

범죄, 멈춰!

폭행 상황을 감지하는 지능형 CCTV 시스템

Capstone 23

목차

1. 프로젝트 배경
2. 핵심 기능
3. 프로젝트 구조와 흐름
4. CCTV(Camera + Single board computer)
5. AI Server
6. Web Server
7. 시연영상
8. 보조 자료

1. 프로젝트 배경

- **실시간 대응 미흡** : 대부분의 CCTV는 영상을 녹화하여 추후 증거로만 활용되는 한계가 있습니다.
- **CCTV 관제 인력 부족** : 행정안전부 기준, 관제 인력 1명 당 적정 모니터링 대수는 **50**대이지만 실제로는 1명 당 약 **100**대 입니다.¹⁾
5년간 CCTV 관제 대수는 **90%** 증가한 반면 관제요원 인원은 **48%** 증가에 그쳤습니다.²⁾

1) CCTV 통합관제센터 운영실태 및 개선방안(입법·정책보고서 Vol. 29, 2019)

2) 행정안전부, 「지자체 영상정보처리기기 통합관제센터 구축 및 운영 규정」

1. 프로젝트 배경

- **육안 관제의 한계** : 연속적으로 영상을 **22분** 이상 쳐다보면 현장의 움직임을 **95%**까지 빠뜨린다는 연구 결과가 있습니다.³⁾
- **영상 용량으로 인한 보관 기간 한계** : 근래의 고화질 CCTV는 고용량이라 오래 보관하지 못합니다. **폭행 상황**이 감지된 구간만을 따로 저장한다면 보관 기간으로 인한 증거 확보의 어려움을 개선 할 수 있습니다.

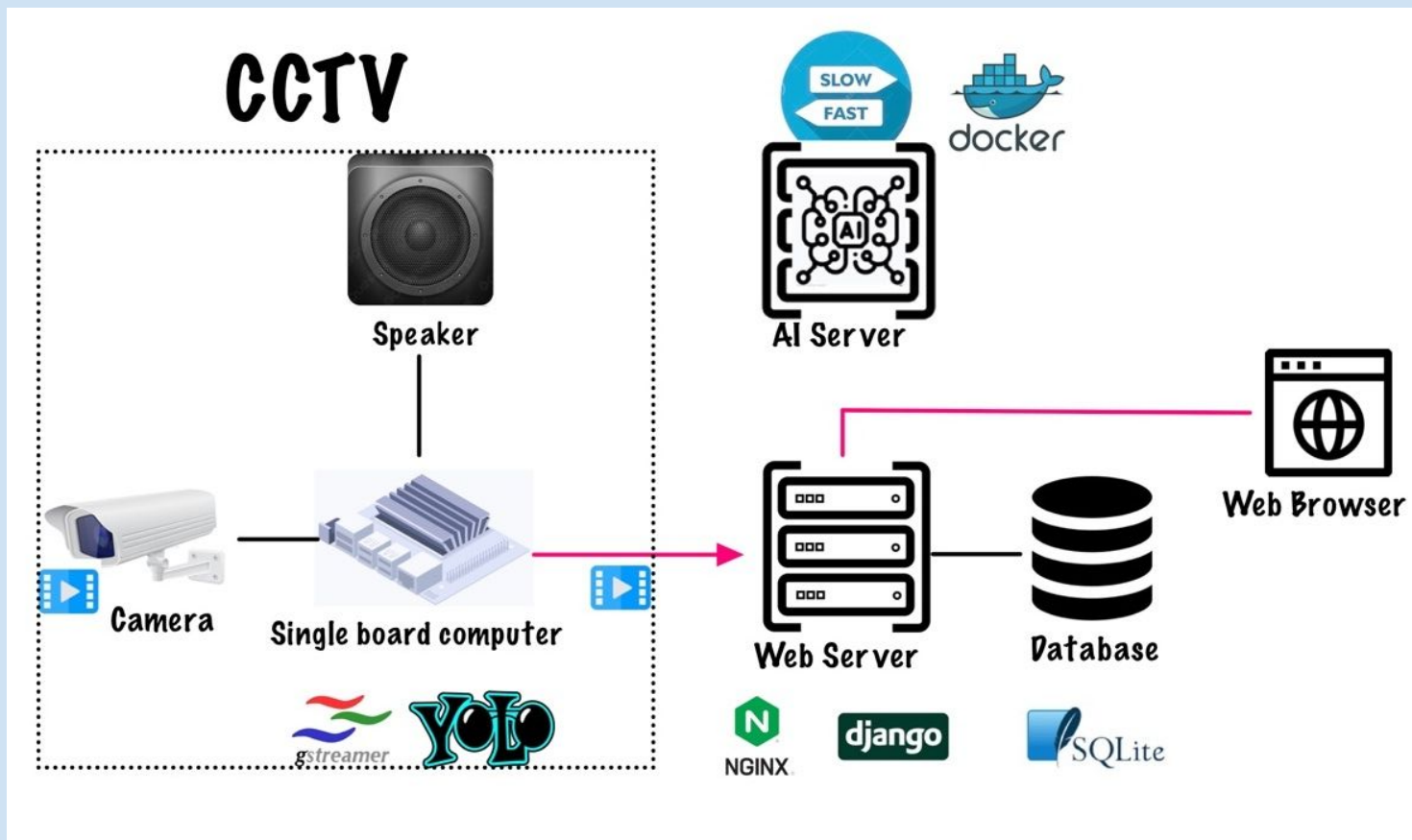
3) T. Ainsworth, Buyer Beware, In Security Oz , vol. 19, pp. 18-26, (2002).

