 지문 꼬리표(text) 작업에 대한 훈련 데이터가 부족 하여 과소 적합(underfitting)된 것으로 판단되었다. 지문 꼬리표에 대한 훈련 데이터를 추가로 수집한 후, 다시 실험하면 성능은 현 실험 결과보다 더 개선될 거라 예상할 수 있다.

#### 4.3 ResNeSt 실험 결과

ResNeSt 모형 훈련 데이터 실험은 YOLOv4와 같은 실험 데이터로 실험을 시행했다. 단, 실험에 사용한 임계 이진화 전처리 값을 80, 100으로 조정했으며, 적응형 임계 전처리 가

시안과 중앙값을 각각 5, 10, 15, 20, 25, 30으로 조정하여 실험했다. ResNeSt 모형 훈련은 ResNeSt 경량화 모형인 ResNeSt-50으로 하였으며, 학습 횟수는 95,000부터 27,000 회까지 5,000회씩을 추가 실험하여 학습 횟수별 인식률을 비교하였다. ResNeSt-50 훈련 데이터 실험 결과는 이미지 전처리를 하지 않은 원본 이미지의 경우 4, 흑백 전처리 22, 임계 이진화 전처리 44, 31, 적응형 임계 전처리 가우시안 적용 55, 58, 62, 54, 60, 61, 중앙값 적용 59, 57, 51, 50, 49, 51로 평가되었다. 가장 좋은 성능은 Table 4에서 보듯이 이미지 전처리 없는 원본 이미지로 140개 인식 객체 중 4개 인식 객체만 오인식하였다. ResNeSt-50 모형의 훈련 데이터 사용시 YOLOv4 8,000회 학습모형의 훈련 데이터 사용과 다

Table 4. Performance Results of ResNeSt

Page	Original	Gray	Thresh										Mean				
			80	100	5	10	15	20	25	30	5	10	15	20	25	30	
answer	8	0	0	5	1	0	1	1	1	1	1	1	0	1	2	2	1
	10	0	1	3	2	3	2	2	3	4	3	3	2	1	3	3	2
	11	0	2	3	2	3	2	3	3	3	3	3	3	2	2	3	3
	12	1	2	2	3	3	3	3	3	3	3	3	3	2	3	3	3
	21	0	0	1	0	0	0	0	0	1	3	0	0	0	0	1	3
	38	0	0	0	1	4	2	0	0	0	0	4	3	1	0	0	0
	39	0	1	0	0	2	2	1	0	4	0	2	2	2	0	0	0
	41	1	0	2	1	4	3	2	1	2	2	3	5	2	1	2	2
	79	0	2	4	1	2	4	4	3	4	2	3	3	4	4	4	2
	81	0	1	3	1	2	2	3	2	5	4	2	2	2	3	2	3
90	0	1	1	1	1	1	1	1	1	1	7	3	2	2	1	2	
grading	8	0	0	0	1	0	1	1	1	1	1	0	0	1	1	0	1
	10	0	1	3	2	3	3	2	2	2	1	3	3	2	2	2	1
	11	0	3	3	3	1	3	5	5	3	1	3	4	3	3	3	3
	12	0	3	3	2	3	3	7	4	4	4	3	3	3	3	3	4
	21	0	0	0	0	0	2	4	4	2	3	0	1	0	0	2	4
	38	0	0	0	2	3	3	3	0	2	3	3	3	4	4	0	2
	39	0	0	0	0	2	2	1	2	2	2	2	2	2	2	2	1
	41	0	0	2	1	5	4	3	3	2	1	4	4	3	4	2	2
	79	0	2	2	1	4	4	6	6	3	4	3	2	4	3	5	1
	81	0	1	2	3	3	2	3	3	5	4	3	2	2	2	2	4
89	1	0	2	3	3	4	2	2	3	4	4	3	3	2	2	4	
90	1	2	3	0	4	5	5	5	3	5	5	4	5	5	5	3	
Sum	4	22	44	31	55	58	62	54	60	61	59	57	51	50	49	51	

Table 5. Evaluation of Detection Performance According to ResNeSt-50 Learning Cycle Change

