

쓰레기 불법 투기 감지를 위한 영상처리 알고리즘 연구

20191868 김용현 20191880 윤종훈 20211942 노승민
20221085 김기홍 20221123 하현경

1. 개요

2. 연구 방법

3. 실험

4. 결론

■ 필요성

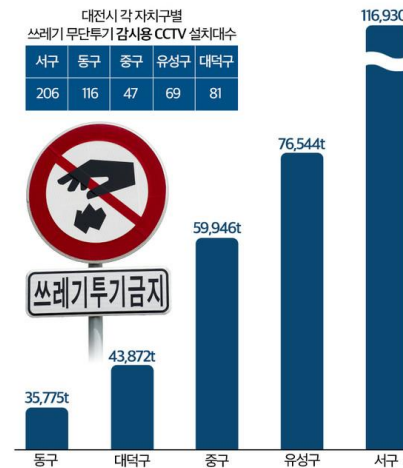
- 쓰레기 무단 투기 문제는 도시 미관뿐만 아니라 공공장소의 위생과 환경에 직접적인 영향을 미침
- 기존의 인력 기반 감시 방식은 감시 사각지대와 피로로 인한 문제점이 있으며, 특정 지역에서는 실시간 대응이 어려운 상황
- 이를 해결하기 위해 CCTV와 영상처리 기반 자동 감지 시스템을 결합하여 무단 투기 행위를 실시간으로 감지하고 기록하는 시스템을 개발



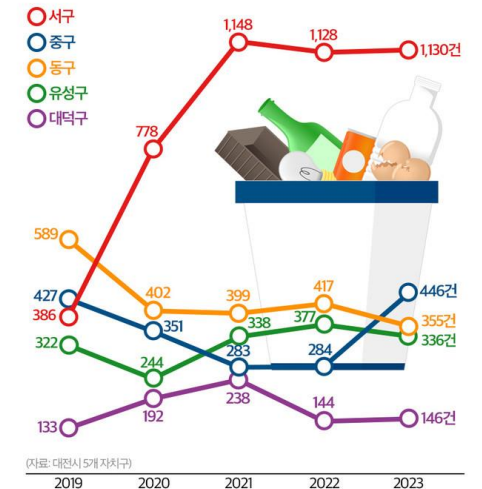
■ 목표

- 다양한 시나리오와 환경을 반영한 데이터셋을 수집하고 전처리하여, 외부 환경 요인의 변동성을 최소화
- 기존의 행동 인식 모델을 무단 투기 감지에 맞게 특화시켜 학습

2023년 대전시 각 자치구별 쓰레기 배출량



대전 각 자치구별 생활쓰레기 무단투기 적발건수

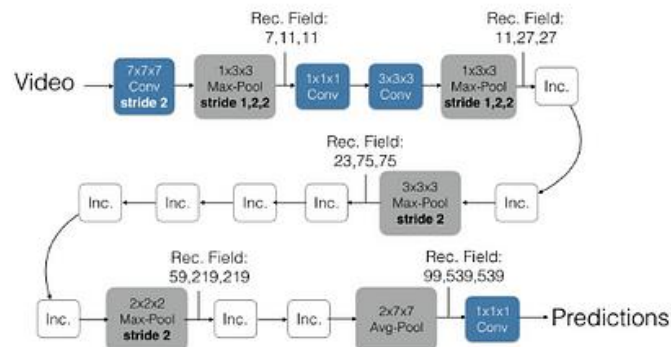


연구 방법 - 불법 투기 행위 인식 신경망

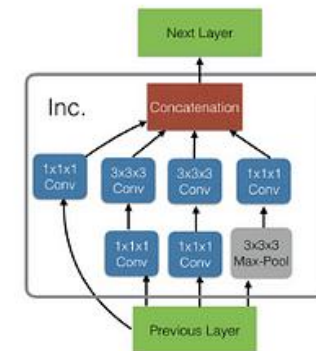
■ I3D(Inflated 3D ConvNet)

- Google DeepMind에서 제안한 비디오 분석 모델
- Inception 구조를 2D에서 3D로 확장하여 비디오의 공간적 및 시간적 정보를 동시에 처리할 수 있도록 설계됨
 - 필터와 풀링 커널이 2D에서 3D로 확장됨
 - 이로써 시간 정보를 처리할 수 있음
 - 높이 x 너비 x 시간
- Inception
 - 다양한 크기의 필터를 병렬로 적용하여 다양한 스케일의 특징을 추출
 - 3D로 확장하여 공간적, 시간적 특징을 동시 추출

Inflated Inception-V1



Inception Module (Inc.)