2. 핵심 기능

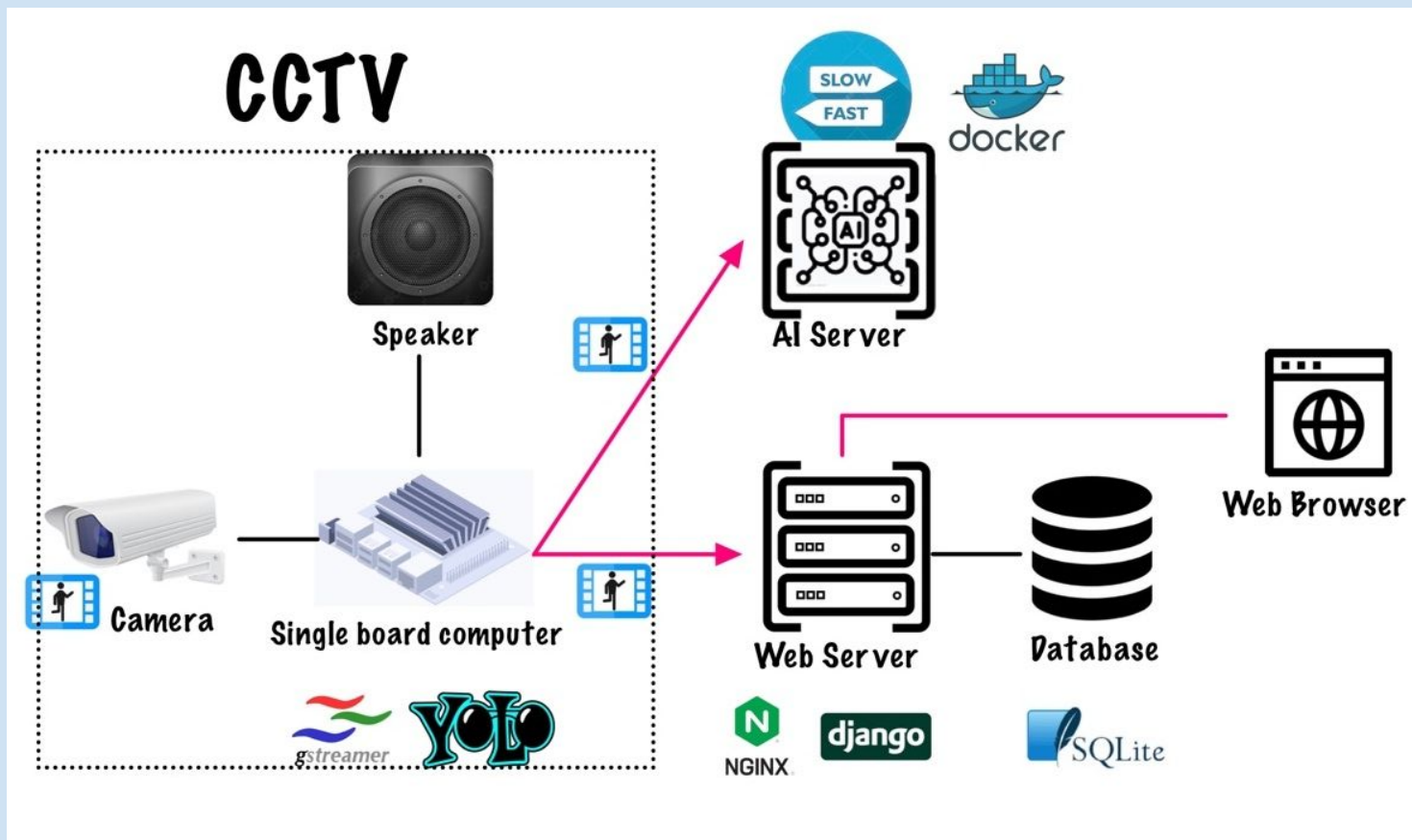
- CCTV에 탑재된 on-device AI가 사람을 검출하면 해당 영상을 AI Server로 전송합니다.
- AI Server가 폭행 상황을 감지하면 현장의 CCTV에서 “범죄 멈춰” 음성을 출력하고, CCTV 관리자(사용자)에게 폭행 장면의 대표 이미지와 촬영 시각, 장소를 메일로 전송합니다.
- 사용자는 언제든지 실시간 CCTV 영상을 확인 할 수 있고, 폭행 상황이 검출된 구간의 영상을 열람할 수 있습니다.