Page	270,000	95,000	100,000	105,000	110,000	115,000	120,000	125,000	130,000	135,000	140,000	145,000	150,000	155,000	160,000	210,000	260,000
answer	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	10	0	0	2	0	0	0	1	0	0	1	0	1	0	0	0	0
	11	0	2	1	1	0	0	1	0	1	0	0	0	0	0	0	0
	12	0	0	0	1	0	0	0	0	0	1	0	0	1	0	1	0
	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	38	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	39	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0
	41	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0	0
	79	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	81	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0	0
90	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Grading	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	11	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1	0
	12	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	38	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	39	0	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0
	41	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	79	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	81	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
89	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	
90	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	
Sum	1	4	5	2	1	1	4	3	3	3	0	1	2	2	2	1	2

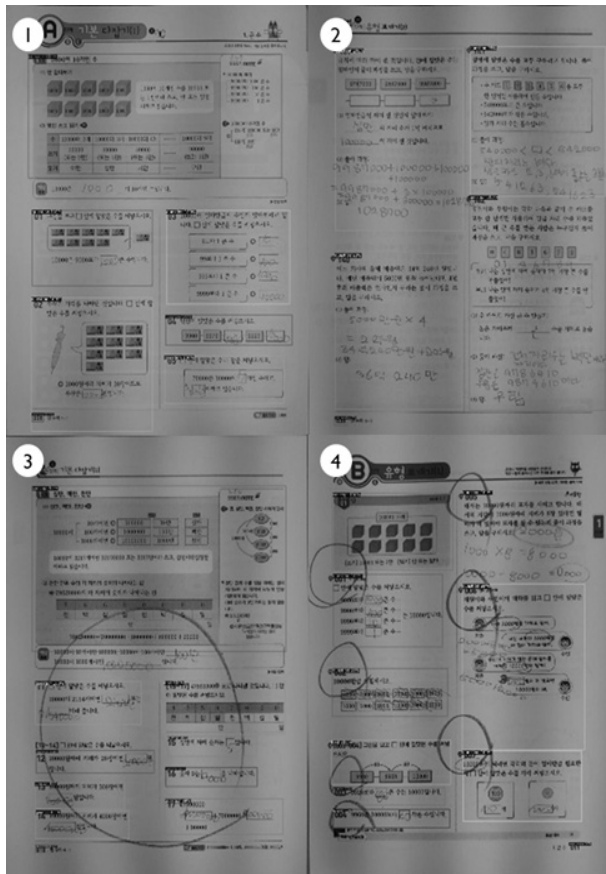


Fig. 5. Image of the Result of the Item Detection Experiment

르게 원본 문항에 답안만 작성한 이미지와 답안과 채점을 포함한 이미지와의 성능 차이가 거의 없음을 볼 수 있다.

ResNeSt-50 모형의 95,000 학습 횟수를 5,000회 간격으로 추가 학습한 결과 Table 5와 같이 학습 횟수가 140,000 회에서 140개 인식 객체를 모두 검출하였고, 학습 횟수가

140,000회보다 높았을 때는 오히려 오인식 객체의 개수가 증가하였다.

Fig. 5는 ①답안 기반 문항영역 인식, ②서술형 답안 기반 문항영역 인식, ③답안+채점1 문항영역 인식, ④답안+채점2 문항영역 인식한 실험 결과 이미지이다.

4.4 실험 결과 비교

YOLOv4 모형 성능과 ResNeSt 모형 성능을 비교한 결과 ResNeSt 모형 원본 이미지가 YOLOv4 모형 흑백 전처리 이미지보다 Table 6과 같이 좋은 성능을 보였다. ResNeSt 모형의 경우 가장 경량화된 ResNeSt-50 모형으로 학습했기 때문에 보다 기대치 평균 정밀도가 높은 ResNeSt-101, ResNeSt-200 모형으로 변경하면 더 좋은 성능을 보일 것이다.