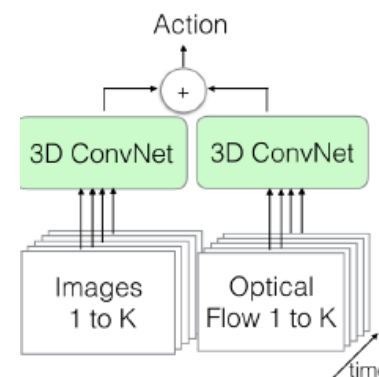


■ I3D(Inflated 3D ConvNet)

• Two-Stream Network

- I3D 모델은 RGB 스트림과 Flow 스트림의 두 가지 스트림으로 구성됨
- RGB 스트림
 - RGB 프레임을 입력으로 받아 공간적 특징을 학습
 - 비디오의 정적 이미지 정보 처리하여, 물체의 모양, 색상, 질감 등을 인식
- Flow 스트림
 - Optical Flow 맵을 입력으로 받아 시간적 특징으로 학습
 - 비디오의 동작 정보를 처리하여, 물체의 움직임을 인식
- 결합
 - 두 스트림의 출력을 결합하여 최종 예측 진행
 - 보통 두 스트림의 특징을 병합하거나, 평균을 취하는 방식으로 이루어짐
 - 비디오의 공간적 및 시간적 정보를 모두 활용할 수 있게 됨

e) Two-Stream 3D-ConvNet



연구 방법 - 불법 투기 행위 인식 신경망

■ I3D(Inflated 3D ConvNet)

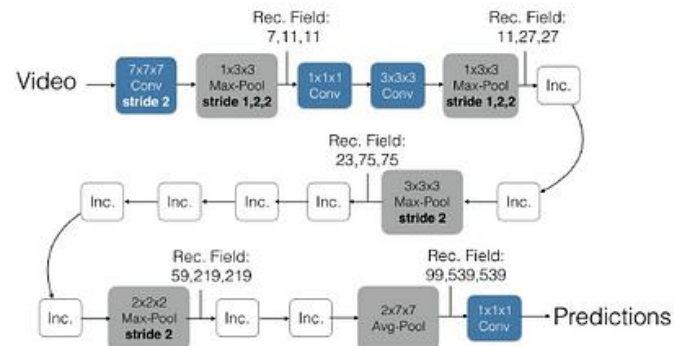
• 입력

- 4D 텐서 형태($N * C * T * H * W$)
- N : 배치크기
- C : 채널 수
 - RGB 비디오의 경우 3(R, G, B)
 - Optical Flow 같은 경우(Flow X, Flow Y)
- T : 프레임 수
 - 보통 16 or 32 프레임 사용
- H : 높이
- W : 너비

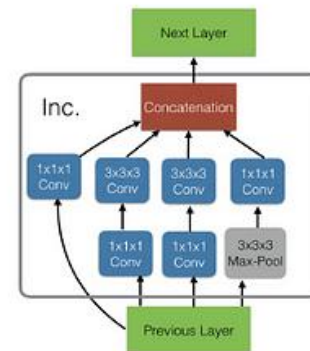
• 출력

- 2D 텐서($N * K$)
- N : 배치 크기
- K : 분류하고자 하는 클래스의 수

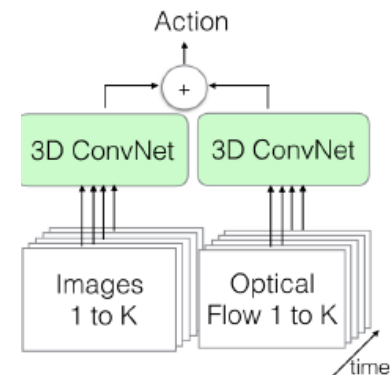
Inflated Inception-V1



Inception Module (Inc.)



