

Data-Centric Wrap-up Report

CV-03

김승기 김준영 신우진 전형우 천지은

1. 프로젝트 개요

해당 프로젝트는 학습 데이터 추가 및 수정을 통한 이미지 속 글자 검출 성능 개선 프로젝트입니다. 글자를 검출하는 OCR 모델은 수정이 불가능하며 모델을 학습시키는데 필요한 loss 함수 또한 수정이 불가능합니다. 해당 프로젝트에서 수정 가능한 부분은 데이터를 변화시킬 수 있는 부분으로 데이터 추가, 수정, 증강이 가능합니다. 정리해보면 데이터를 중심으로 text detection의 성능을 올리는 프로젝트입니다.

2. 프로젝트 팀 구성 및 역할

팀원	역할
전체	EDA, Dataset Split, Online augmentation
김승기	offline augmentation, git convention
김준영	2-stage detector, neck, optimizer, Mmdetection baseline
신우진	swin based model
전형우	1-stage detector, YoloV8
천지은	offline augmentation, optimizer, pseudo labeling, ensemble

3. 프로젝트 목표

3-1. 팀 목표

팀의 목표는 프로젝트의 취지에 맞게 data를 직접 만져보는 것입니다. Data를 수정하는 방법에는 현재 있는 data를 직접 고치는 방법과 augmentation을 통해 추가적으로 다양한 data를 생성하는 방법이 있습니다. 이 중 주어진 학습 data를 직접 고쳐 성능을 올리는 것을 팀의 목표로 삼아 프로젝트를 진행했습니다.

4. 프로젝트 결과

4-1. EDA

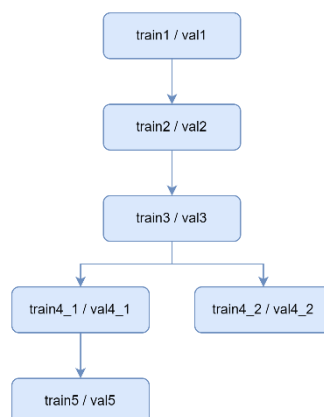
Text detection을 진행하기 위해 검출해야 하는 text들의 크기, 종횡비등 text들에 관련된 EDA를 진행했지만 큰 insight를 얻지 못했습니다. 그래서 301개의 training images를 전부 보면서 어떤 특징들을 가지고 있는지 살펴보았습니다. 구겨진 진료비 영수증 이미지, 배경이 있는 이미지, 이미지를 스캔 또는 인쇄를 하며 생기는 노이즈가 있는 이미지, 가로 방향의 이미지, 흐린 이미지, 어두운 이미지, 글씨들이 도장에 가려져 있는 이미지들을 확인할 수 있었습니다.

Baseline code로 학습된 모델의 결과가 위에서 살펴본 EDA의 결과에 어떻게 반영이 되었는지 살펴보았고 세로로 쓰여진 가로쓰기를 잘 잡지 못하는 문제, 세로쓰기를 검출하는 bbox가 불안정한 문제, bbox를 하나도 못 잡은 이미지가 존재하는 문제와 같이 추가적인 문제들을 발견할 수 있었습니다.

위의 결과들을 바탕으로 해당 프로젝트에서 해결할 문제를 다음과 같이 정리를 할 수 있었습니다.

- ☐ 배경에 bbox가 생기는 문제
- ☐ 글씨들이 도장에 가려져 검출이 안되는 문제
- ☐ 세로로 쓰여진 가로쓰기를 잘 잡지 못하는 문제
- ☐ 세로쓰기를 검출하는 bbox가 불안정한 문제
- ☐ Bbox를 하나도 못 잡는 이미지가 존재하는 문제

4-2. Relabeling



EDA를 통해 확인한 문제를 주어진 training dataset에 대해 다시 labeling을 하고자 했습니다. 그러나 주어진 dataset에 대한 annotation은 UFO(Upstage Format for OCR)로 되어 있지만 해당 annotation을 지원하는 annotation tool인 Upstage annotation tool을 사용하지 못해 외부 annotation tool에 맞는 format으로 주어진 annotation을 변경해야 했습니다. 해당 프로젝트에서는 annotation tool로 CVAT를 사용했고 이전에 사용해본 경험이 있는 COCO format으로 주어진 annotation을 변경했습니다.

