

2020 인공지능 그랜드 챌린지 참가기

이스트소프트 A.I. PLUS Lab은 과학기술정보통신부에서 주최하는 <2020 인공지능 그랜드 챌린지>에 참가했습니다. 비전 파트는 응급환자 신속대응 기술 개발을 목표로 하는 행동인지 트랙에 참여하여 1위를 했습니다. 이번 글에서는 대회 참가기를 소개하고자 합니다.

1. 인공지능 그랜드 챌린지 소개

인공지능 그랜드 챌린지는 기존의 연구계획서로 수행기관을 선정하는 전통적인 연구개발 선정방식 대신, 정부는 도전문제만 제시하고, 연구팀(혹은 개인)이 자율적으로 해결방안을 제시하여, 우수한 해결책을 제시하는 팀을 선발하여 후속 R&D를 지원하는 방식입니다. 국내 뿐만 아니라 해외 연구자들도 참여 가능하고 상위 팀에게 주어지는 후속 연구비 지원 총액도 120억원에 달하는 국내 최대 규모의 인공지능 대회입니다. 올해는 복합재난상황 이해 및 대응 기술 개발을 목표로 하는 3차 대회와 지역사회 생활문제 대응 기술개발을 목표로 하는 4차 대회가 동시에 개최 되었습니다. 주어진 기간동안 온라인으로 다양한 솔루션을 제출하여 실력을 겨루는 방식으로 진행되었습니다.

챌린지에 참가하여 문제를 풀어내는 과정과 회사에서 AI 프로젝트를 진행하는 과정은 매우 유사합니다. 그랜드 챌린지에는 업체 뿐만 아니라 학교 연구실과 소규모 스타트업도 있었는데요 아직 업체에서 규모있는 AI 프로젝트를 수행한 경험이 없는 분들에게 도움이 될 수 있도록 대회 참가 과정이 실제 규모 있게 진행되는 AI 프로젝트 수행과 어떤 유사점이 있는지도 알려 드리겠습니다.

2. 대회 준비

2.1 문제 파악


행동인지 트랙은 주어진 동영상 내에서 갑자기 실신하는 사람을 검출하고, 해당하는 사람의 위치를 제시하는 문제가 출제되었습니다. 대회 규격에 대한 상세 내용은 <그림1>에 요약되어 있습니다.

[임무] 동영상 내에서 도움이 필요한 사람 찾기


목표 주어진 동영상에서 인공지능 기술을 이용하여 실신하는 사람을 찾고, 해당인물에 바운딩 박스(좌표)를 도출하는 모델을 제시하라

예시


예시 1




예시 2



예시 1



예시 2



데이터 셋 가이드 라인

<행동정의>
병이나 충격으로 의식을 잃고 쓰러져 한동안 일어나지 못하는 것

<이벤트의 세부동작>
쓰러짐 : 걷다가 혹은 서 있다가 등 일반적인 행동 중 갑자기 바닥으로 쓰러져 미동없이 누워있는 행위

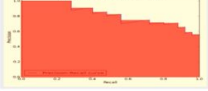
<태깅기준>
- 바닥에 쓰러져서 10초 이상 유지하는 현상을 쓰러짐이라 지칭
- 바닥에 완전히 누운 시점부터로 함
- 쓰러졌다가 일어나는 경우 누워있는 시점까지를 태깅함
- 화면 시작부터 쓰러짐의 형태를 보이는 대상도 10초 이상 유지할 경우 쓰러짐으로 정의
- 화단, 벤치 등 사물 위에 누워 있는 행위도 쓰러짐으로 정의

평가기준

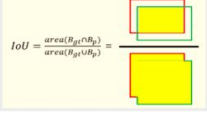
- 제출한 바운딩 박스와 GT를 비교하여 일치비율(I.o.U.)이 0.75 이상인 경우 검출성공
- 평가점수는 mAP(mean Average Precision)

AP(Average Precision)

AP, FP를 나눈 뒤 confidence score의 기준을 바꾸어 Recall, precision을 계산



IoU(Intersection of Union)

$$IoU = \frac{area(R_{gt} \cap R_p)}{area(R_{gt} \cup R_p)}$$


데이터 규격

- FHD(1920×1080px), 15fps, 20초 고정길이, 동영상 500개 클립

시간제한
6시간

사양
V100(VRAM 32GB), 8 Core, RAM 90G

<그림1. 행동 인지 트랙 대회 규격>

챌린지 문제를 소개받는 것은 사업부서나 고객사로부터 문제를 의뢰받는 경우와 유사합니다.

