

# 영수증 글자 검출

## Receipt Text Detection

---

OCR 3조

진성준 이현진 서재현

# Table of Contents

---

- 01** 대회 소개
- 02** 데이터 소개
- 03** Data Augmentation
- 04** Backbone
- 05** Post processing
- 06** 대회 결과
- 07** 회고

# 01 대회 소개

## 영수증 글자 검출(Receipt Text Detection)

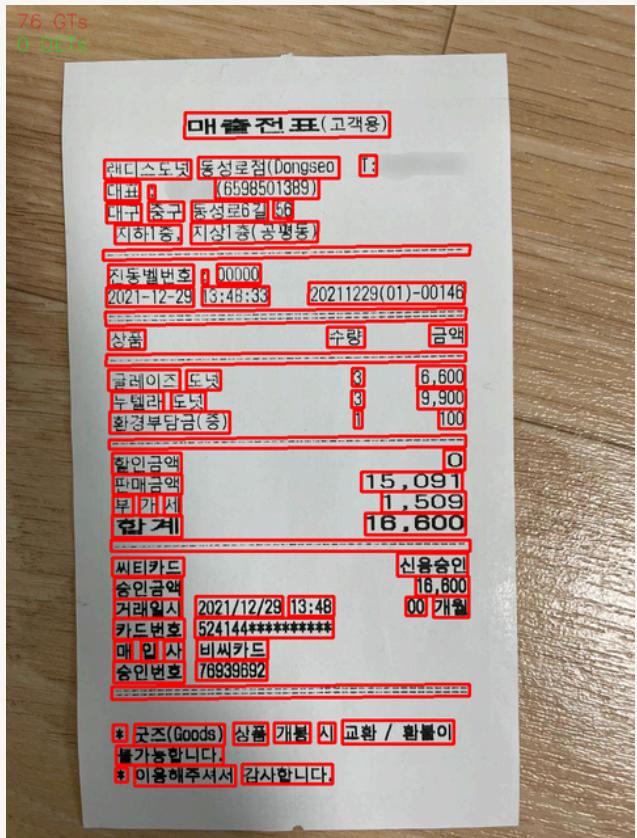
- 인공지능 모델을 이용하여 제공된 영수증 이미지에서 문자의 위치를 정확하게 검출하는 문제에 도전하는 대회
- 평가방법
  - CLEval Metric



# 02 데이터 소개

- JPG 형태의 영수증 이미지 파일
- 이미지 파일 내 텍스트 위치 정보를 표기한 JSON 파일

## Image(예시)



## JSON

- images
  - {파일이름}
  - words
    - {points}
      - [X, Y]
      - [X, Y]
      - [X, Y]
      - [X, Y]