4-2-1. Train1 / Val1

처음 제공 받은 100장의 training image로 해당 프로젝트에서는 사용하지 않았습니다.

4-2-2. Train2 / Val2

추가적으로 받은 training images 201장을 더한 총 301개의 dataset입니다. 위에서 확인한 EDA는 해당 dataset으로 실시한 결과입니다.

4-2-3. Train3 / Val3

가이드라인을 새롭게 만들어 해당 가이드라인에 따라 5명의 팀원이 60장씩 나눠서 relabeling을 진행했습니다. 다시 labeling을 진행한 결과 bbox를 하나도 못 잡는 이미지가 있던 문제를 해결했고 도장에 가려져 있는 text의 경우 도장이 매우 진해서 사람도 찾기 힘든 경우를 제외하고는 검출하는 것을 확인했습니다.

그러나 새롭게 발견된 문제들이 있었습니다. QR 코드에 bbox가 생기는 문제, blur 처리된 부분에 bbox가 생기는 문제, bbox가 text에 비해서 크거나 작게 그려지는 문제가 있었으며 bbox의 크기 문제는 bbox가 서로 겹쳐지는 문제로 이어졌습니다.

이렇게 새롭게 발견된 문제들에 대해서는 하나의 가이드라인으로 다시 labeling을 진행했지만 서로 다른 사람이 진행함에 따라 생기는 문제라고 생각해 이후 진행되는 relabeling에서는 이 점을 보완했습니다.

해당 dataset에서 해결된 문제와 해결하지 못한 문제, 새롭게 발견된 문제를 정리하면 다음과 같습니다.

- ☐ 배경에 bbox가 생기는 문제
- ☒ 글씨들이 도장에 가려져 검출이 안되는 문제
- ☐ 세로로 쓰여진 가로쓰기를 잘 잡지 못하는 문제
- ☐ 세로쓰기를 검출하는 bbox가 불안정한 문제
- ☒ Bbox를 하나도 못 잡는 이미지가 존재하는 문제
- ☐ QR 코드에 bbox가 생기는 문제
- ☐ Blur 처리된 부분에 bbox가 생기는 문제
- ☐ Bbox가 text보다 크거나 작게 그려지는 문제

4-2-4. Train4_1 / Val4_1 & Train4_2 / Val4_2

위에서 서로 다른 사람이 진행하면서 문제들을 해결하기 위해 한 명의 팀원이 이미지를 전체적으로 보면서 문제점을 파악하고 글자들의 밀도가 높은 부분에 대해 relabeling을 train3 / val3을 가지고 진행했습니다. QR 코드를 인식하는 문제를 해결하기 위해 text의 밀도가 높은 부분에 대해 relabeling을 진행했습니다. 그러나 이 부분은 inference시에는 blur 처리가 되어 있으므로 또 다른 팀원이 train4_2 / val4_2에서는 해당 부분을 blur 처리하고 blur 처리된 부분에 대해 annotation을 모두 삭제하는 방법도 같이 진행했습니다.

다시 labeling을 진행하면서 발견한 문제로는 대괄호를 감싼 부분에서 다른 기준들이 적용된 것을 확인했고 세로 쓰기 부분에 대해서도 다른 기준들이 적용된 것을 확인했습니다.

완전히 해결되지는 않았지만 전체적으로 bbox가 text에 비해서 크거나 작게 그려지는 문제, QR 코드에 bbox가 생기는 문제, Blur 처리된 부분에 bbox가 생기는 문제를 어느정도 줄일 수 있었습니다.

그러나 하나로 잡아야 할 text를 여러 bbox로 잡으면서(one-to-many match) 평가에 penalty가 적용됨에 따라 성능이 저하되었습니다.

해당 dataset에서 해결된 문제와 해결하지 못한 문제, 새롭게 발견된 문제를 정리하면 다음과 같습니다.