연구소 관점에서는 문제 정의가 명확하지 않은 경우가 많습니다. <그림1>과 같이 대회 규격이 어느 정도 정해져서 와도 대회 규격에 맞는 프로그램을 작성하기 위해서는 매우 엄밀한 정의가 필요합니다. 사업부나 외부 고객사와도 초기에는 회의를 통해 문제 정의와 평가 방법에 대해 많은 논의를 거쳐 상세 기능을 정하게 됩니다. 대회의 경우에도 초기 사업 설명회 이후 참여팀들의 피드백과 질의를 통해 문제 규격이 엄밀해 집니다. 규격을 엄밀히 하는 것의 마지막 단계는 여러 상황을 아우를 평가 데이터 샘플과 그에 대한 라벨, 그리고 평가 코드입니다.

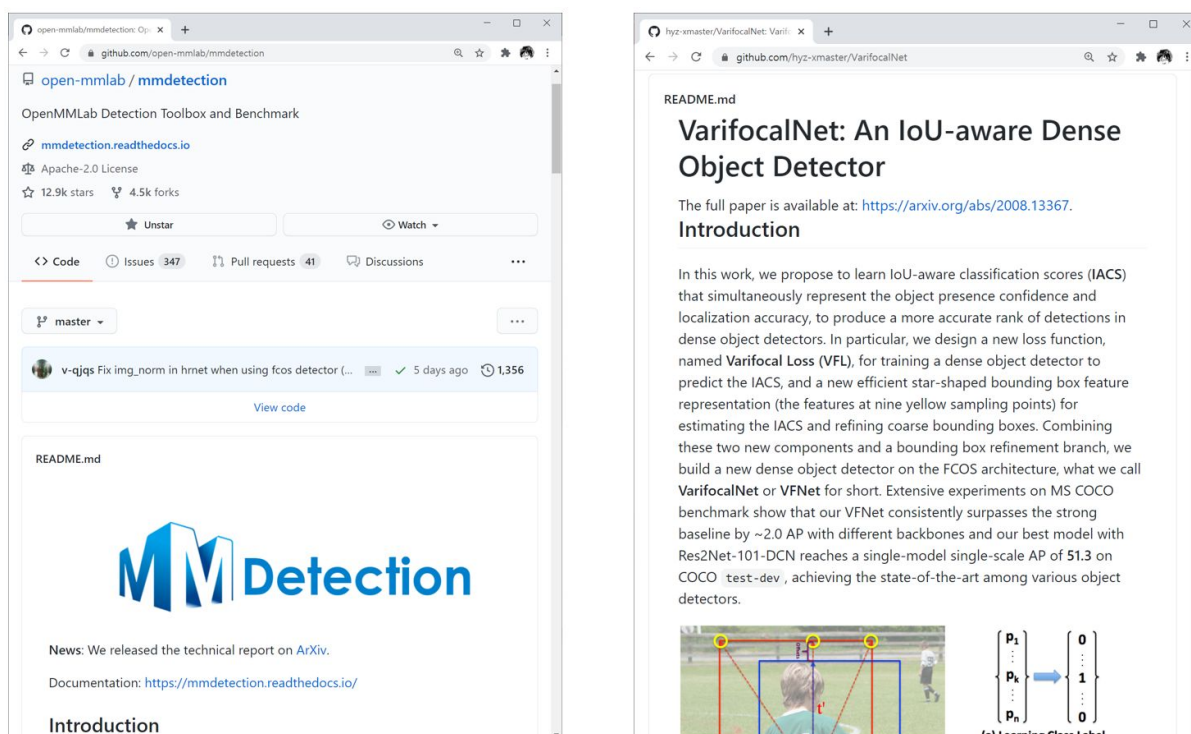
해외 유명 챌린지 플랫폼인 카글에서 진행되는 대회는 대부분 훈련 데이터를 제공합니다. 반면 그랜드 챌린지에서는 일반적으로 훈련 데이터를 제공하지 않습니다. 사업부서나 고객사에서 연구소에 문제를 의뢰할 때도 훈련 데이터를 제공해 주지 않는 경우가 대부분입니다. 특히 신규 서비스 개발인 경우 데이터 확보가 더 어렵습니다. 기업 연구소의 경우 서비스될 환경과 유사한 훈련데이터를 예측하고 수집하는 것도 핵심 역량에 포함됩니다. 훈련 데이터를 제공하지 않는 그랜드 챌린지에서는 특히 이런 역량이 중요합니다. 글로벌 선도 기업이나 학교의 연구소에서는 문제를 정의하고 평가 방식을 고안하고 데이터를 마련하는 일을 많이 합니다. 작금의 기술로 풀기 어려우나 사회적으로 유의미한 문제를 선정하고 참여팀들은 해당 문제를 풀기에 적합한 데이터 셋을 마련하는 그랜드 챌린지의 방향성은 그래서 여러 모로 긍정적인 면이 있다고 생각합니다. 다만 아쉬운 점은 코드화 가능한