## Train

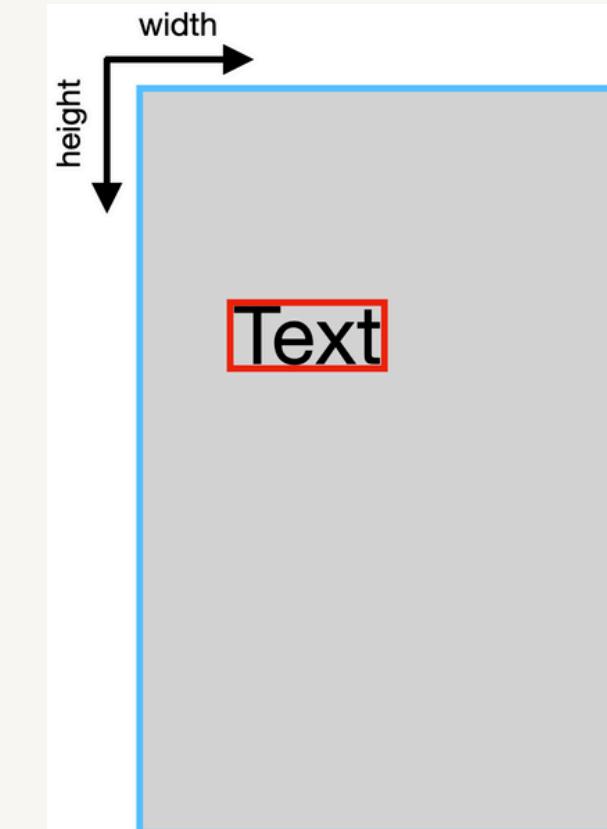
- 3,273장

## Val

- 404장

## Test

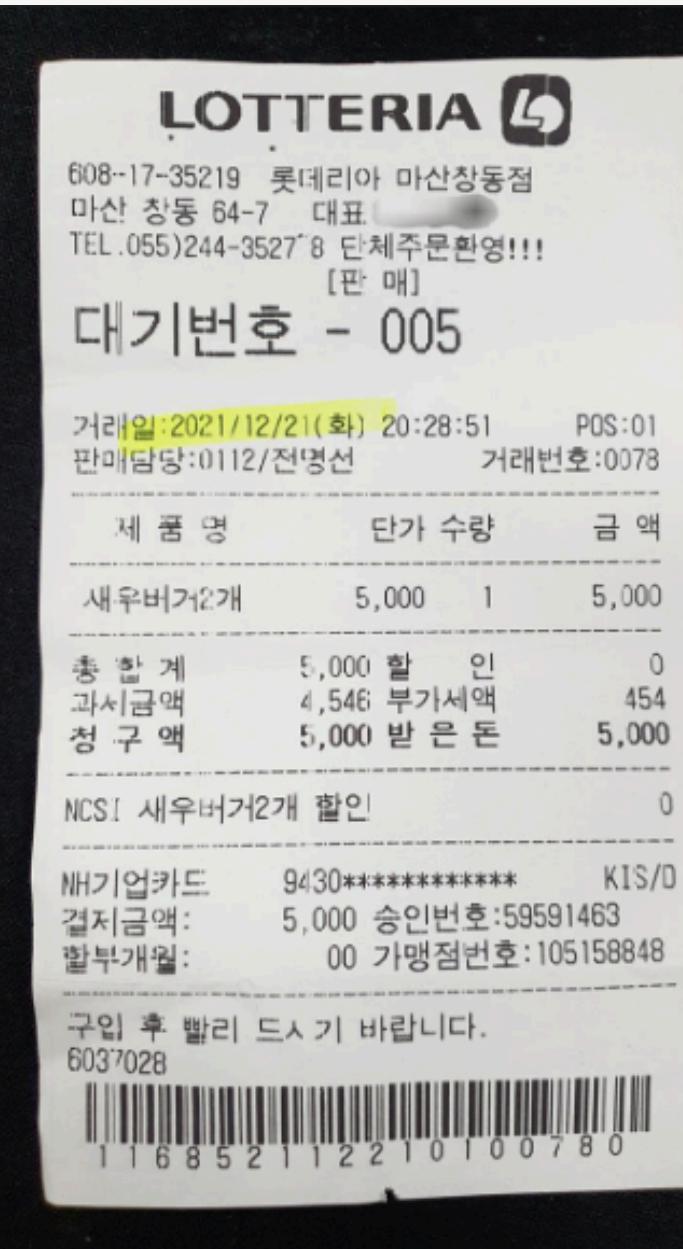
- 413장



```
{"images": {  
    "drp.en_ko.in_house.selectstar_000007.jpg": {  
        "words": {  
            "0001": {  
                "points": [  
                    [356, 990],  
                    [464, 990],  
                    [464, 1012],  
                    [356, 1012]  
                ]  
            },  
            "0002": {  
                "points": [  
                    [278, 988],  
                    [344, 988],  
                    [344, 1018],  
                    [278, 1018]  
                ]  
            },  
            "0003": {  
                "points": [  
                    [188, 984],  
                    [268, 984],  
                    [268, 1020],  
                    [188, 1020]  
                ]  
            }  
        }  
    }  
},  
"76 GTs": 0,  
"0 DETs": 0}
```

# 03 Data Augmentation

## 원본 이미지



대부분의 경우,  
깔끔한 영수증

왼쪽 이미지처럼  
형광펜 혹은 워터마크가 있는  
영수증도 존재

# 03 Data Augmentation

## Trial 1



텍스트 : JPEG 압축 구현, 잉크 밀림 구현

노이즈 : 문서 전체 노이즈

기타 :