3. 프로젝트 구조와 흐름

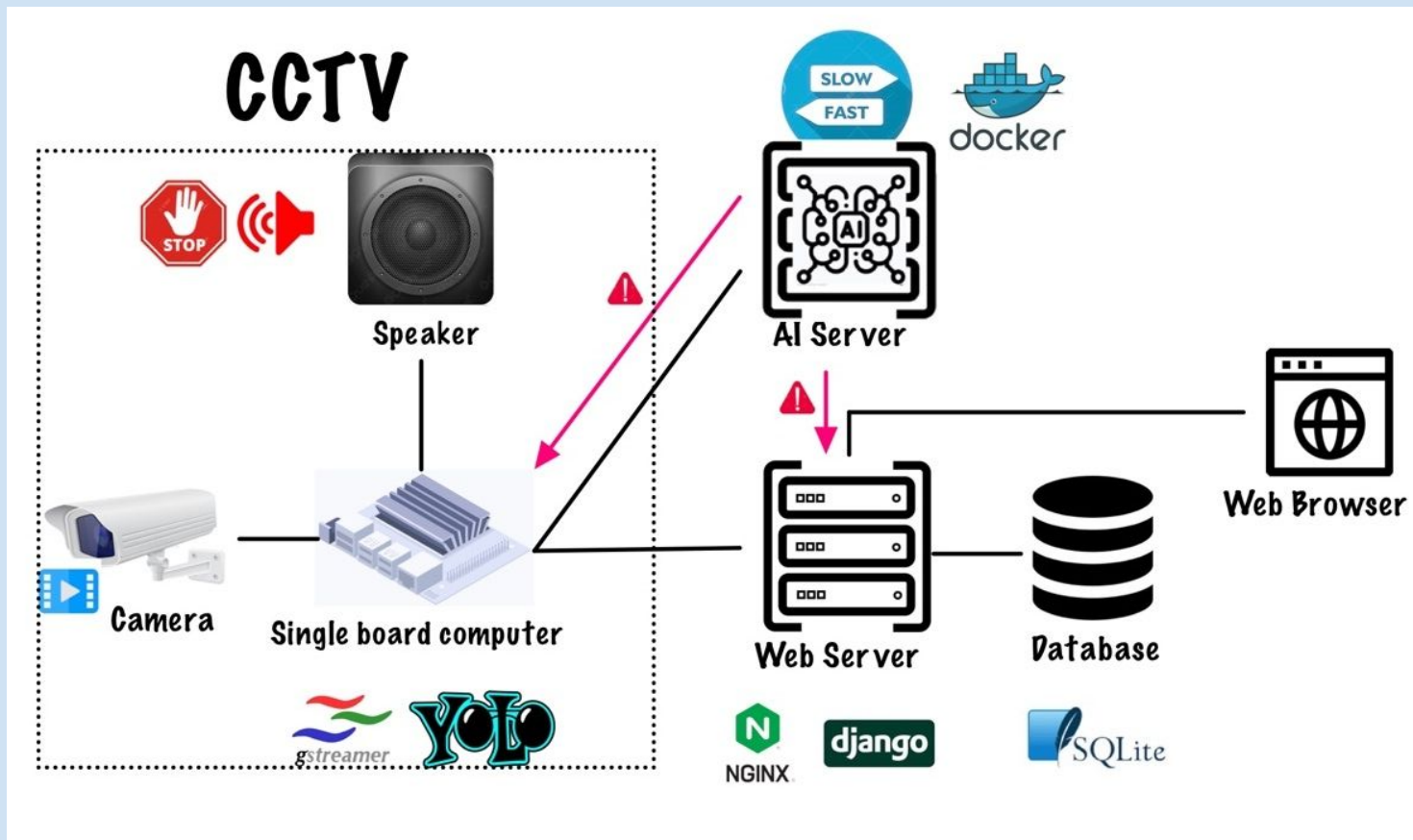
3. 프로젝트 구조와 흐름



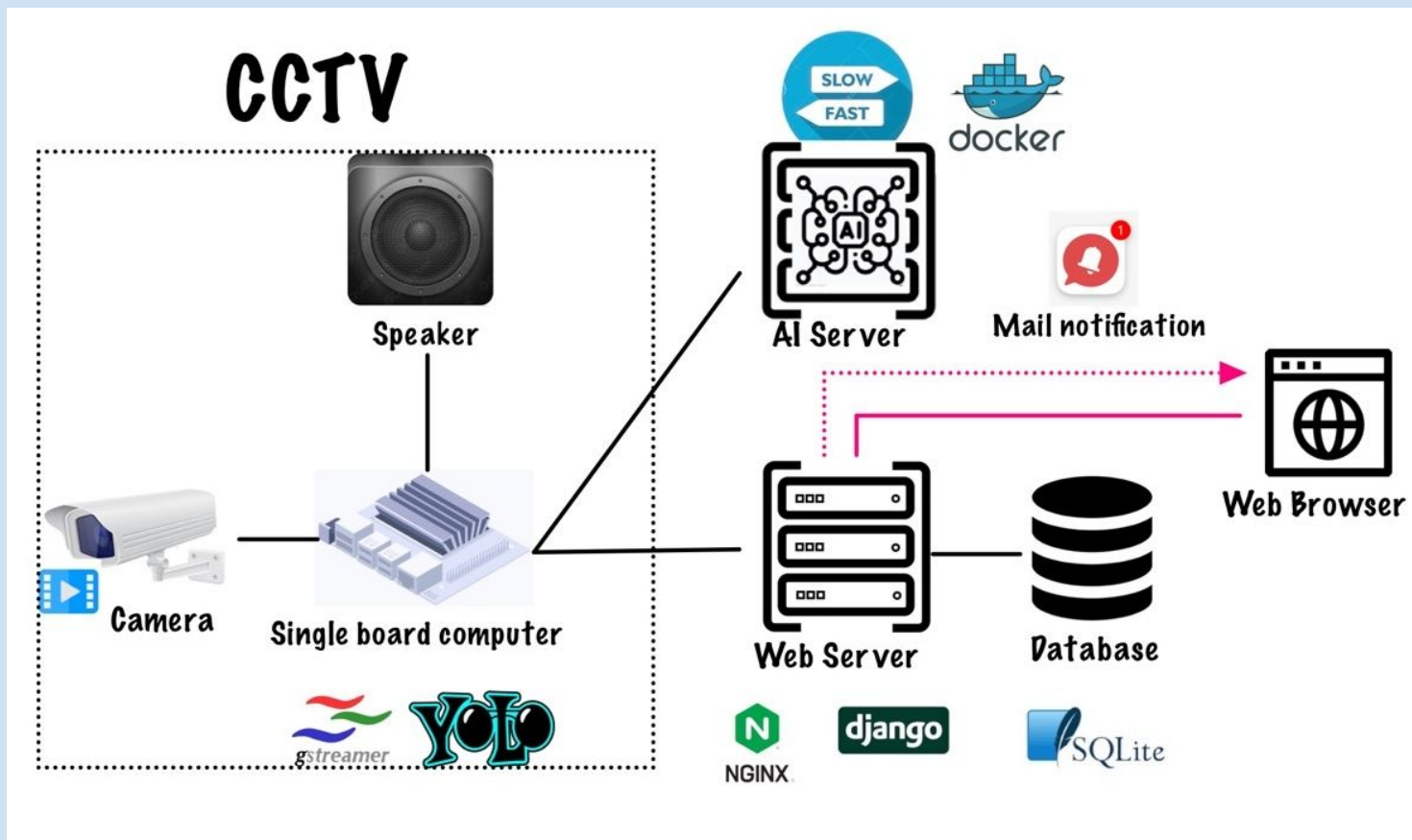
3. 프로젝트 구조와 흐름



3. 프로젝트 구조와 흐름



3. 프로젝트 구조와 흐름



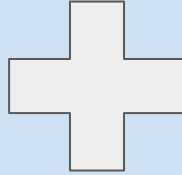
4. CCTV

(Camera + Single board computer)

4. CCTV(Camera + Single board computer)



Jetson Nano Developer Kit⁴⁾



Logitech C920 HD webcam⁵⁾

4) <https://developer.nvidia.com/embedded/jetson-nano-developer-kit>

5)