4.5 실험의 한계 및 개선방안 도출

Table 6 모형별 성능 비교와 모형별 오인식에 대한 분석을 토대로 문항 검출 모형의 실험 한계와 개선방안을 네 가지로 정리하였다. 첫째, YOLOv4 모형 실험 결과 지문 꼬리표 데이터가 부족하여 과소 적합된 것으로 판단됨에 따라 지문 꼬리표에 대한 추가 이미지 데이터를 수집한 후, 재학습 진행이 필요하다. 둘째, 본 실험에서는 일정한 조명하에서 문항 이미지를 촬영하여 실험하였으며, 실험에 사용하는 이미지의 전처리 방법에 따라 성능이 변화하고 있음을 보여주고 있다. 향후 연구에서는 다양한 조명이나 잡신호가 성능에 미치는 영향에 관해서는 추가 연구가 필요하다. 셋째, 객체 인식 검출 모형마다 문항에 대한 특징지도 추출 방법이 상이하므로 YOLOv4, ResNeSt-50 모형 이외의 모형들에 대한 추가 탐색이 필요하다. 넷째, 본 연구는 문항을 검출하는 인식 모형에 관해 연구하였다. 향후 본 연구를 토대로 지능형 교육 시스템에서 활용하는 연구가 필요하다.

Table 6. Comparison of YOLOv4 Model and ResNeSt Model Performance

구분	Page	ResNeSt-Original	YOLOv4-Gray
answer	8	0	0
	10	0	0
	11	0	2
	12	1	4
	21	0	0
	38	0	0
	39	0	0
	41	1	0
	79	0	1
	81	0	3
grading	90	0	0
	8	0	0
	10	0	0
	11	0	2
	12	0	3
	21	0	0
	38	0	0
	39	0	0
	41	0	0
	79	0	1
81	1	0	
90	1	0	
Sum		4	16

### 5. 결 론

최근 인공지능 기술이 전 산업 분야에 적용되어 사람이 처리하던 업무를 대체하고 있다. 특히 사람의 눈과 같은 역할을 하는 컴퓨터 비전 기술은 빅데이터 딥러닝 학습을 통해 많은 산업 분야에서 사람보다 더 좋은 성능을 나타내기도 한다. 교육 분야에 적용되는 주요 에듀테크 기술은 AR/VR, AI, 로보틱스, 블록체인 등으로 파악된다. 체험 기반의 실감형 교육을 가능하게 하는 AR/VR 기술의 경우 2025년에는 2018년에 비해 7배 높은 성장을 보일 것으로 전망하며, 인공지능 기술 시장 또한 2018년 8억 달러에서 2025년 61억 달러 규모의 시장으로 전망하고 있다. 하지만 교육 부분의 디지털 전환이 다른 부분에 비해 상대적으로 더디게 진행되어 인공지능 기술의 접목도 더디게 접목되고 있다[5]. 이에 본 연구에서는 딥러닝 기반의 교재 문항 검출을 통해 전통적 교육방식의 대표 교육 형태인 교재 및 문제지 기반의 교육을 콘텐츠의 디지털 전환 없이 지능형 교육 시스템과의 상호작용이 가능한 토대를 마련하는 연구이다. 본 연구에서 제안된 교재 문항 인식 검출 모형으로는 YOLOv4와 ResNeRt 두 개를 선택하여 다양한 이미지 전처리 조건으로 실험한 결과 ResNeRt가 YOLOv4에 비해 높은 성능을 보였다. 하지만 본 연구의 대상이 초등학교 4학년 수학 3종의 교재 데이터를 기반으로만

연구한 것이고, 다양한 촬영 환경을 고려하지 않았기 때문에 대상 범위를 넓히고, 다양한 촬영 환경을 고려하여 추가 성능 검증할 필요가 있다. 교재에서 문항을 실시간으로 검출할 수 있다는 것은 많은 시간과 비용을 들여 오프라인 콘텐츠를 디지털 콘텐츠로 전환 없이 지능형 교육 시스템에 적용할 수 있다는 것이다. 현재 오프라인 기반의 학원, 출판사, 전통방식의 교육 서비스 사업자는 전통적 교육방식에서 에듀테크 서비스로의 변화하는데 비용과 시간의 투자가 필요하며, 아직은 디지털 네이티브 학생들에게 전통적 교육방식으로 교육하는 실정이다. 이런 측면에서 본다면 향후 온·오프라인이 결합한 에듀테크 서비스를 제공하는 경우, 본 연구의 결과를 활용한다면 빠른 기간 내에 효과적이며 효율적인 학습 서비스를 제공할 수 있을 것이다.

