| **GAN을 이용한 Object Detection Data Augmentation**    **요 약**  딥 러닝 관련 연구분야에 있어서 GAN(Generative Adversarial Networks)을 통한 학습 데이터 증강은 빼놓을 수 없는 필수 요소이다. 따라서 본 문서는 CCTV 속 사고(실족, 폭행, 돌진, 유기 등)행동 데이터를 DC(Deep Convolutional) GAN을 이용해 데이터를 증대 시키고, 이를 다시 딥 러닝 모델에 학습시켜 상황에 대한 구체적 분석이 가능한, 개선된 성능의 AI 설계를 제안한다. |
| --- |

**1. 서론**

**1.1. 연구배경**

딥러닝에 있어서 데이터의 수는 굉장히 중요하다. 데이터는 어디에서나 사용되고 있고 좋은 데이터가 있어야 좋은 모델을 만들 수 있다. 치우친 샘플로 모델을 만드는 경우 엉뚱한 결과를 초래한다. 학습 데이터의 질에 따라 학습 결과가 좌우되므로 좋은 데이터를 가지고 학습을 시켜야한다.

하지만 데이터는 충분하지 않다. 딥러닝은 어마어마한 데이터를 학습을 위해서 요구한다. 쉽게 모을 수 있는 데이터가 있는 반면 암진단 데이터와 같은 의료 데이터는 모으기가 엄청나게 어렵다. 적은 데이터로 우수한 결과를 내기 위해서는 데이터 증강 기술이 매우 중요하다.

학습 데이터에 다양한 환경이나 특성을 넣어 변형을 하는 데이터 증강 기술은 딥러닝에 있어서 가장 핵심적인 기술이다. 괜찮은 데이터를 만들면 만들수록 딥러닝의 성능은 좋아진다. 이렇게 매우 중요한 부분이지만 데이터 증강기술은 다른 분야의 연구에 비해 소홀한 경향이 있다. 따라서 우리는 딥러닝의 핵심기술인 Data Augmentation에 집중하여 의미있는 데이터를 얻고 이를 실제로 적용시켜 성능이 개선되는지에 대한 연구를 진행하고자 한다.

GAN(Generative Adversarial Networks)을 통해서 여러 기법을 적용 시킬 것이다. GAN은 영상 변환을 해주는 방법이다. 예를 들어 아침 영상을 밤 영상으로 만들거나 스케치를 한 사진을 기반으로 실제 사진처럼 보이게 변환해줄 수 있다. GAN을 이용해서 원본 데이터를 변형시켜 데이터를 증강 시키고 딥러닝에 학습시킬 것이다. 여러 기법은 다음과 같다. 일부 영역을 잘라내기, 좌우 또는 상하 반전시키기, 회전시키기, 위치 바꾸기, 크기바꾸기와 같은 기본적인 증강기술과 노이즈를 추가하는 방법 그리고 블럭을 랜덤으로 처리, 블럭 생성, 매우 많은 파라미터를 가지고 원본 데이터를 변형시켜 좋은 데이터를 생성할 것이다. 많은 기법을 사용해서 다양한 가설을 세우고 기존 문제를 학습시킨 것과 Data Augmentation을 통해 데이터를 증강 시킨 후 성능을 비교하여 연구를 발전시키고자 한다.

Tensorflow, Pytorch, OpenCV, 등의 Machine Learning 라이브러리가 상용화되면서, Object Classification과 관련된 프로젝트는 튜토리얼들을 따라하며 쉽게 진행할 수 있고, 검색해 보면 이미 여러 개가 존재한다. 우리는 조금 더 도전적 가치가 있는 연구를 진행하고자 Multiple Object Tracking, Person Re-Identification, 등의 다양한 딥러닝 관련 영상처리 연구 주제에 대해 조사를 진행했다. 여러 논문과 GitHub Repository들을 탐색하며, 일부 주제들에서 공통적으로 GAN을 이용해서 학습 데이터를 증대시키는 모습을 확인할 수 있었다. 이에 착안하여, 사람의 연속적인 행동 데이터 또한 GAN으로 증대시켜서 딥러닝 모델에 학습시킨다면 사고 상황을 인식할 수 있는 CCTV를 만들 수 있다고 판단하여 주제를 선정하게 되었다. 이 연구를 통해 CCTV 통합관제센터에서 더욱 면밀하게 사고 상황을 인식하고 CCTV를 효율적으로 이용 및 관리할 수 있는 것에 목적이 있다.

**1.2. 연구목표**

CCTV 영상 데이터가 많기 때문에 학습할 데이터 또한 많다고 생각할 수 있지만, 사실 사고 발생 시점의 영상만이 유의미한 학습 데이터라고 할 수 있다. 그러므로 사고 상황의 영상이 더욱 많이 필요하다. 하지만 긴 CCTV 영상들 속에서 사고 순간을 개별적으로 추출하거나, 사고 영상을 얻기 위해 사고가 발생하기를 기다리는 것 또한 현실적이지 못하다. 따라서 우리는 이번 연구를 통해 CCTV 속 이상행동을 기반으로 사고 영상의 데이터를 추출하여, GAN을 통해 데이터를 증강시키고, 데이터 학습을 통해 이상 행동에 대해 인식할 수 있는 상황 인식 CCTV의 성능 개선에 대해 연구하는 것이 목표이다.