<https://www.logitech.com/en-us/products/webcams/c920-pro-hd-webcam.960-000764.html#specs>

[CCTV] On-device AI

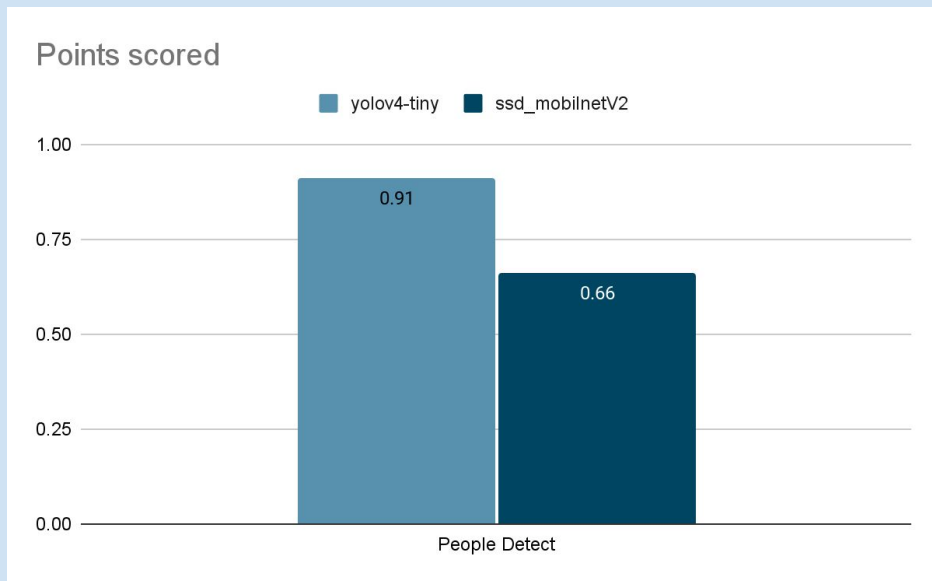
- AI Server의 폭행 상황 검출 정확도를 높이고, Computing resource를 절약하기 위해 사람이 검출 될 때만 AI Server로 영상을 전송합니다.
- Jetson Nano에서 사람을 검출하기 위해 두가지 모델을 선정하여 테스트했습니다.
 - SSD mobilenet v2⁶⁾
 - YOLOv4-tiny⁷⁾

6) [MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications](#)

7) [YOLOv4: Optimal Speed and Accuracy of Object Detection](#)

[CCTV] Object Detection 모델 비교

자체 제작한 Test set⁸⁾에서 사람 검출 정확도⁹⁾를 평균낸 결과는 다음과 같습니다.



YOLOv4-tiny 모델의 검출 정확도가
ssd_mobilenetV2 모델 보다 약 **25%** 높아
YOLOv4-tiny 모델을 선택했습니다.

8) p.12 AI Server 파트에서 설명

9)_top 1 accuracy

[CCTV] YOLOv4-tiny 적용



테스트 영상



시연 영상

5. AI Server

5. AI Server 모델 및 데이터 소개

- CCTV로부터 수신한, 사람이 검출된 영상을 바탕으로 폭행 상황여부를 판단합니다.
- 행동 인식에 특화된 인공지능 모델(**SlowFast**¹⁰⁾)을 사용합니다.

이 모델은 배경 정보와 행동 정보를 검출하는 두개의 네트워크에서 계산된 정보를 종합하여 영상 속 행동을 판단합니다.

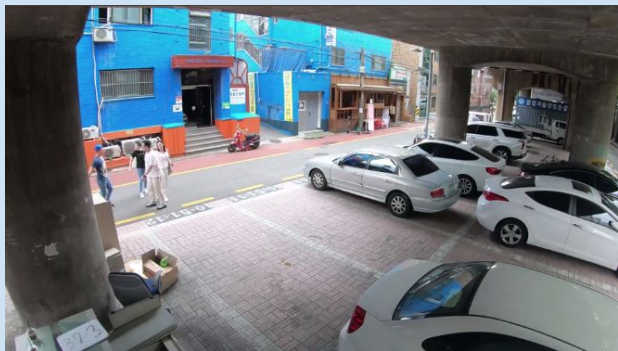
- AI Hub에서 제공하는 공공 데이터셋 중 **이상 행동 CCTV 영상 AI데이터**¹¹⁾로 학습을 진행했습니다.

10) [SlowFast Networks for Video Recognition\(ICCV, 2019\)](#)

11) [이상행동 CCTV 영상 AI데이터](#)

[AI Server] 이상치 데이터 제거

- 공공 데이터 셋에서 5000개로 이루어진 폭행 영상을 확인해본 결과 폭행이라고 할 수 없는 영상들이 폭행으로 오분류되어 있었습니다.
- 모델의 정확도 향상을 위해서는 이러한 이상치 데이터를 제거 해야 했습니다.