엄밀한 평가 방식 공유, 문제를 충분히 설명할 정도의 평가 데이터 샘플 및 라벨 공유가 이루어지지 않고 있는 점입니다. 개선되리라 기대합니다.

2.2 전략 도출

문제 분석 후 공략방법을 도출했습니다. 이 과정은 챌린지나 회사 프로젝트나 동일합니다. 개발 기간, 투입 가능 인력, 투입 가능 비용등 가용 리소스를 먼저 산정합니다. 여러 정황을 고려하여 저희는 비록 이번 챌린지는 동영상 문제이지만 일반 이미지 탐지(object detection)를 기본 공략으로 사용하기로 했습니다. 문제 규격 중 10초 룰(10초간 쓰러짐이 지속되어야 실신으로 인정)외에는 동영상에 특화된 부분이 없다고 판단했습니다. 비전 파트는 시아나운서 프로젝트를 수행하면서 동영상을 어느 정도 다뤄 보기는 했으나 타 업무 진행으로 대회 준비를 짧게 해야 되는 상황을 고려하면 능숙하게 다룰 수 있는 이미지 탐지 기술로 공략하는 것이 더 좋을 것이라 판단했습니다.

코드 베이스는 기존부터 이미지 탐지에 사용하던 mmdetection 툴을 사용하고 모델은 VarifocalNet을 사용하기로 했습니다. 멀티 스테이지 디텍터가 정확도가 뛰어나기 때문에 mmdetection을 선택했고 segmentation 없이 box만 도출하면 되기 때문에 속도도 빠르고 box detection 정확도가 매우 높은 pre-trained 모델을 제공하는 VarifocalNet을 선택했습니다. VarifocalNet은 <그림2>의 왼편에 보이는 github 저장소에서 box map 51.3 이라는 수치를 보여주는 모델을 제공하고 있습니다.