- 불확실한 데이터 학습을 위한 추가 선 생성
- 배경 색 및 밝기 변경
- offline augmentation -> 데이터 3배 증강

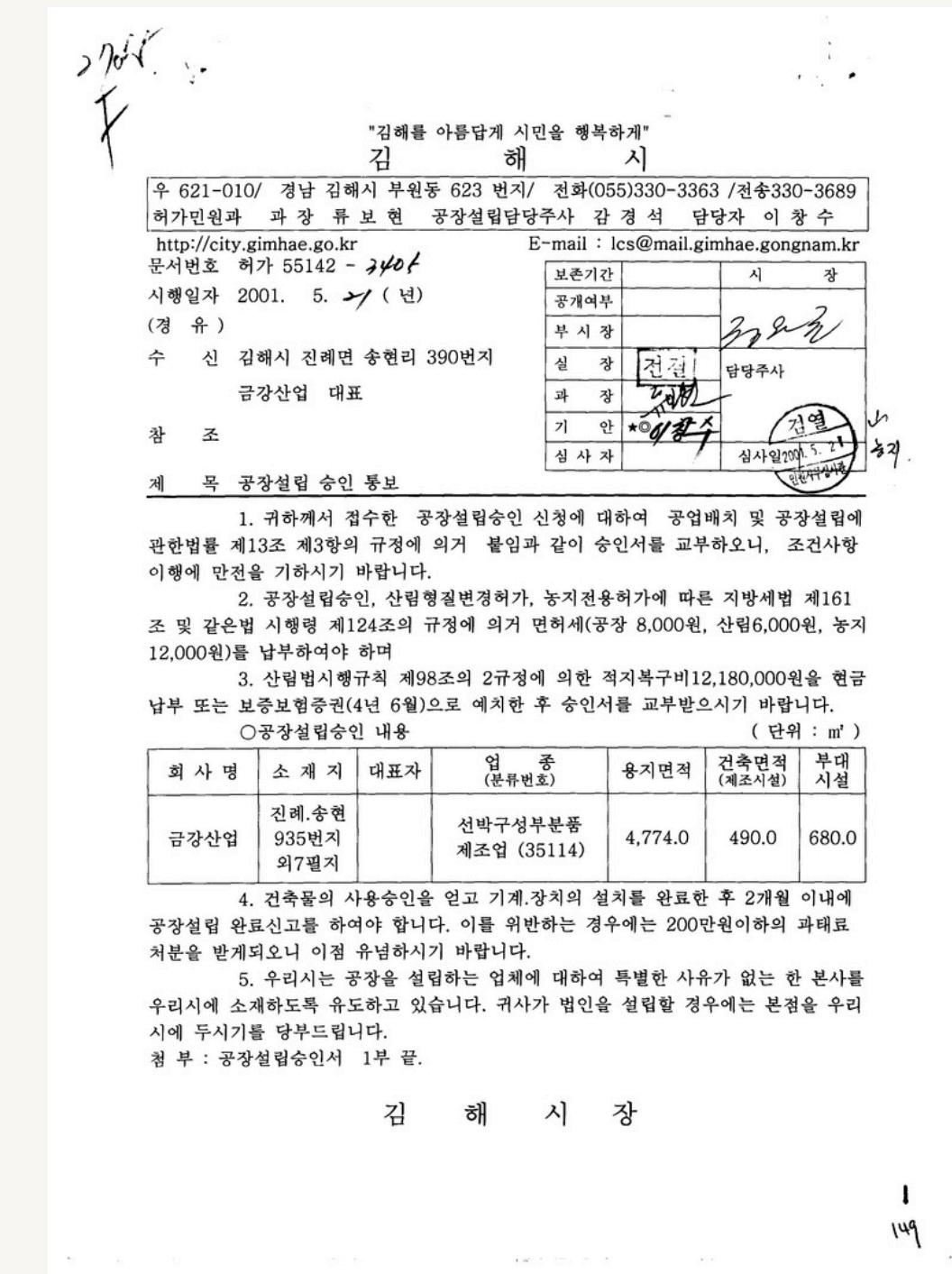
변경 전	평가지표	변경 후
0.9603	H-mean	0.9555
0.9759	Precision	0.9799
0.9474	Recall	0.9356

# 03 Data Augmentation

## trial 2 : 추가 데이터 - “공공행정문서 OCR”

The screenshot shows the AI Hub website's 'Data Search' section. A search bar at the top contains the query '공공행정문서 OCR'. Below it, a card for the 'Public Administration Document OCR' dataset is displayed. The card includes a preview image of a document, tags like '#광학글자인식', '#활자체', '#공공행정 문서', '#글자체', '#공공행정 업무유형', and '#문자형태', and filters for '분야' (한국어), '유형' (이미지). It also shows download statistics: 구축년도: 2020, 갱신년월: 2021-10, 조회수: 12,100, 다운로드: 1,812, 용량: 384.89 GB. Buttons for '다운로드' (Download) and '샘플 데이터' (Sample Data) are present. At the bottom, there is a file download link: [원천]train1.zip | 21.83 GB | key: 46302.

- 생성시점이 오래되어 스캔, 촬영 화질이 좋지 않음
- 다양한 문자형태(수가, 인쇄체, 타자체 등)를 포함
- 공공행정문서에 특화된 OCR 모델 개발을 위한 데이터셋



# 03 Data Augmentation

## trial 2 : 추가 데이터 - “공공행정문서 OCR”

```
{"images": {  
    "drp.en_ko.in_house.selectstar_000007.jpg": {  
        "words": {  
            "0001": {  
                "points": [  
                    [356, 990],  
                    [464, 990],  
                    [464, 1012],  
                    [356, 1012]  
                ]  
            },  
            "0002": {  
                "points": [  
                    [278, 988],  
                    [344, 988],  
                    [344, 1018],  
                    [278, 1018]  
                ]  
            },  
            "0003": {  
                "points": [  
                    [188, 984],  
                    [268, 984],  
                    [268, 1020],  
                    [188, 1020]  
                ]  
            }  
        }  
    }  
}.
```