<예시 1>

폭행 장면이 없는데 폭행 상황으로 분류되어 있음



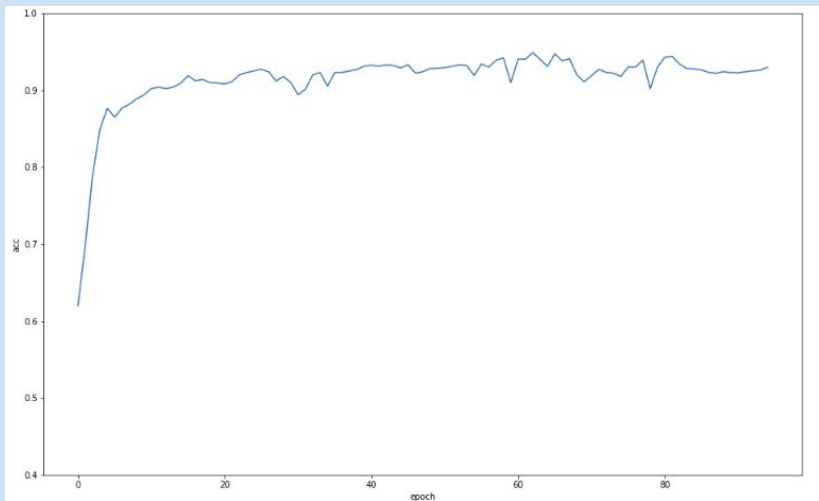
<예시 2>

손을 내미는 영상인데 폭행 상황으로 분류되어 있음

[AI Server] 이상치 데이터 제거 과정

- 5000개가 넘는 폭행 영상을 직접 확인하여 이상치 데이터를 제거하는 것은 불가능하다고 판단하여 아래의 방법을 사용했습니다.
- training set에서 확실한 폭행 영상을 500개 추출해 **정제된 미니 데이터 셋**을 만들었습니다.
- 정제된 미니 데이터 셋으로 **이상치 데이터 검증용 모델**을 만들었습니다.
 - 1) 해당 모델을 사용하여 **training set**의 모든 폭행 영상을 검사해 결과값을 얻습니다.
 - 2) 결과값이 0에 가까울수록 폭행영상을 정상 상황으로 오판, 1에 가까울수록 잘 판단한 것입니다.
 - 3) **결과값 하위 10%**의 영상들을 이상치 데이터라고 판단하여 **training set**에서 제거했습니다.

[AI Server] 이상치 데이터 제거 후 학습



Validation Set 정확도

- Validation Set에 대한 정확도가 **최대 94.6%**입니다.
(이상치 데이터 제거 전 90%, 동일 Validation Set)
- “이상 행동 CCTV 영상 AI데이터”와는 다른 특성을 가지는 데이터에도 좋은 결과를 보이는지 확인하기 위해 자체적으로 **Test set**을 구축했습니다.

[AI Server] 직접 만든 test set

normal 영상 약 200개



abnormal 영상 약 100개



Normal 영상은 폭행 상황이 아니지만 육체적 활동이 많은 영상으로 구성했습니다.
(ex. 배드민턴 치기, 말뚝박기, 하이파이브, 어깨동무 등등)

[AI Server] Test set으로 정확도¹¹⁾ 평가 결과

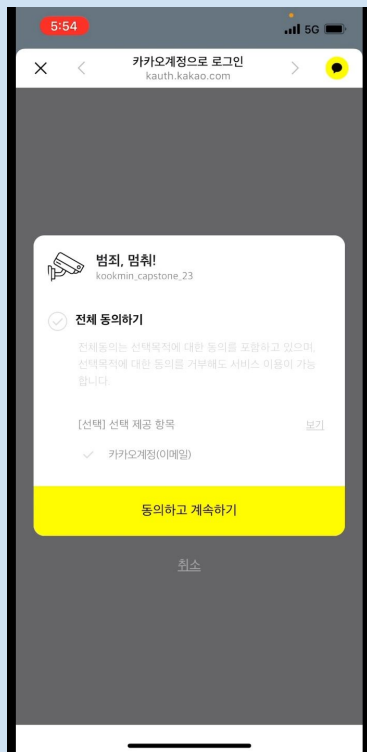
- 이상치 데이터를 제거한 모델들의 성능이 **test set**에 대해서도 전반적으로 좋은 성능을 보였습니다.
- **Test set**을 기준으로 가장 좋은 성능을 가지는 모델은 이상치 데이터를 제거한 데이터 셋으로 **40 epoch** 학습시킨 모델이었습니다.

| Preprocessing | Test Data | 5 Epoch | 10 Epoch | 15 Epoch | 20 Epoch | 25 Epoch | 30 Epoch | 35 Epoch | 40 Epoch | 45 Epoch | 50 Epoch | 55 Epoch | 60 Epoch | 65 Epoch | 70 Epoch | 75 Epoch | 80 Epoch | 85 Epoch | 90 Epoch | 95 Epoch |
|---------------|-----------|---------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| BEFORE | Normal | 0.980 | 0.946 | 0.926 | 0.985 | 1.000 | 0.995 | 0.980 | 0.985 | 0.975 | 0.961 | 0.985 | 0.912 | 0.956 | 0.956 | 0.966 | 0.975 | 0.990 | 1.000 | 0.961 |
| | Abnormal | 0.825 | 0.893 | 0.932 | 0.845 | 0.699 | 0.680 | 0.845 | 0.786 | 0.893 | 0.854 | 0.864 | 0.893 | 0.854 | 0.883 | 0.825 | 0.825 | 0.602 | 0.583 | 0.835 |
| | Total | 0.928 | 0.928 | 0.928 | 0.938 | 0.899 | 0.889 | 0.935 | 0.919 | 0.948 | 0.925 | 0.945 | 0.906 | 0.922 | 0.932 | 0.919 | 0.925 | 0.860 | 0.860 | 0.919 |
| AFTER | Normal | 0.956 | 0.971 | 0.946 | 0.985 | 0.961 | 0.971 | 0.966 | 0.985 | 0.975 | 0.922 | 0.961 | 0.985 | 0.990 | 0.995 | 0.951 | 0.951 | 0.956 | 0.946 | 0.931 |
| | Abnormal | 0.874 | 0.883 | 0.932 | 0.864 | 0.883 | 0.738 | 0.922 | 0.922 | 0.913 | 0.903 | 0.883 | 0.864 | 0.825 | 0.845 | 0.932 | 0.864 | 0.932 | 0.883 | 0.913 |
| | Total | 0.928 | 0.941 | 0.941 | 0.945 | 0.935 | 0.893 | 0.951 | 0.964 | 0.954 | 0.915 | 0.935 | 0.945 | 0.935 | 0.945 | 0.945 | 0.922 | 0.948 | 0.925 | 0.925 |

11) [acc by confusion matrix](#)