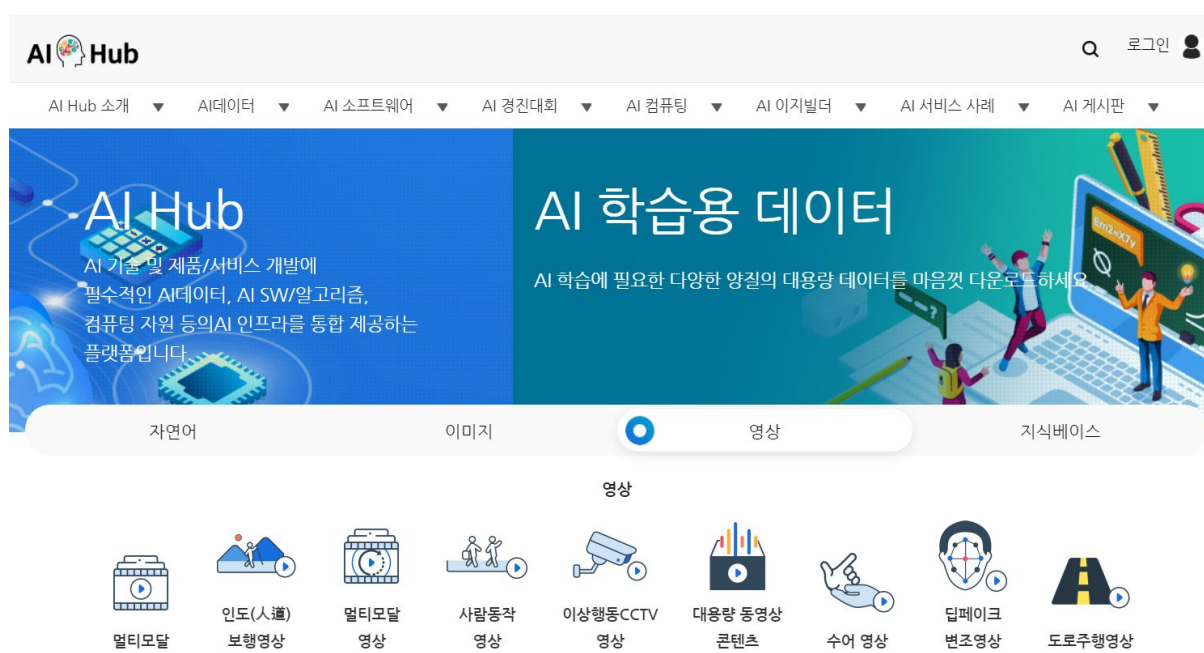


<그림2: 코드 베이스>

2.2 학습 이미지 마련

공략법이 도출되면 먼저 훈련 데이터를 확보해야 합니다. 이 부분도 회사 프로젝트와 챌린지 모두 동일합니다. 데이터 수집 작업을 연구원들이 할 수도 있고 회사 내의 데이터 라벨링 팀과 나누어 하거나 외부 회사에 비용을 지불하고 맡길 수도 있습니다. 이번 챌린지의 경우 쓰러진 사람 이미지는 대량으로 수집하기 어렵고 촬영 비용도 부담하기 어려워서 연구소에서 직접 하였습니다.

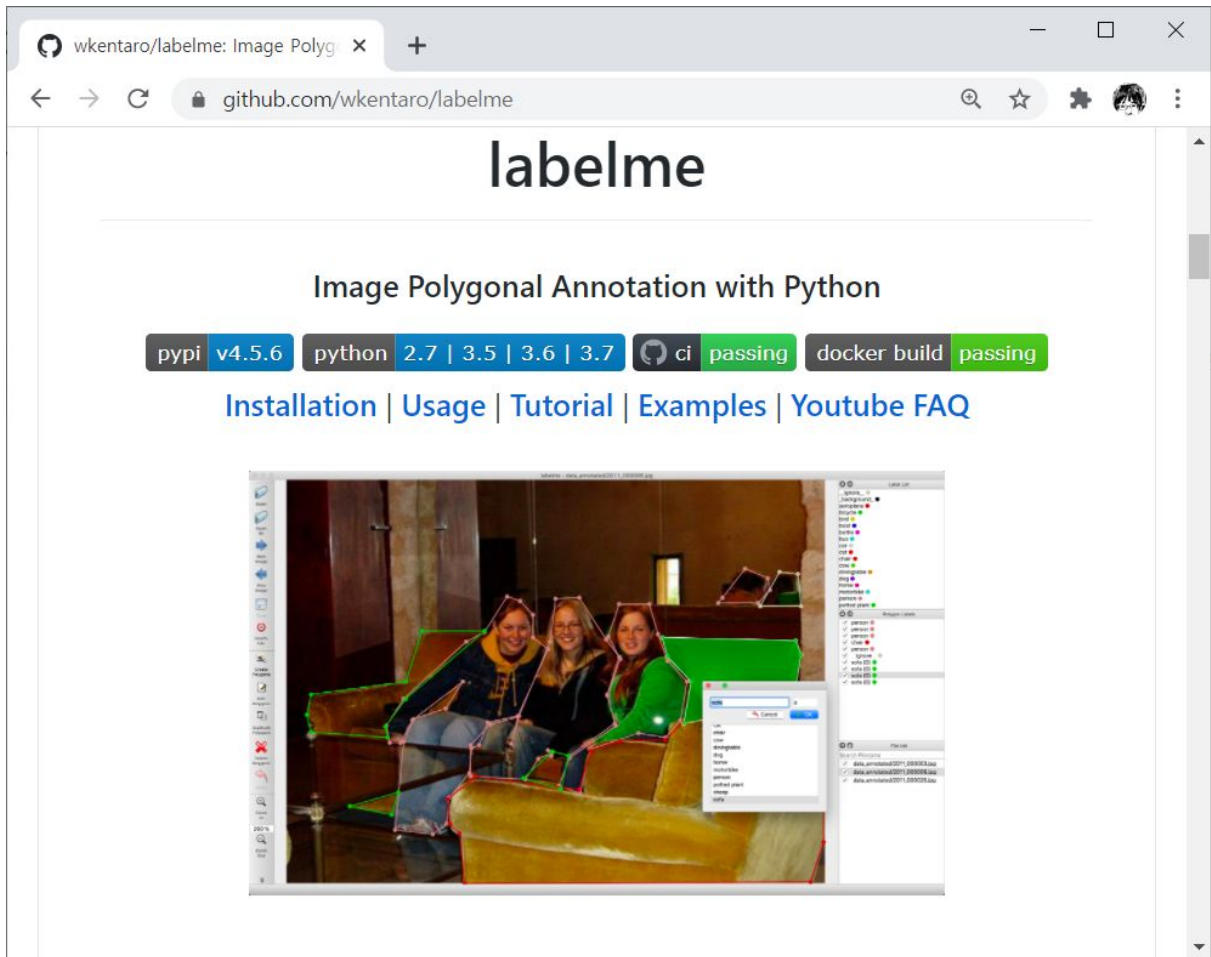
한국정보화진흥원에서 운영하는 AI 통합 플랫폼 AI Hub(<https://www.aihub.or.kr/>)에 있는 데이터를 기본으로 사용했습니다. AI Hub에는 <그림3>과 같이 ‘사람동작 영상’, ‘이상행동 CCTV 영상’, ‘공항 이상행동 CCTV 영상’ 이 있습니다. 동영상 데이터이기 때문에 영상 데이터와 함께 제공되는 라벨 파일을 분석하여 쓰러짐 부분만 추출하는 작업이 추가로 필요했습니다. validation 데이터로는 대회 주최측에서 제공한 샘플 데이터와 직접 촬영한 데이터를 사용했습니다.



