

OCR 시스템에서 YOLOv4를 활용한 텍스트 이미지 전처리 연구

김하윤, 유상인, 주혜경, 최여진
한국폴리텍대학 서울강서캠퍼스 데이터분석과
let_hykim@naver.com, tkddls3626@naver.com,
hyegyeong25@gmail.com, choiyj0303@naver.com

A Study on Preprocessing Image Text Using Yolov4 in OCR System

Ha-Yoon Kim, Sang-Yin Yu, Hye-gyeong Ju, Yeo-jin Choi
Department of Data Analysis, Seoul Gangseo Campus of Korea Polytechnics

요 약

본 연구는 유료 OCR 서비스를 이용하여 야외 촬영 이미지의 텍스트를 검출하는 프로젝트에서 야외 촬영 텍스트를 학습시킨 Yolov4 모델을 통한 전처리 작업을 제안한다. 텍스트 감지를 통한 이미지 텍스트 전처리 진행은 불필요한 OCR 실행을 줄여 리소스를 절약하고 유료 서비스의 경우 비용 절감 효과까지 도모할 수 있다는 장점이 있다.

1. 서론

COVID-19(코로나19)의 여파로 국내 중위권 성적 학생의 비중은 전면 비대면 수업이 진행됨에 따라 각각 75.9%, 66.1%의 중, 고등학교에서 감소했다.

<표 1>

사교육걱정없는세상, 2020 코로나19 학력 격차 실태

중위권 (B,C,D등급)	학기	고1 전국				중2 전국			
		2018-2	2019-1	2019-2	2020-1	2018-2	2019-1	2019-2	2020-1
유위권 감소율 (<0)	해당학교 개수	199개교	220개교	177개교	270개교	440개교	438개교	403개교	646개교
	(소서울상 일국대내) 해당학교 비중에	48.7%	53.9%	43.3%	66.1%	51.7%	51.4%	47.3%	75.9%

2020년 한국교육학술정보원에서 실시한 설문 조사 결과는 이러한 학습 격차 심화의 이유로 62.92%가 학생의 자기 주도적 학습 능력 차이인 것으로 밝혔다. 이에 본팀은 학생들의 자기주도적 학습 능력 향상을 통한 학습 격차 심화 현상을 완화시키기 위한 방법으로 OCR(광학 문자 인식) 기반의 학습 문제 출제 어플리케이션을 개발하였다. 사용자는 해당 어플리케이션을 통해 이미지를 업로드한 후 OCR 시스템을 거쳐 추출된 텍스트를 기반으로 한 학습 문제를 풀이할 수 있다. 본팀이 차용한 OCR 시스템은 Google Cloud Vision API로, Google사에서 제공하는 머신러닝 기반의 이미지 분석 API이다.

Google Cloud Vision API는 강력한 OCR 기능을 제공하지만 무료 체험 기간 이후에는 월 1,000장이 넘어갈 경우 1,000 장당 \$2.5 ~ \$5 달러의 비용이 발생한다는 단점이 있다.

본팀의 프로젝트와 같이 야외 실제 촬영 이미지를 입력받는 OCR 시스템에서는 입력으로 들어오는 이미지의 유효성 검사가 어렵다. 따라서 유효하지 않은, 이미지 데이터로서 효용성이 없는 이미지는 텍스트가 인식되지 않거나 올바르게 읽지 않은 결과값을 초래하게 된다. 때문에 야외 촬영 이미지를 입력받는 OCR 시스템에서는 올바른 입력 이미지를 받기 위해 반복적인 OCR 함수의 호출이 필요하며 Google Cloud Vision API를 사용할 경우 반복적인 API 호출로 이어져 개발자에게 비용 부담을 유발한다.

이에 본팀은 객체 탐지 모델인 Yolov4에 텍스트 이미지 데이터를 학습시켜 OCR 시스템에서의 텍스트 이미지를 전처리하기 위한 방법을 고안하였다.