대회 데이터

```
{"images": [{}{'image.make.code': '5350178',  
    'image.make.year': '1999',  
    'image.category': '인.허가',  
    'image.width': 2478,  
    'image.height': 3492,  
    'image.file.name': '5350178-1999-0001-0266.jpg',  
    'image.create.time': '2020-12-17 00:23:54'}],  
    'annotations': [{}{'id': 0,  
        'annotation.type': 'rectangle',  
        'annotation.text': '"김해를',  
        'annotation.ttype': 'textType1',  
        'annotation.bbox': [839, 309, 174, 53]},  
        {'id': 1,  
            'annotation.type': 'rectangle',  
            'annotation.text': '아름답게',  
            'annotation.ttype': 'textType1',  
            'annotation.bbox': [1032, 309, 193, 59]},  
            {'id': 2,  
                'annotation.type': 'rectangle',  
                'annotation.text': '시민을',  
                'annotation.ttype': 'textType1',  
                'annotation.bbox': [1244, 312, 156, 59]},  
                {'id': 3,  
                    'annotation.type': 'rectangle',  
                    'annotation.text': '행복하게"',  
                    ...  
                    {'id': 170,  
                        'annotation.type': 'rectangle',  
                        'annotation.text': '배',  
                        'annotation.ttype': 'textType1',  
                        'annotation.bbox': [2043, 920, 194, 130]}]}]
```