e) Two-Stream 3D-ConvNet

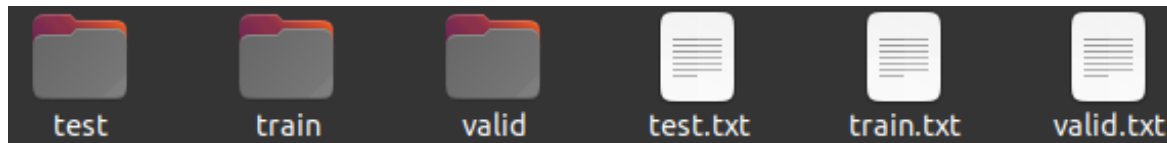


연구 방법 - 쓰레기 투기 데이터 셋

- [AI-Hub] 이상행동 CCTV 영상
 - Network : I3D(Inflated 3D ConvNet)
 - Task : 비디오 행동 인식
 - DataSet : 이상행동 CCTV 영상
 - Class : 08. 투기(dump) 클래스 (총 623개 / 50:07:11)
- 데이터 셋 분류
 - Train (75%)
 - 393개 비디오
 - Valid (15%)
 - 100개 비디오
 - Test (15%)
 - 102개 비디오



```
# dataset settings
dataset_type = 'VideoDataset'
data_root = 'data/kinetics400/videos_train'
data_root_val = 'data/kinetics400/videos_val'
ann_file_train = 'data/kinetics400/kinetics400_train_list_videos.txt'
ann_file_val = 'data/kinetics400/kinetics400_val_list_videos.txt'
ann_file_test = 'data/kinetics400/kinetics400_val_list_videos.txt'
```



실험 - I3D 신경망 Fine Tuning

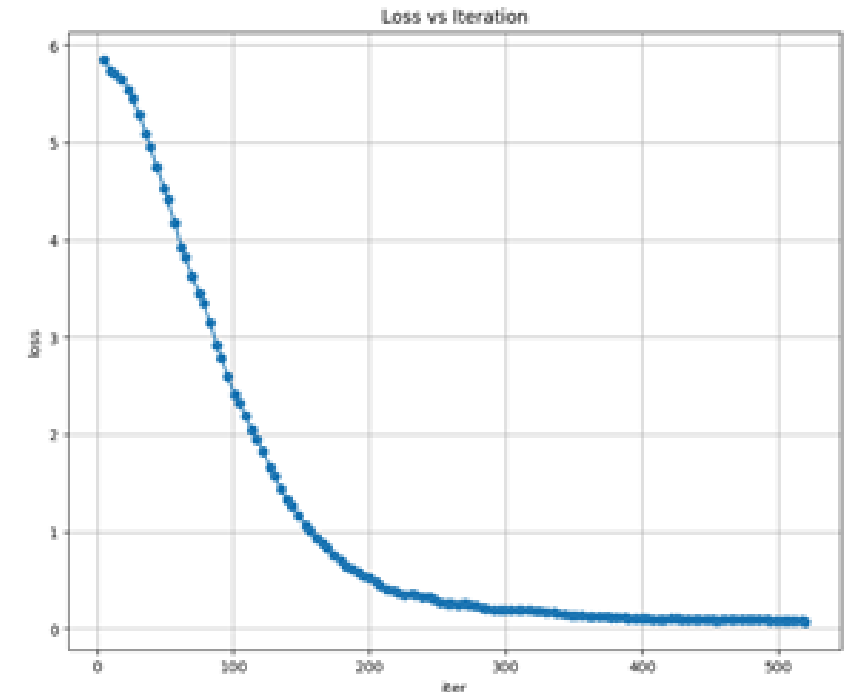
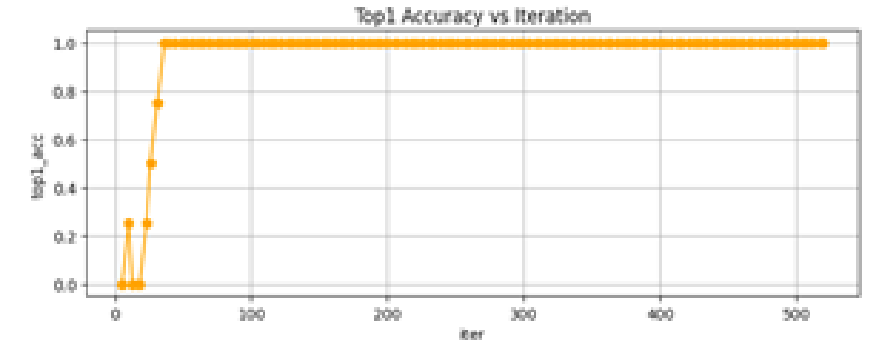
■ Fine Tuning