<그림3. 학습 데이터에 사용할 이미지를 제공받은 AI Hub>

2.3 학습 이미지 라벨링

라벨링 툴은 labelme라는 오픈소스 툴을 사용했습니다. 툴을 이용한 라벨링 작업은 회사 내의 데이터 라벨링 팀에 의뢰를 하였습니다.



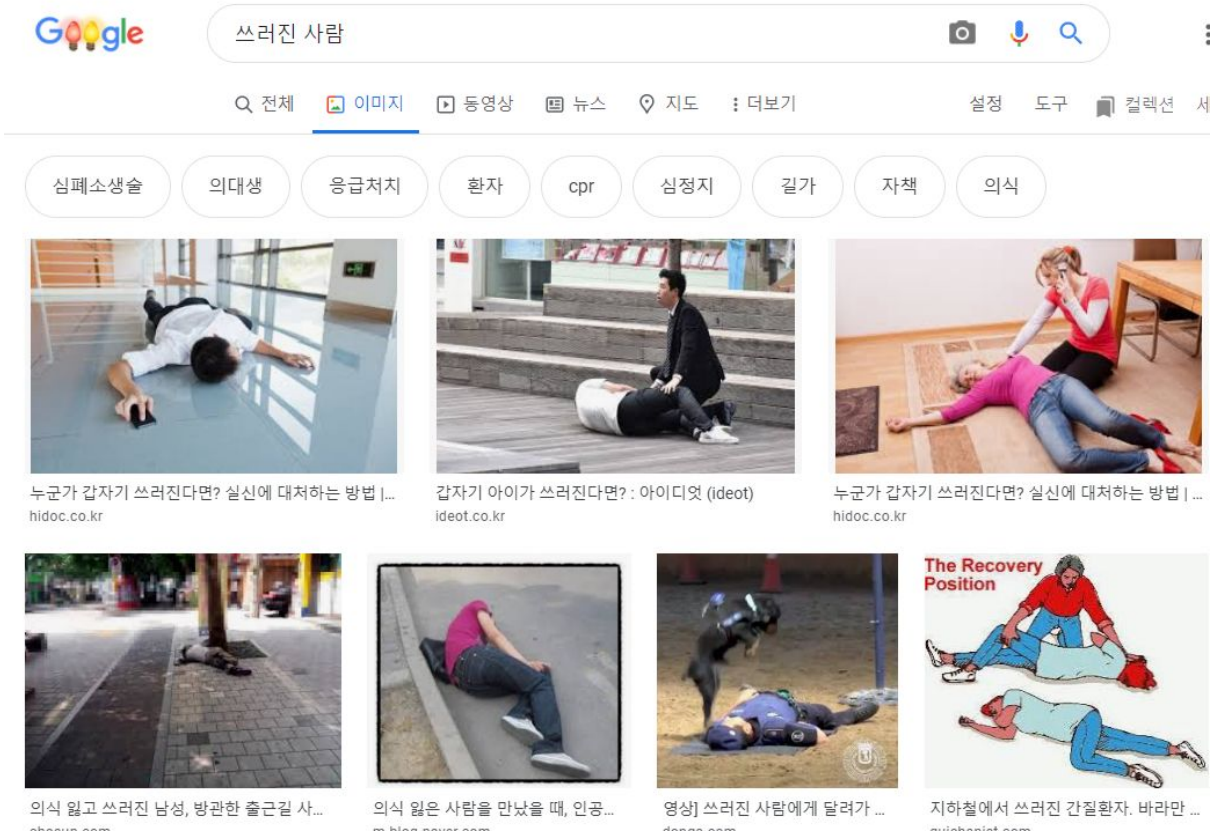
<그림4. 이미지 라벨링 툴 labelme의 github 저장소>

2.4 학습 및 초기 모델 결과 검토

학습된 모델의 결과를 보니 기어가는 강아지나 앉아 있는 사람도 쓰러졌다고 판단 하는 경우가 많았습니다. AI Hub 데이터는 모두 연출된 화면이기 때문에 다양한 장면이 없었습니다. 학습 때 보지 못한 기어가는 개나 앉아 있는 사람이 입력되었을 때는 학습 때 봤던 서 있는 사람과 쓰러진 사람 사람중에서는 쓰러진 사람에 더 가깝다고 판단하고 있었습니다. 실제 프로젝트를 진행할 때도 모델의 정확도가 낮은 이유가 모델의 아키텍처나 훈련 기법이 고도화 되지 않아서라기보다 학습 데이터 자체에 문제가 있는 경우가 훨씬 많습니다. 모델 연구나 최적화 연구등 단위 연구가 아니라 AI 기술을 종합해서 실제 문제를 풀어야 하는 상황에서는 이와 같이 학습 데이터 분포를 실제 서비스 환경과 맞추는 노력과 기술이 무엇보다 중요하다고 볼 수 있습니다.