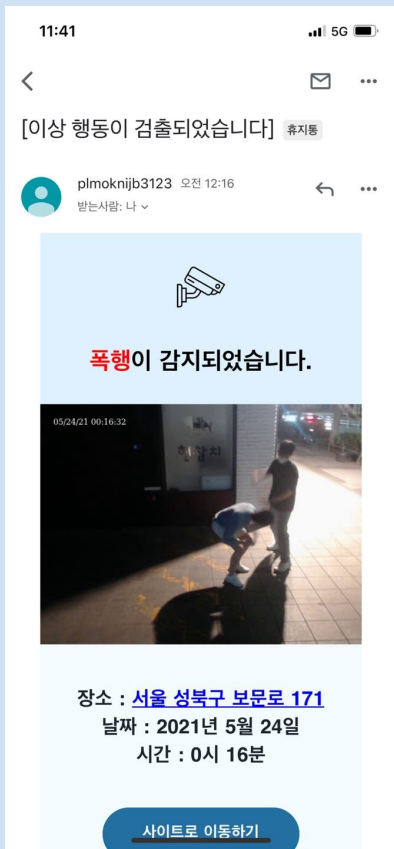
6. Web Server

6. Web Server 1) 로그인/ 회원가입 기능



- 로그인/회원가입 기능 : 카카오 로그인 연동
- 장점 : 카카오톡 사용자라면 별도의 로그인/회원가입 절차를 거치지 않습니다.

[Web Server] 2) 사용자 알림 기능

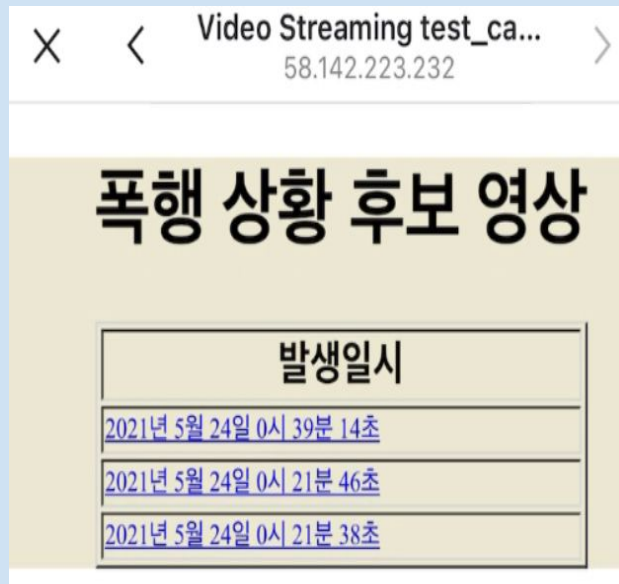


- 폭행 상황이 감지되면 사용자에게 자동으로 이메일을 전송합니다.
CCTV 설치 장소, 날짜, 시각 뿐만 아니라 폭행이 검출된 영상의 대표 이미지를 한눈에 확인할 수 있습니다.
- 사용자는 이메일에 첨부된 “사이트로 이동하기” 버튼을 클릭하면 실시간 **CCTV** 영상과 문제의 폭행 후보 영상을 확인할 수 있습니다.

[Web Server] 3) 실시간 CCTV 화면과 녹화된 영상 열람

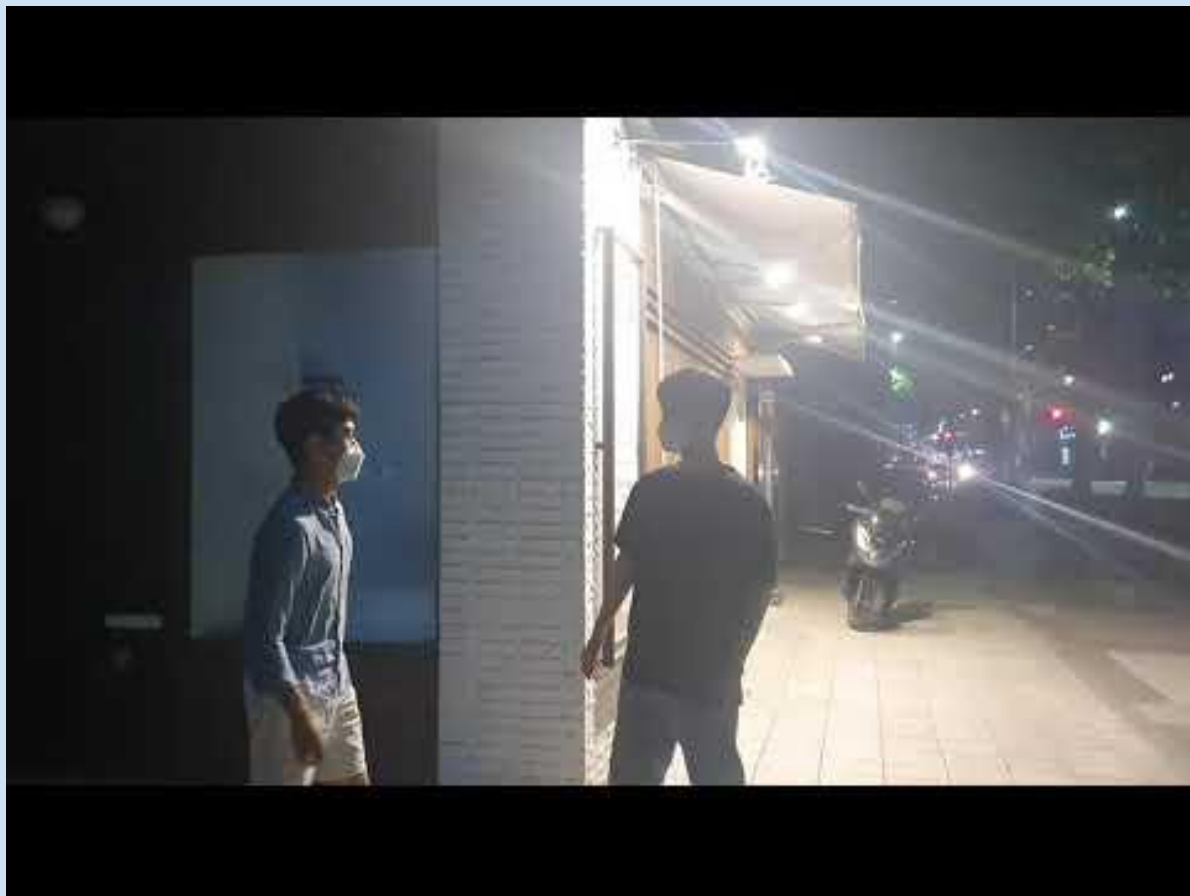


실시간 CCTV 영상



녹화된 폭행 후보 영상 확인

7. [시연 영상]



Make your neighborhood more safe..