- 사전 학습된 I3D 신경망(ResNet50)을 바탕으로, 투기(dump) 클래스를 인식하도록 재학습
- 하이퍼파라미터
 - 학습률(learning rate) : $1.0e-5$
 - 최적화 방법 : SGD
 - 손실 함수 : 교차 엔트로피(Cross-Entropy)
 - 배치 크기(Batch Size) : 4 (2080Ti GPU x 8 ea)
 - 에포크(Epoch) : 40
- 학습 환경
 - 다중 GPU 환경에서 학습 진행하여 연산 속도 최적화
 - 학습 중 손실 함수 수렴 여부를 통해 모델의 안정성 확인

실험 - Fine Tuning 결과

■ Fine Tuning 결과

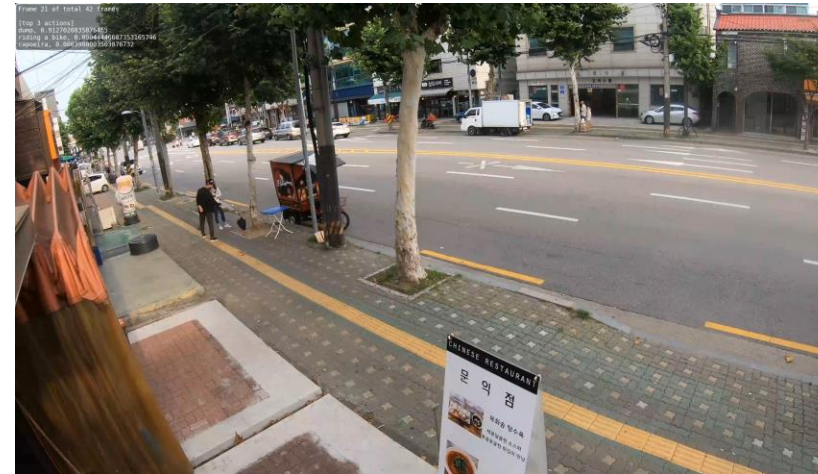
- Top 1 Accuracy
 - 100 iteration 이후 Top 1 Accuracy가 1.0로 높은 모습을 보임
 - 100 iteration 이후에 Dump 클래스의 점수가 가장 높음을 의미
- 손실 값
 - 안정적이게 수렴하는 모습을 보임
 - 500 iteration 이후 0에 가깝게 수렴하는 모습



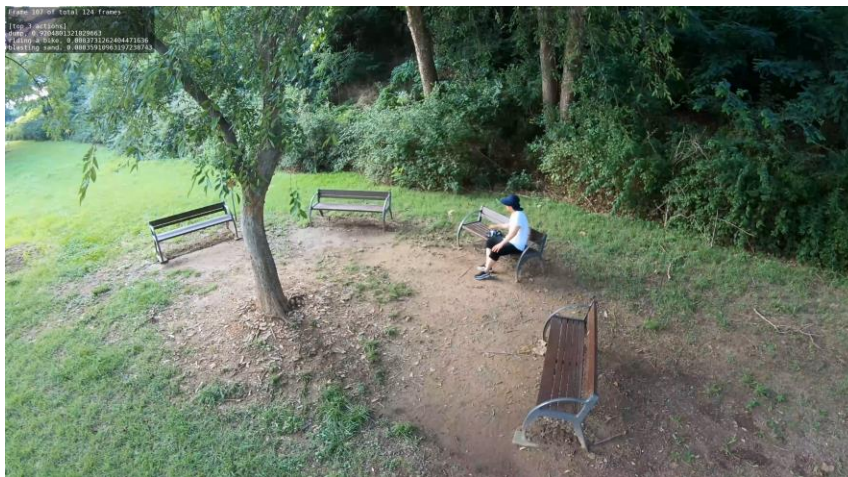
실험 - 추론 결과



투기 행위 신뢰도 : 0.78 (야간)



투기 행위 신뢰도 : 0.98 (주간)



투기 행위 신뢰도 : 0.91 (주간)



투기 행위 신뢰도 : 0.92 (주간)

■ 기대효과

- 공공장소 관리 효율성 향상
 - 자동화된 탐지 시스템을 통해 감시 인력의 부담 감소
 - 즉각적인 대응이 가능
- 운영 비용 절감
 - 기존 CCTV 인프라를 활용함으로써 설치 및 운영 비용을 절감
 - 감시 사각지대를 줄여 추가 감시 자원의 투입 최소화
- 기술 확장성
 - 다양한 불법 행위 탐지와 행동 분석 시스템으로 확장 가능

감사합니다