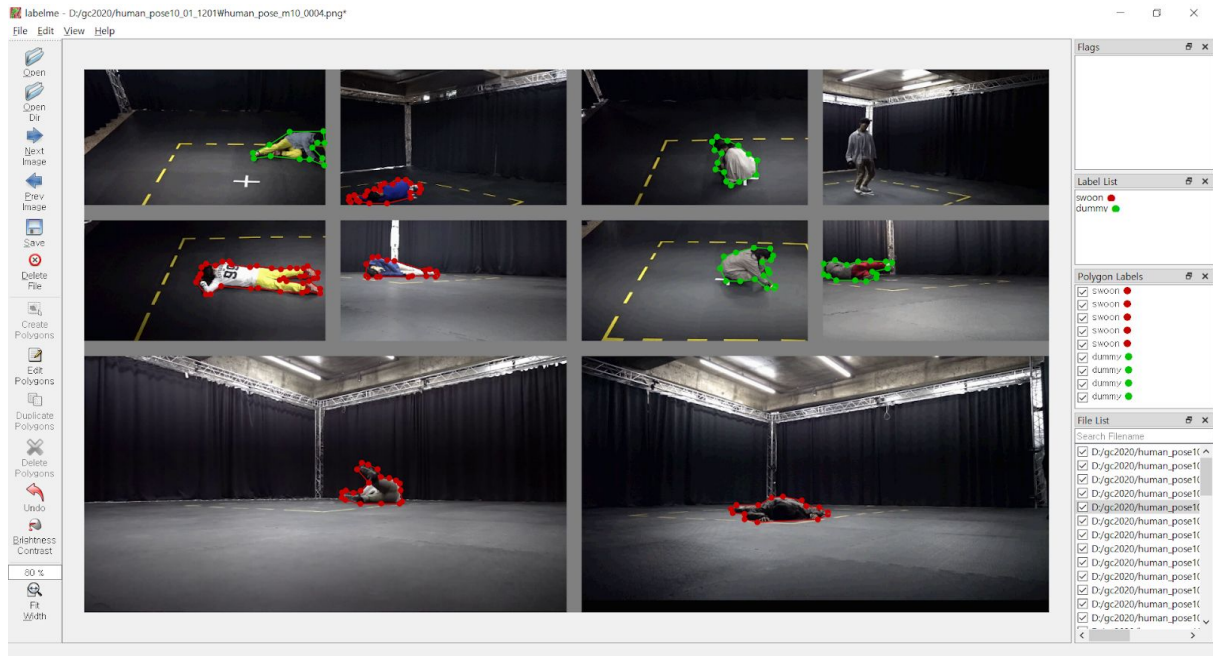
2.5 학습데이터 보강

두 가지 처방을 하였습니다. 첫째 처방은 다양한 화면을 보여주기 위해 <그림5>와 같이 구글 이미지와 유튜브 동영상에서 캡처한 이미지를 추가하는 것이었습니다.



<그림5. 쓰러진 사람 이미지 검색>

두 번째 처방은 쓰러지기 전 상황도 탐지하도록 하는 것이었습니다. 쓰러지기 전 상황도 별도의 라벨을 부여하여 학습시키기로 했습니다. <그림 6> 에서 붉은 색 라벨이 실제 대회에서 요구하는 쓰러짐에 대한 라벨이고 녹색 라벨이 정확도 향상을 위해 추가로 학습시키는 dummy 라벨입니다. 그리고 <그림6>에서 볼 수 있듯이 학습 속도 향상을 꾀하기 위해 AI Hub에서 다운받은 이미지를 크기를 다양하게 하여 한개의 이미지로 병합하여 학습 데이터로 사용하였습니다. 대회 규격은 사각형 박스만 제출하는 것이지만 정확도를 떨어뜨리지 않는 회전 오그멘테이션 등이 가능하도록 폴리곤 라벨링을 하였습니다.



<그림6. labelme를 사용한 라벨링. 쓰러짐 라벨과 쓰러지기 전 동작 모두 라벨링>

3. 챌린지 진행

3.1 제출 및 점수 확인

챌린지 기간동안 솔루션을 여러번 제출하고 점수판을 통해 점수를 확인할 수 있습니다. 이 과정은 프로젝트시 연구소 내부 테스트를 끝내고 모듈을 실제 서비스에 반영하는 과정에 대응됩니다. 어떤 챌린지는 참가자들이 실제 평가 데이터를 받아보고 정답만 작성해서 제출하는 경우도 있고 이번 그랜드 챌린지 처럼 평가 데이터는 볼 수 없고 솔루션만 제공 후 최종 메트릭 값만 볼 수 있는 챌린지가 있습니다. 실제 프로젝트 수행 시에도 이와 유사합니다. 서비스 반영 시 개별 데이터를 연구소에서는 직접 볼 수 없고 메트릭 또는 사용자 피드백만 받는 경우가 있습니다. 외부 고객사에 제품을 납품해야 되는 경우 개별 데이터를 쉽게 볼 수 없는 경우가 종종 있습니다.

이번 챌린지는 많은 참여팀들이 초반에 평가 플랫폼 환경에서 코드가 정상 동작하지 않아 애를 먹었습니다. 실제 프로젝트에서도 외부 고객사에 납품을 하던가, 개발팀이 직접 접근이 불가능한 서비스 서버에 코드를 배포하는 경우 흔히 발생하게 됩니다. 이 경우는 보통 서비스 오픈 직전이거나 서비스 오픈 직후기 때문에 운영팀 분들의 피드백과 에러 로그를 전달받아 정상 동작하는 코드로 빠르게 수정해야 됩니다.