“공공행정문서 OCR” 데이터

## Annotation

- 대회 : Polygon
- “공공행정문서 OCR” : bbox

H-mean

0.9572

Precision

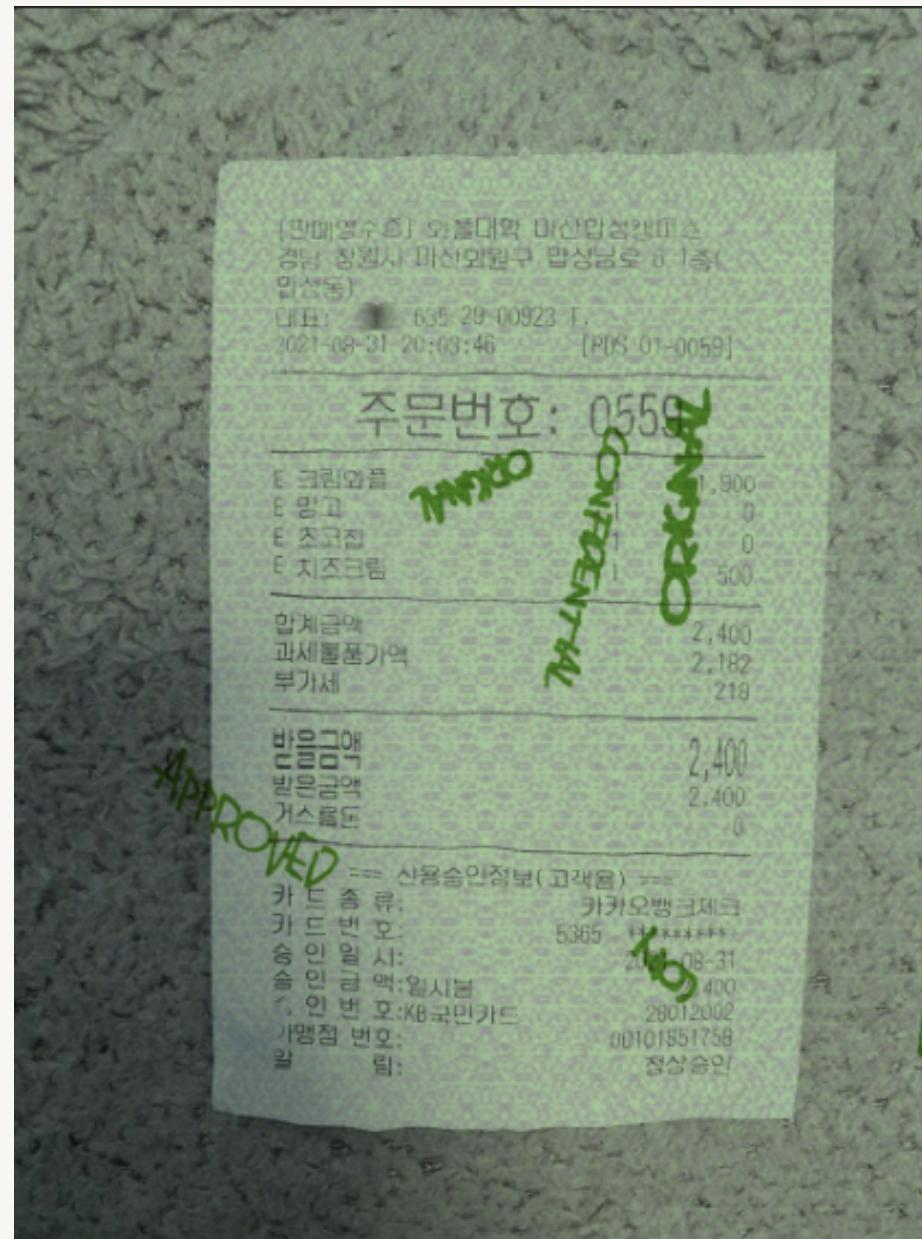
0.9791

Recall

0.9392

# 03 Data Augmentation

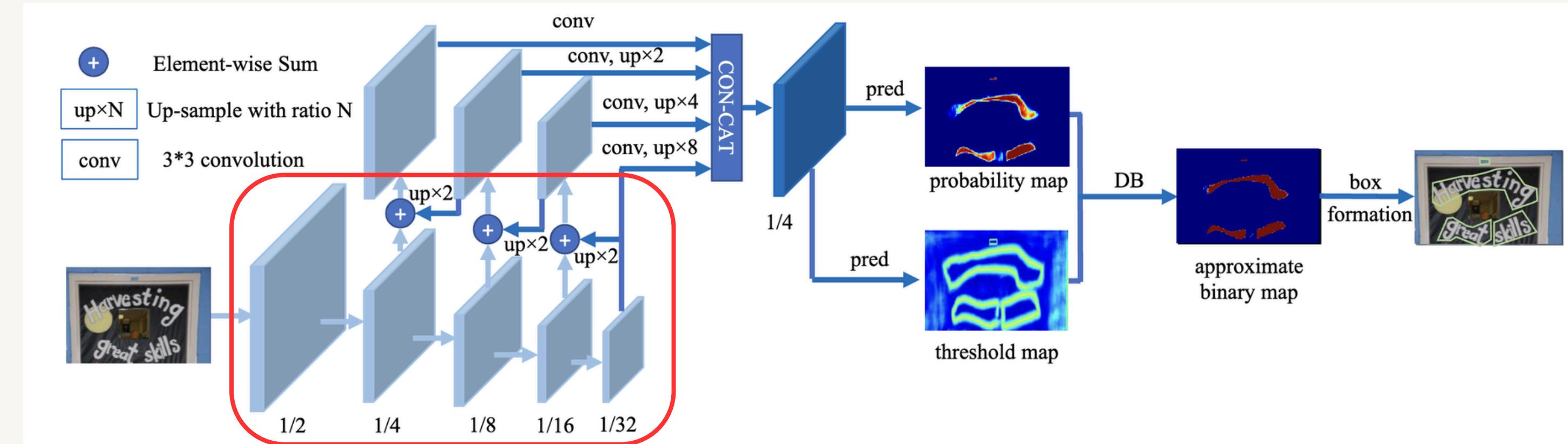
## 최종 : Augraphy augmentation



텍스트와 배경에 약간의 노이즈  
를 추가하고, 형광펜과 워터마크  
를 추가하여 데이터 증강

변경 전	평가지표	변경 후
0.9603	H-mean	0.9685
0.9759	Precision	0.9708
0.9474	Recall	0.9673

# 04 Backbone



Backbone model

Encoding 과정에서 사용되는 사전학습 된 Image Model

timm을 이용한 baseline code를 통해 다양한 Backbone model을 실험

resnet18, resnet50, EfficientNet\_b0, Efficientnet\_b3, mobilenet, ConvNext 등

# 04 Backbone

Trial	사용 모델	val_precision	val_recall	val_hmean	비고
	Resnet18	0.9759	0.9474	0.9603	baseline
	Resnet50	0.9772	0.9274	0.9495	
	EfficientNet_b0	0.9789	0.9666	0.9721	
	EfficientNet_b3	0.9788	0.9683	0.9730	
	MobileNet	0.9816	0.9414	0.9603	
	<b>ConvNext</b>	<b>0.9816</b>	<b>0.9797</b>	<b>0.9805</b>	