- ☐ 배경에 bbox가 생기는 문제
- 글씨들이 도장에 가려져 검출이 안되는 문제
- ☐ 세로로 쓰여진 가로쓰기를 잘 잡지 못하는 문제
- ☐ 세로쓰기를 검출하는 bbox가 불안정한 문제
- Bbox를 하나도 못 잡는 이미지가 존재하는 문제
- QR 코드에 bbox가 생기는 문제
- Blur 처리된 부분에 bbox가 생기는 문제
- Bbox가 text보다 크거나 작게 그려지는 문제
- ☐ One-to-many match

4-2-5. Train5 / Val5

대괄호, 세로쓰기에 대해 annotation 기준이 다르게 적용된 문제를 두 명의 팀원이 각각 대괄호, 세로쓰기를 맡아 다시 labeling을 진행했습니다.

해당 labeling을 진행한 후 세로쓰기를 검출하는 bbox가 불안정한 문제는 어느정도 해결된 것으로 확인했습니다.

해당 dataset에서 해결된 문제와 해결하지 못한 문제, 새롭게 발견된 문제를 정리하면 다음과 같습니다.

- ☐ 배경에 bbox가 생기는 문제
- 글씨들이 도장에 가려져 검출이 안되는 문제
- ☐ 세로로 쓰여진 가로쓰기를 잘 잡지 못하는 문제
- 세로쓰기를 검출하는 bbox가 불안정한 문제
- Bbox를 하나도 못 잡는 이미지가 존재하는 문제
- QR 코드에 bbox가 생기는 문제
- Blur 처리된 부분에 bbox가 생기는 문제
- Bbox가 text보다 크거나 작게 그려지는 문제
- ☐ One-to-many match

4-2-6. Train6 / Val6

배경에 bbox가 생기는 문제를 해결하기 위해 배경이 존재하는 이미지들만 모아서 fine tuning을 진행했습니다. 그 결과 배경에 bbox가 생기는 문제를 어느정도 해결할 수 있었습니다.

해당 dataset에서 해결된 문제와 해결하지 못한 문제, 새롭게 발견된 문제를 정리하면 다음과 같습니다.

- 배경에 bbox가 생기는 문제
- 글씨들이 도장에 가려져 검출이 안되는 문제
- ☐ 세로로 쓰여진 가로쓰기를 잘 잡지 못하는 문제
- 세로쓰기를 검출하는 bbox가 불안정한 문제
- Bbox를 하나도 못 잡는 이미지가 존재하는 문제
- QR 코드에 bbox가 생기는 문제
- Blur 처리된 부분에 bbox가 생기는 문제
- Bbox가 text보다 크거나 작게 그려지는 문제
- ☐ One-to-many match

4-3. Ensemble

위에서 아직 해결 못한 문제를 해결하기 위해 WBF ensemble을 사용했습니다. 기존 WBF ensemble을 각 모델에 가중치를 주어 평균을 내는 방식을 사용하는데 해당 방법으로 평가를 진행했더니 단일 모델보다 더 좋지 않은 성능을 보여 WBF ensemble을 변형하여 사용했습니다. 하나의 모델만이 예측한 경우 잘못 예측한 것이라고 볼 수 있으므로 해당 bbox는 삭제하도록 변경했습니다.

이렇게 진행했더니 기존에 어느정도 해결했던 문제들을 조금 더 해결할 수 있었고 ensemble의 특징에 맞게 bbox가 이전보다 더 정교해지는 결과를 얻었습니다. 이에 따라 세로로 쓰여진 가로쓰기를 잘 잡지 못하는 문제, one-to-many match 문제를 해결할 수 있었고 최종적으로 해결하려고 했던 문제들을 완벽히는 아니지만 어느정도 해결했습니다.

- 배경에 bbox가 생기는 문제
- 글씨들이 도장에 가려져 검출이 안되는 문제
- 세로로 쓰여진 가로쓰기를 잘 잡지 못하는 문제
- 세로쓰기를 검출하는 bbox가 불안정한 문제
- Bbox를 하나도 못 잡는 이미지가 존재하는 문제
- QR 코드에 bbox가 생기는 문제
- Blur 처리된 부분에 bbox가 생기는 문제
- Bbox가 text보다 크거나 작게 그려지는 문제
- One-to-many match