대회 평가 데이터셋은 15fps인데 첫번째 제출 버전은 빠른 결과를 얻기 위해 5fps로 추론하고 나머지 사진들은 이전에 추론한 값과 동일하게 답을 내는 버전을 제출했습니다. 제출을 해서 점수를 받아 보니 평가 속도인 map값이 0.75 근처가 나왔습니다. 다른 참여팀들과 점수 차이가 많이 났기 때문에

추가적인 알고리즘을 도입하지 않고 남은 기간 동안 몇 가지 설정만 바꾸어 점수를 올리고 하였습니다.

3.2 점수 올리기

이후 제출에서는 fps를 7.5 로 올리기, iou 값을 기본값인 0.6에서 0.8 로 올리기 등의 설정을 바꾸면서 최고 점수가 나오는 솔루션을 찾아 갔습니다. 실제 프로젝트에서도 이런 설정값 미세 조정을 통해 서비스 퀄리티를 높이는 경우가 많습니다. 서비스를 오픈하고 나면 연구실 환경과 다른 데이터 분포에서 모델이 동작하게 됩니다. 서비스 환경의 데이터 분포를 파악하고 거기에 맞춰 모델을 새로 학습시키는 경우도 있지만 이번 챌린지 대회처럼 서비스 퀄리티 피드백만 가지고 실제 서비스 환경에 적합한 설정값을 찾아 가능 경우도 있습니다. 설정값 변경만으로 map 0.75에서 map 0.799까지 올릴 수 있었습니다.

4. 참가 결과

5번째 제출한 설정값이 map 0.799 값을 기록했고 이 점수로 대회를 1위를 기록하며 대회를 마감하였습니다.



<그림6. 대회 최종 결과>

대회 최종 순위 확정 후 각 트랙별 1위 팀들의 공략법을 공유하는 기술 교류회가 열렸습니다.

https://github.com/kts12345/agc2020_4th_tr1 에서 기술 교류회 발표 자료, 발표 동영상 등 모든 자료를 다운로드 받을 수 있습니다. 발표 자료 외에도 소스코드를 도커 형태로 공유하였기 때문에 누구나 손쉽게 저희 솔루션을 테스트 해 볼 수 있습니다. 특히 4차 1트랙 참여팀들은 각 팀의 validation set에 대해 저희 솔루션을 돌려서 여러 가지 비교를 해 볼 수 있어서 내년도 챌린지 참가에 도움이 될 것으로 기대됩니다.

5. 참가 소감 및 향후 계획

정부에서 AI Hub를 통해 제공하는 풍부한 데이터를 기본 학습데이터로 삼고 기본 학습 데이터에 부족한 부분을 google image와 유튜브 데이터 수집을 통해 보완한 것이 이번 대회에서 좋은 점수를 받는데 큰 역할을 한 것이라고 생각합니다.

이번 챌린지는 대회 시작 전부터 Q&A 게시판이나 메일 등을 통해 대회 규격의 명확화 등에 대해 주최측에 많은 피드백을 주었어야 했고 대회도 중단되었다 재개 되는 등 여러 이슈가 있었습니다. 일단 프로젝트가 시작되면 업무를 발주한 고객사와 솔루션을 개발해야 되는 연구팀이 서비스의 성공적인 오픈이라는 목표를 향해 함께 달려 가듯 개발팀과 주최측이 의미 있는 챌린지가 되도록 서로 어느 정도 희생하고 어려운 결단을 내렸어야 했습니다. 그 과정에서 많은 분들이 속상한 일을 겪어야 하기도 했습니다. 저희 팀은 다행히 여러 운이 따라 주어 큰 난관 없이 챌린지 진행이 가능했는데요 다음 부터는 대회 준비 부분에 더욱 의견을 개진하여 AI 본연의 실력 겨루기에 집중할 수 있는 대회가 되도록 해야 겠다고 생각했습니다.

※ 참고

<http://www.ai-challenge.kr/>

<https://aihub.or.kr/>

<https://github.com/open-mmlab/mmdetection>

<https://github.com/hyz-xmaster/VarifocalNet>

<https://github.com/wkentaro/labelme>