# 05 Post Processing



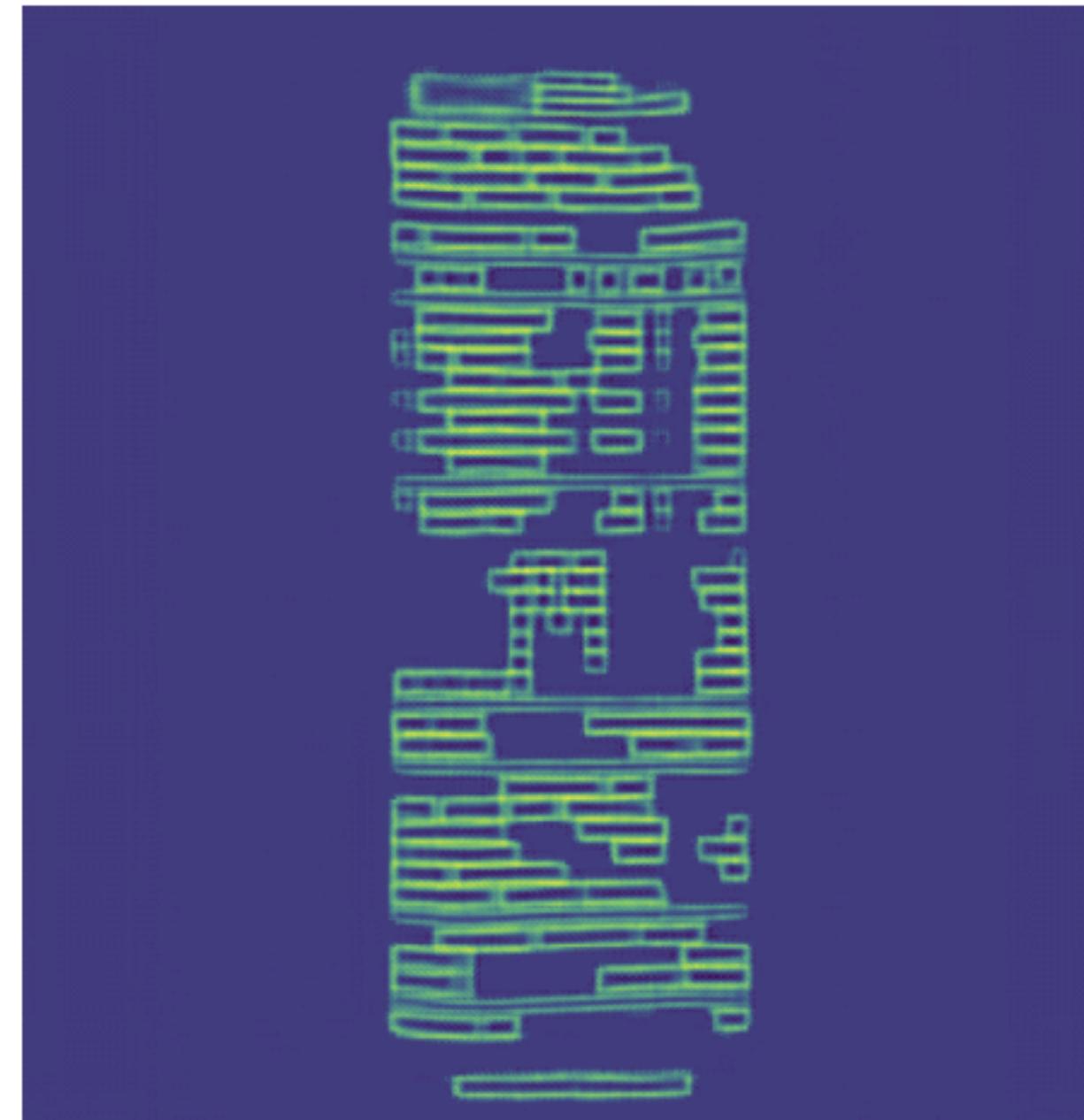
" ----- "

위와 같은 선도 검출을 해야 하지만,  
검출해내지 못하는 경우가 발생



Model에 문제가 있는지,  
Post Processor에 문제가 있는지  
확인 필요

# 05 Post Processing



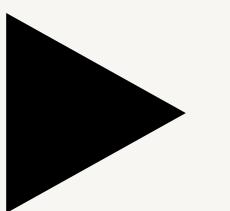
Model의 probability map과 threshold map을 시각화하여 살펴보면 “-----” 선과 글자들을 잘 검출하고 있음



Post Processor 파라미터에 문제 가 있음을 파악

# 05 Post Processing

```
postprocess:  
  thresh: 0.3  
  box_thresh: 0.4  
  max_candidates: 300  
  use_polygon: False
```



```
postprocess:  
  thresh: 0.225  
  box_thresh: 0.4  
  max_candidates: 300  
  use_polygon: True
```