첫 번째 목표는 사고 영상 속 행동 데이터 추출하는 것이다. 먼저 사고 원본 영상에서 행동 데이터를 추출해야 그 데이터를 기반으로 행동 데이터를 증강 시킬 수 있다. 따라서 객체 탐지 시스템에 대한 연구를 기반으로 우리가 가지고 있는 AI HUB 이상행동 CCTV 영상 데이터 속 이상 행동에 대해 탐지하고 행동 데이터를 추출해내는 것이 첫 번째 목표이다.

둘 째, GAN으로 행동 데이터 증강 시키는 것이다. 추출한 행동 데이터를 기반으로 데이터를 다양한 기법을 통해 데이터를 증강 시킬 것이다. 기본적인 증강기술을 통해 변형시키고, 새로운 증강기술에 대해 연구하여 고품질의 데이터를 원본으로부터 생성해서 데이터를 증강 시킬 것이다.

셋 째, 데이터 학습으로 높은 행동 인식 정확도를 얻는것이다. 적은 양으로 데이터를 학습시켰을 때와 Data Augmentation을 진행 한 후 학습 시킨 때를 비교하여 성능이 어느 정도 좋아졌는지, 유의미하게 데이터를 증대시켰는지 확인하는 것이다.

**2. 관련연구 및 데이터**

**2.1. 객체 추적**

**2.1.1. 컴퓨터 비전 기반 모션 추적**

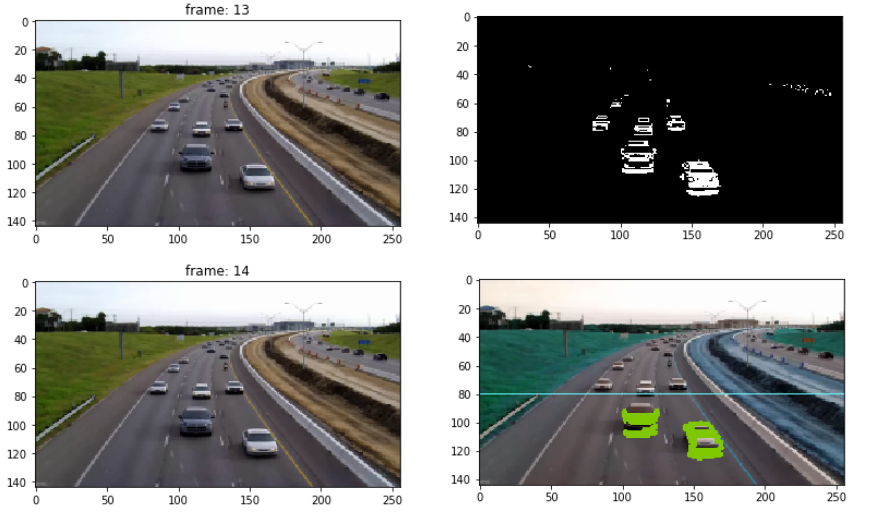
제일 구현하기 간단한 방식의 사물 추적으로, 영상 속 특정 프레임과 그 다음 프레임의 차이 값을 이용한다. 프레임 간의 차이 값은 비교가 용이하도록 가우시안 변환, 임계값 설정, 다일레이션 등의 다양한 필터를 거쳐 변환된다. 결과적으로 모션 추적 방식에서는 실시간으로 움직이는 사물의 테두리를 강조하거나 사물위에 박스를 그려 사물을 추적한다. 컴퓨터 비전 기반 모션 추적은 별도의 학습과정이 필요 없고 파이썬의 cv2 라이브러리만을 이용하여 구현 가능하지만, 오직 프레임 간의 차이를 이용하기 때문에 여러 객체를 구분하여 추적하거나 객체 자체 정보를 얻는데에는 어려움을 겪는다.

**2.1.2. 딥러닝 기반 실시간 다중 객체 추적**

****

영상에서의 자동 객체 인식 및 추적 분야에 있어 딥러닝 활용도는 점점 증가하고 있다. 무인 감시 및 범죄 예방 등 다양한 분야에서 활용되고 있고, 관련 기술에 대한 연구가 많이 진행되고 있다. 대표적인 딥러닝 기반 객체 추적 시스템은 객체 검출 알고리즘과 추적 알고리즘의 연쇄 된 형태로 구성되어 있다. 따라서 축적 시스템의 성능은 객체를 검출하는 것에 의존적인 한계가 있다. 따라서 딥러닝 기반 검출기에 대한 의존성을 낮추고 또한 추적기의 동작 속도를 증가시키는 방법에 대해 연구한다. Deep Sort 알고리즘의 객체 특징 정보 비교 과정을 변형하여 성능을 향상시키는 방법에 대해 제안하고 있다.