### References

- [1] J. W. Schofield, R. Eurich-Fulcer, and C. L. Britt, "Teachers, computer tutors, and teaching: The artificially intelligent tutor as an agent for classroom change," *American Educational Research Journal*, Vol.31, No.3, pp.579-607, 1994.
- [2] Y. Kong, "Edutech Industry Trends and Implications," *Software Policy & Research Institute*, Monthly Software Oriented Society, No.70, 2020.
- [3] 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, pp.1137-1149, 2017.
- [4] Y. H. Lee and Y. Kim, "Comparison of CNN and YOLO for object detection," *Journal of the Semiconductor & Display Technology*, Vol.19, No.1, pp.85-92, 2020.
- [5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Computer Vision and Pattern Recognition, pp.770-778, 2016.
- [6] X. Wu, D. Sahoo, and S. C. H. Hoi, "Recent advances in deep learning for object detection," *Neurocomputing*, Vol.396, pp.39-64, 2020.
- [7] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The Pascal Visual Object Classes (VOC) challenge," *International Journal of Computer Vision*, Vol.88, No.2, pp.303-338, 2010.
- [8] J. S. Lee, S. K. Lee, D. W. Kim, S. J. Hong, and S. I. Yang, "Trends on object detection techniques based on deep learning," *Electronics and Telecommunications Trends*, Vol.33, No.4, pp.23-32, 2018.
- [9] Alexey Bochkovskiy, Chien-Yao Wang, Hong-Yuan Mark Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," *Computer Science Engineering ArXiv*, 2020, [Internet], <https://arxiv.org/abs/2004.10934>

- [10] J. Solawetz, "Getting started with VoTT annotation tool for computer vision," 2020.07, [Internet], <https://blog.roboflow.com/vott/>
- [11] P. Henderson, and V. Ferrari, "End-to-end training of object class detectors for mean average precision," *Computer Science Engineering ArXiv*, 2017, [Internet], <https://arxiv.org/abs/1607.03476>



**김 태 종**

<https://orcid.org/0000-0002-2967-3319>  
 e-mail : ktj4820@gmail.com  
 2000년 금오공과대학교 산업공학(학사)  
 2003년~현 재 월드버텍 주식회사 대표이사  
 2017년~현 재 에듀테크협회 이사

2020년 한국방송통신대학원 이러닝학(석사)  
 관심분야 : 이러닝, 에듀테크, 인공지능, Deep Learning



**한 태 인**

<https://orcid.org/0000-0002-7250-1168>  
 e-mail : hanten55@daum.net  
 1982년 고려대학교 통계학과(학사)  
 1985년 고려대학교 통계학과(석사)  
 1999년 고려대학교 컴퓨터학과(이학박사)  
 2020년 한국방송통신대학교 이러닝학과 부교수

2020년~현 재 한국열린사이버대학교 디지털비즈니스학과 석좌교수  
 관심분야 : 이러닝, 에듀테크, 전자상거래, 정보통신정책



**박 지 수**

<https://orcid.org/0000-0001-9003-1131>  
 e-mail : jisupark@jj.ac.kr  
 2013년 고려대학교 컴퓨터교육과(박사)  
 2015년~2018년 충남대학교 초빙교수  
 2018년~2019년 경기대학교 교양학부 조교수

2019년~2020년 동국대학교 융합교육원 교수  
 2020년~현 재 전주대학교 컴퓨터공학과 조교수  
 2020년~현 재 한국정보처리학회 이사 및 JIPS 간사  
 2017년~현 재 HCIS 저널 편집위원  
 관심분야 : 분산 시스템, 모바일 클라우드 컴퓨팅, e-Learning, SW교육, 빅데이터 분석, IoT