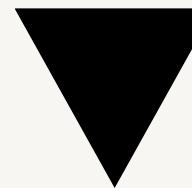
Thresh는 text 영역으로 검출 시  
사용하는 확률값의 임계값

Box Thresh는 text 영역에서 글  
자들이 한 단어인지 판단하는 (하  
나의 박스로 묶이는지 판단)임계값

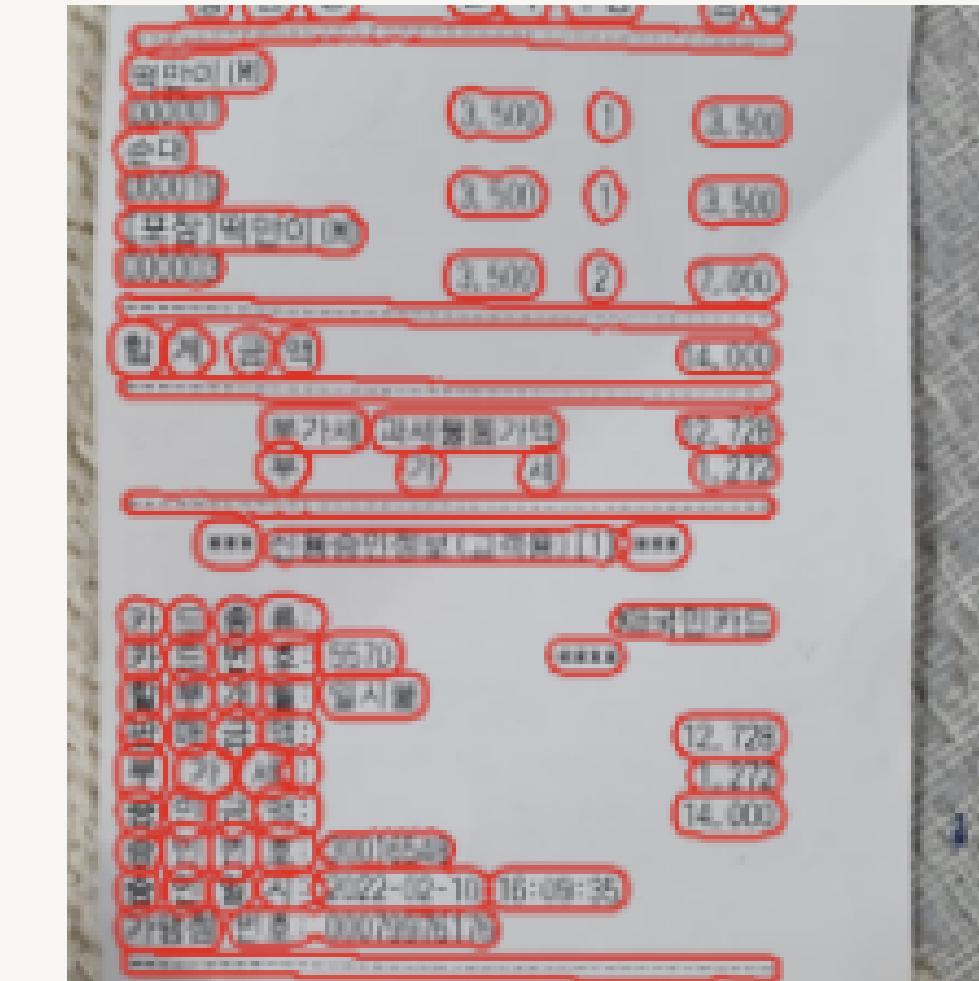
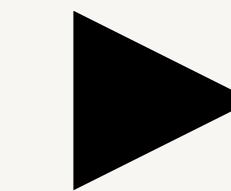
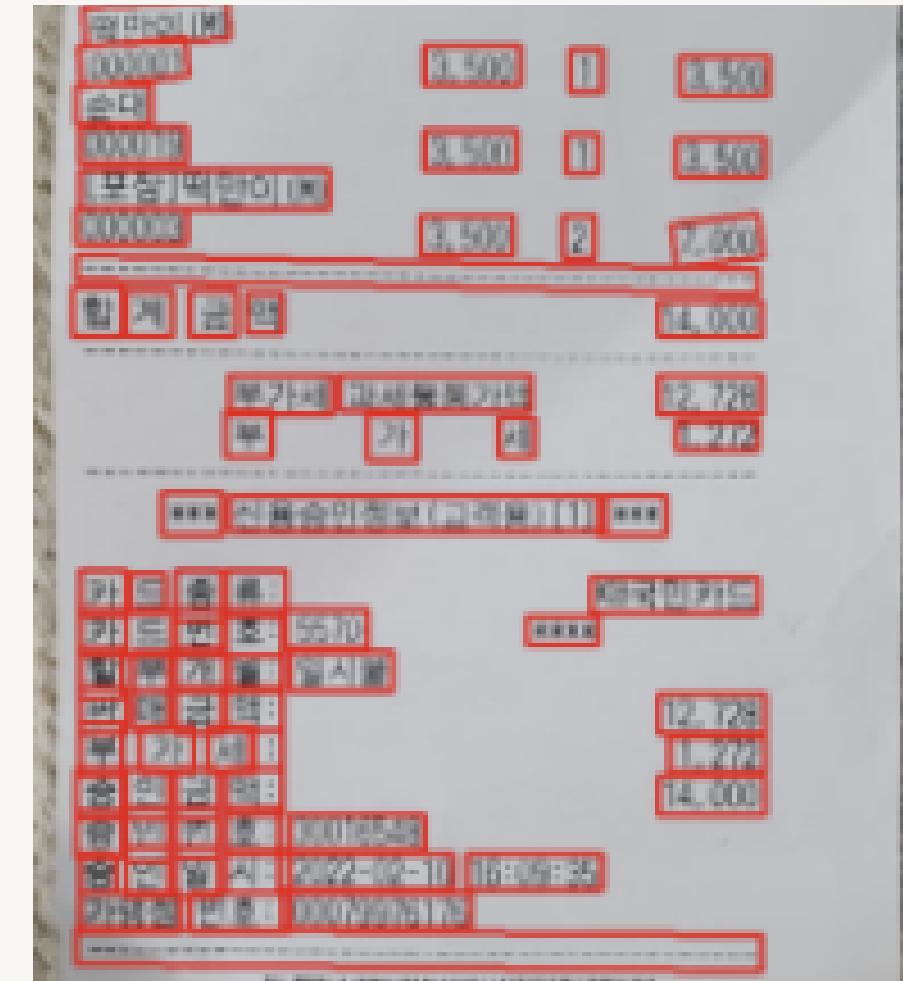
검출 결과를 시각화 했을 때, 검출된 것들의  
박스 정확도는 높은것으로 판단됨(Precision)  
“-----”과 같은 선이 검출되지 않아  
recall이 낮기 때문에, Thresh 값을 낮춰 더  
많은 영역이 검출되도록 값 변경