Crime, stop!

8. 보조 자료

SlowFast, YOLO 연산량 표,

| model | flow | pretrain | top-1 | top-5 | GFLOPs×views |
|-------------------------------|------|----------|-------------|-------------|--------------|
| I3D [5] | | ImageNet | 72.1 | 90.3 | 108 × N/A |
| Two-Stream I3D [5] | ✓ | ImageNet | 75.7 | 92.0 | 216 × N/A |
| S3D-G [61] | ✓ | ImageNet | 77.2 | 93.0 | 143 × N/A |
| Nonlocal R50 [56] | | ImageNet | 76.5 | 92.6 | 282 × 30 |
| Nonlocal R101 [56] | | ImageNet | 77.7 | 93.3 | 359 × 30 |
| R(2+1)D Flow [50] | ✓ | - | 67.5 | 87.2 | 152 × 115 |
| STC [9] | | - | 68.7 | 88.5 | N/A × N/A |
| ARTNet [54] | | - | 69.2 | 88.3 | 23.5 × 250 |
| S3D [61] | | - | 69.4 | 89.1 | 66.4 × N/A |
| ECO [63] | | - | 70.0 | 89.4 | N/A × N/A |
| I3D [5] | ✓ | - | 71.6 | 90.0 | 216 × N/A |
| R(2+1)D [50] | | - | 72.0 | 90.0 | 152 × 115 |
| R(2+1)D [50] | ✓ | - | 73.9 | 90.9 | 304 × 115 |
| SlowFast 4×16, R50 | | - | 75.6 | 92.1 | 36.1 × 30 |
| SlowFast 8×8, R50 | | - | 77.0 | 92.6 | 65.7 × 30 |
| SlowFast 8×8, R101 | | - | 77.9 | 93.2 | 106 × 30 |
| SlowFast 16×8, R101 | | - | 78.9 | 93.5 | 213 × 30 |
| SlowFast 16×8, R101+NL | | - | 79.8 | 93.9 | 234 × 30 |

| Model | GFLOPs | mAP@0.5 (%) | NANO (FPS) | TX2 (FPS) | NX (FPS) | RTX (FPS) | i3 (FPS) | i5 (FPS) | i7 (FPS) | i9 (FPS) |
|--|--------|-------------|------------|-----------|----------|-----------|----------|----------|----------|----------|
| State-of-the-art light-weight YOLOs: input dimension 416 × 416, with two feature maps (13 × 13 and 26 × 26) | | | | | | | | | | |
| YOLOv3-tiny | 5.57 | 33.1 | 17 | 45 | 49 | 166 | 20 | 25 | 26 | 74 |
| YOLOv3-tiny-prn | 3.47 | 33.1 | 23 | 58 | 55 | 166 | 27 | 31 | 33 | 94 |
| YOLOv4-tiny | 6.91 | 40.2 | 18 | 44 | 46 | 165 | 16 | 22 | 23 | 64 |
| SF-YOLO (ours): input dimension 416 × 416, with two feature maps (13 × 13 and 26 × 26) | | | | | | | | | | |
| small | 2.59 | 33.1 | 34 | 77 | 77 | 170 | 32 | 39 | 41 | 118 |
| medium | 3.69 | 37.7 | 26 | 62 | 70 | 168 | 26 | 33 | 33 | 96 |
| large | 5.02 | 40.4 | 20 | 49 | 58 | 165 | 21 | 27 | 28 | 79 |

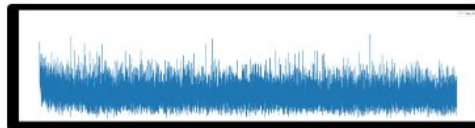
SlowFast 파라미터 사이즈 132 MB
YOLOv4 tiny 파라미터 사이즈 23 MB

멘토링 활용

- 멘토님과 Slack을 구성하여 진행했습니다.

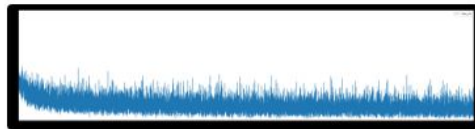
아래 그래프는 lr을 변경하기 전 loss 그래프입니다.

image.png ▾



이 그래프는 lr을 개선 한 loss그래프 입니다. 하지만 아직도 수렴하지 않은 양상을 보입니다.

image.png ▾



공공 데이터 셋을 살펴보니 애매한 데이터가 많이 존재하여 loss에 영향을 준다고 판단했습니다.

공공 데이터 셋을 정제 하려고 하는데 혹시 좋은 방법이 있을지 여쭙어 보고 싶습니다!

항상 감사합니다!



Eunyoung 1:22 AM

오 ㅎㅎㅎ 자료 공유 감사여 일단 데이터셋에서 어떤 문제를 보시고 계신가요? 제가 데이터를 다 보지 않아서. ㅎㅎ 기억나는거는 녹색 바탕에서 짙은 데이터들인데 그런 데이터들을 줄여보면 어떨까 궁금하네여 일단 loss가 줄지 않는거는 특정 테스트 데이터에서는 계속 잘못된 값이 나오는거 같은데 어떤 데이터에서 잘 안되는지 분석하고 방법을 찾아보는게 어떨까 하는 생각이 드네여



멘토링 관련

- notion을 활용하여 실험 정보를 멘토님과 공유했습니다.