2. 관련연구

YOLO(You Only Look Once)는 One-stage 방식의 모델로 구조를 단순화하여 빠르게 동작할 수 있도록 하는 모델이다. YOLO의 장점은 전체 이미지에 단일 신경망을 적용하여 다량의 작업을 한번에

처리할 수 있고, 보편적인 Object Detection의 특징으로 영역을 나누어 바운딩 박스 및 일치 확률을 예측할 수 있다는 점이다. 예측된 확률은 바운딩 박스에 가중치를 부여하여 더욱 정확한 예측치를 얻을 수 있다. 또한 한 장의 이미지에 대해 수천 장이 요구되는 R-CNN보다 빠르다는 장점이 있다. 특히 Yolov4는 GPU에서 실시간으로 동작하면서, 기존에 사용하던 하나의 GPU만으로 매우 빠른 학습이 가능한 새로운 모델이다. Yolov4는 최신 딥러닝 기법을 적극적으로 활용하여 모델을 설계하고, 학습시켜 정확도와 속도 측면에서 모두 성능을 크게 끌어올린 모델이다.[1]

3. Yolov4 전처리 모듈 제안

제안하는 Yolov4 전처리 모듈은 OCR 시스템에서 텍스트 이미지 전처리를 가능하게 한다. 본 연구에서는 Yolov4 모델에 약 200여장의 텍스트 이미지 데이터셋을 학습시켜 OCR 시스템의 전처리 작업을 수행하는 데 이용하였다. <그림 1>은 Yolov4 모델에 학습시킨 텍스트 이미지 데이터셋의 일부이다.



<그림 1>

본 연구에 사용될 Yolov4 모듈은 텍스트 이미지를 OCR 시스템으로 입력 시 전처리 작업만 수행할 것이기 때문에 이미지를 분석할 때 내부의 텍스트 존재 여부만 감지하면 된다. 때문에 학습 설정 파일의 .data 파일의 classes 속성은 <그림 2>와

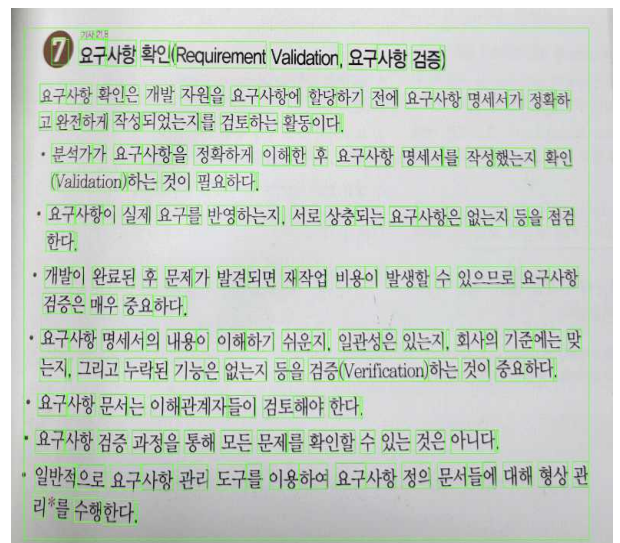
같이 'Text', 'notText'의 두가지로 설정한다.

```

1 classes = 2
2 train = data/list/train.txt
3 valid = data/list/valid.txt
4 test = data/list/test.txt
5 names = data/ClassNames.names
6 backup = backup
    
```

<그림 2>

학습된 Yolov4 모듈은 입력받은 이미지에서 텍스트를 검출하여 바운딩 박스로 표시한 후 OCR 시스템에 해당 이미지를 넘겨 OCR 작업을 진행한다. 위의 과정으로 학습시킨 Yolov4 모듈의 텍스트 감지 결과는 <그림 3>과 같다.



<그림 3>

본 연구의 Yolov4는 예측 결과에서 'Text'로 인식되는 바운딩 박스가 발견되지 않으면 텍스트가 검출되지 않는 이미지로 간주하고 OCR 소스를 실행시키지 않게 된다.

4. 결론

본 연구를 통해 구현한 Yolov4 모듈을 사용하여 전처리 작업을 통해 불필요한 OCR 실행을 줄여 리소스를 절약할 수 있음을 알 수 있다. 뿐만 아니라 Google Cloud Vision API와 같이 호출 건당 비용이 부과되는 OCR시스템을 사용할 경우에는 전처리 모듈을 사용함으로써 유효한 텍스트

이미지가 들어올 때까지 OCR을 실행시키지
않으므로 프로젝트 비용 또한 절약할 수 있을
것으로 사료된다.

참고문헌

[1] Roger S. Pressman "Software Engineering A
Practliners' Approach" 3rd Ed. McGraw Hill

- 본 논문은 과학기술정보통신부 정보통신창의인재
양성사업의 지원을 통해 수행한 ICT멘토링 프로젝
트 결과물입니다.