# 05 Post Processing

```
postprocess:  
    thresh: 0.3  
    box_thresh: 0.4  
    max_candidates: 300  
    use_polygon: False
```



```
postprocess:  
    thresh: 0.225  
    box_thresh: 0.4  
    max_candidates: 300  
    use_polygon: True
```



Box로 예측하는 것 보다 Polygon  
으로 예측하는 것이 상대적으로 더  
정밀하게 예측함

# 05 Post Processing

변경 전	평가지표	변경 후
0.8818	H-mean	0.9712
0.9651	Precision	0.9725
0.8194	Recall	0.9710

Post Processor의

Thresh 값을 조정하여  
recall을 높일 수 있었고,

Polygon으로 예측하여  
Precision을 높일 수 있었음

# 06 대회 결과

38	Finished	재현	0.9801	0.9838	0.9771	상세보기	2024-04-30 08:56			
----	----------	----	--------	--------	--------	------	------------------	--	--	--

34	Finished	재현	0.9805	0.9816	0.9797	상세보기	2024-04-29 17:43			
----	----------	----	--------	--------	--------	------	------------------	--	--	--

# 07 회고

## 진성준



- OCR 프로젝트의 프로세스를 체험 할 수 있었던 좋은 기회
- Data Processing 능력의 부족함을 체감
- 프로젝트를 도와주는 유ти리티를 체험 할 수 있어 좋았음

## 이현진



- Hydra 구조를 이용한 DL 과정을 직접 실습해 볼 수 있어 유익했음
- Augraphy를 이용해볼 수 있었음
- wandb를 이용해 실험 진행 상태를 팀과 공유 하며 진행할 수 있는 것이 흥미로웠음

## 서재현



- Hydra를 이용해 config 파일을 수정해가면서 편하게 여러 실험을 할 수 있어서 좋았음
- 검출기 뿐만 아니라 인식기를 활용해 최종 text 까지 인식해보면 좋을 것 같다

# Reference

[https://hydra.cc/docs/tutorials/basic/running\\_your\\_app/working\\_directory/](https://hydra.cc/docs/tutorials/basic/running_your_app/working_directory/)

<https://huggingface.co/timm>

[https://augraphy.readthedocs.io/en/latest/doc/source/example\\_usage.html](https://augraphy.readthedocs.io/en/latest/doc/source/example_usage.html)

[https://augraphy.readthedocs.io/en/latest/doc/source/list\\_of\\_augmentations.html](https://augraphy.readthedocs.io/en/latest/doc/source/list_of_augmentations.html)

[https://albumentations.ai/docs/api\\_reference/augmentations/](https://albumentations.ai/docs/api_reference/augmentations/)

<https://pytorch.org/docs/stable/optim.html>

감사합니다

---