2021Capstone

Quick Find

All Updates

Settings & Members

WORKSPACE

멘토님 질의 응답 내용

학교 제출용 회의록(금요일)

Capstone - 23

- 임태현
- 김소망
- 박건희
- 함윤석
- 우리끼리 회의록
- 가 학교 제출용 회의록(금요일)
- 가 멘토님 질의 응답 내용

모델 실험

- 도커

Capstone


PRIVATE

Add a page

Templates

Import


Trash



모델 실험

데이터 구성

- AI HUB에서 제공하는 이상행동 CCTV 영상 AI데이터를 사용
- 링크 : <https://aihub.or.kr/aidata/139>



```
<?xml version="1.0" encoding="UTF-8" standalone="true"?>
<annotation>
  <folder>assault</folder>
  <filename>7-3_cam02_assault01_place04_day_summer.mp4</filename>
  <source>
    <database>NIA2019 Database v1</database>
    <annotation>NIA2019</annotation>
  </source>
```


멘토링 내용

<멘토님과 진행한 “google meet”에서 발췌한 내용입니다>

○ 질문: 영상에서 2인 이상의 사람이 검출 될 때만 AI server로 영상을 전송해서 행동 인식에 소비되는 컴퓨팅 자원 소모량을 줄이고 싶은데 이 아이디어는 어떤가요?

○ 답변: 좋은 것 같아요. HW 성능을 고려하면, Face Detection 같은 가벼운 모델을 사용하는 것을 추천합니다.

적합한 경량 인공지능 모델(YOLOv4-tiny)을 찾아서 적용하여 사람이 검출되는 영상만 AI Server로 전송했습니다.

멘토링 내용

<멘토님과 진행한 “google meet”에서 발췌한 내용입니다>

○ 질문: 인공지능 학습 결과 정확도가 낮고 Loss가 수렴하지 않습니다. 원인을 알 수 있을까요?

○ 답변: Loss의 양상을 봤을 때 학습률(Learning Rate)이 너무 큰것 같습니다. 학습률을 조금씩 낮춰가며 테스트 해보세요.

< “Slack”의 대화 내용에서 발췌한 내용입니다>

○ 질문: 저희 컴퓨터 사양이 좋지 않아 모델의 모든 파라미터를 학습시킬수 없는 상황입니다. 따라서 모델의 일부분만 학습중입니다. 하지만 이로 인해 낮은 정확도를 보입니다. 혹시 좋은 해결 방법이 있을까요?

○ 답변: 공유해준 자료를 보니 FC층만 가중치 갱신이 일어나는 것 같습니다. 학습시간은 줄일 수 있지만 정확도는 보장 할 수 없지요. BackBone에서 FC층에 가까운 층은 가중치 갱신을 할 수 있도록 설정 후 학습을 진행해보세요. 학습시간은 조금 늘겠지만 정확도는 증가할거예요.

멘토링 내용

- 멘토님의 핵심 답변 내용인
 - 학습률(Learning Rate)이 너무 크다.
 - FC층만 가중치 갱신이 일어나는 것 같다. BackBone에서 FC층에 가까운 층은 가중치 갱신을 할 수 있도록 설정 후 학습을 진행해라

위 두가지 부분에 대한 답변을 듣고 기존 Learning Rate를 점진적으로 줄여가며 test했습니다.

또 BackBone의 가장 마지막 층부터 가중치 업데이트를 허용해가며 test 했습니다.

그 결과 정확도 개선(60%→90%)에 성공 했습니다.

멘토링 내용

○ 질문: 저번에 모델을 학습시켰을때는 punch,kick,normal로 나누어서 학습을 진행했습니다. 하지만 loss가 수렴하지 못해서 abnormal,normal로 나누어서 학습을 진행 했습니다.

하지만 이도 마찬가지로 loss가 수렴하지 못했습니다. 멘토님이 조언해 주신대로 learning rate를 변경하여 학습을 진행해봤습니다.

loss의 진동폭이 줄어 들었지만 그래도 아직 진동하는 것을 볼 수가 있습니다. 이를 극복하려면 어떻게 해야 할까요?

○ 답변: 데이터셋에서 문제가 있다고 생각이 듭니다. 기억나는거는 녹색 바탕에서 찍은 데이터들인데 그런 데이터들을 줄여보면 어떨까 궁금하네요 일단 loss가 줄지 않는거는 특정 테스트 데이터에서는 계속 잘못된 값이 나오는거 같은데 어떤 데이터에서 잘 안되는지 분석하고 방법을 찾아보는게 어떨까 하는 생각이 듭니다.

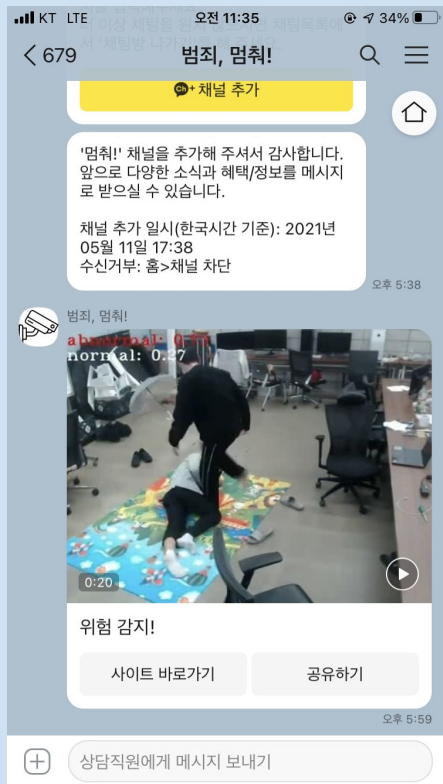
18~20page 에서 설명드린 방법으로 노이즈를 제거하여 정확도 개선(90%→94%)에 성공 했습니다.

[보완할 점] - AI Model



- 고품질의 데이터 셋을 구축한다면, 범죄 상황 외에도 **실신** 같은 위급 상황 또한 검출하도록 개선할 수 있습니다.

[보완할 점] - Web Server



- **메일보다 높은 접근성 :**
사업자등록을 한다면 카카오톡 플러스 친구 채널을 “비즈 채널”로 전환하여 검출 내역을 **카카오톡 메시지**로 자동으로 전송할 수 있습니다.
- 웹 페이지 **UI**를 개선하면 사용자에게 더욱 향상된 편의성을 제공할 수 있습니다.