**2.1.3. Vehicle Detection Model**

****

CCTV 영상에 있는 차량을 자동으로 추적 및 감지하는 모델로 하루 동안 교통 분기점에 몇 대의 차량이 있는지, 몇 시에 차량이 막히는지, 어떤 종류의 차량이 통과하는지, 그리고 사고를 감지하는 것과 같이 다양하게 사용할 수 있는 차량을 추적하는 모델을 OpenCV와 Python을 통해 구현한다. 영상에서 움직이는 물체를 어떻게 탐지하는지, 실제로 어디에 사용하는지, 그리고 영상 개체 탐지를 위해 알아야하는 필수 개념(Frame Differencing, Image Thresholding, Contours Finding, Image Dilation)에 대해 알 수 있다. 움직이는 물체를 탐지하기 위해서는 Frame Differencing을 통해 알 수 있다. Image Thresholding을 통해 그레이스케일 영상의 픽셀 값은 흑백으로 표현된다. 그리고 Contours Finding은 동일한 색이나 강도를 가진 영상에서 영역의 모양을 식별하는데 사용된다. Image Dilation은 이미지 확장기능을 적용한 후 이미지에서 윤곽선을 찾아 영상의 객체를 추적할 수 있다.

**2.1.4. Person Re-Identification**

****

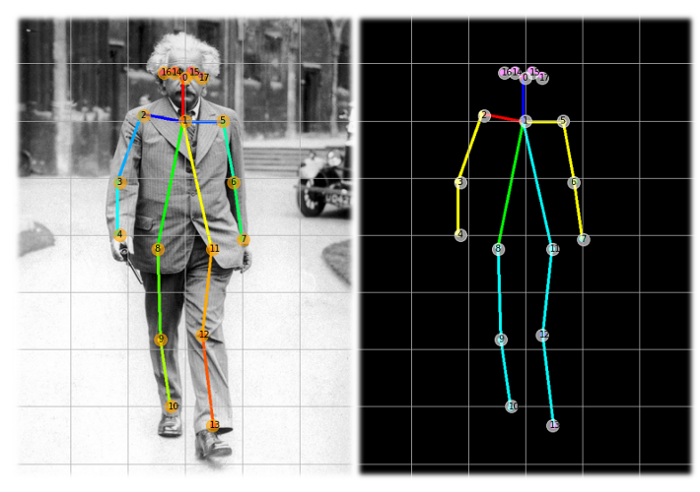
Person Re-Identification(person Re-Id)은 인접하나 겹치지 않는 CCTV 영상들 속에서 동일한 사람의 이미지를 추출하는 다중 카메라 객체 추적 방식이다. 특정 되어진 사람을 다른 영상 혹은 카메라에서도 계속 추적하며, 카메라 위치 정보와 연계한다면 동선을 파악할 수 있다는 점에서 주목받고 있다. 그러나 딥 러닝 모델에 학습시킬 Re-Id 데이터 셋이 다른 모델들에 비해 부족할 경우 한계가 드러나지만 후술할 GAN으로 개선이 가능하다.

**2.2. 객체 인식**

**2.2.1. COCO Dataset**

COCO는 Common Objects in Context의 약어로, 그 뜻 그대로 일상에서 볼 수 있는 환경 또는 사물의 이미지와 문맥(설명)을 포함한 Data Set이다. COCO는 주로 객체인식, 객체 분리, 키 포인트(관절, 눈 등) 인식, 등의 다양한 분야의 머신 러닝 학습에 이용된다.

**2.2.2. Pose Estimation**

****

Pose Estimation은 사람에 대한 사진이나 영상을 보고 사람의 자세를 추측하는 기술이다. 사람의 목, 어깨, 팔꿈치와 같은 관절 부분을 Key Point로 찾아내어 2차원이나 3차원 모델을 얻어 낼 수 있다. 조사 결과 가장 많이 사용된 Pose Estimation 기법은 Carnegie Mellon의 Robotics Institute가 제안한 Open Pose로 모든 Key Point를 찾고 연결하는 Deep Pose와 달리 Key point들이 들어갈 사람 몸을 찾는 PAF(Part Affinity Fields)를 기반으로 한다. 이 때문에 Open Pose는 이미지나 영상 속 사람의 수와 상관 없이 Key Point들을 찾을 수 있다. Open Pose는 원래 Caffe framework를 사용해서 개발 되었지만 최근에는 Tensorflow를 기반으로 개발된 Open Pose도 존재한다.

**2.3. GAN**

Generative Adversarial Network(GAN)은 적대적 학습을 하는 딥 러닝 과정을 거친 생성 모델이다. GAN에서는 실제 데이터를 보고 존재하지 않는 데이터를 만드는 생성자(Generator) 모델과 생성자가 만든 가짜 데이터와 실제 데이터를 구별하고자 하는 구분자(Discriminator) 모델이 적대적으로 경쟁한다. 이 두 모델을 적대적으로 계속 학습시켜서 구분자가 구별할 수 없는 가짜 데이터를 생성할 수 있는 생성자를 만드는 것이 GAN의 목적이다.

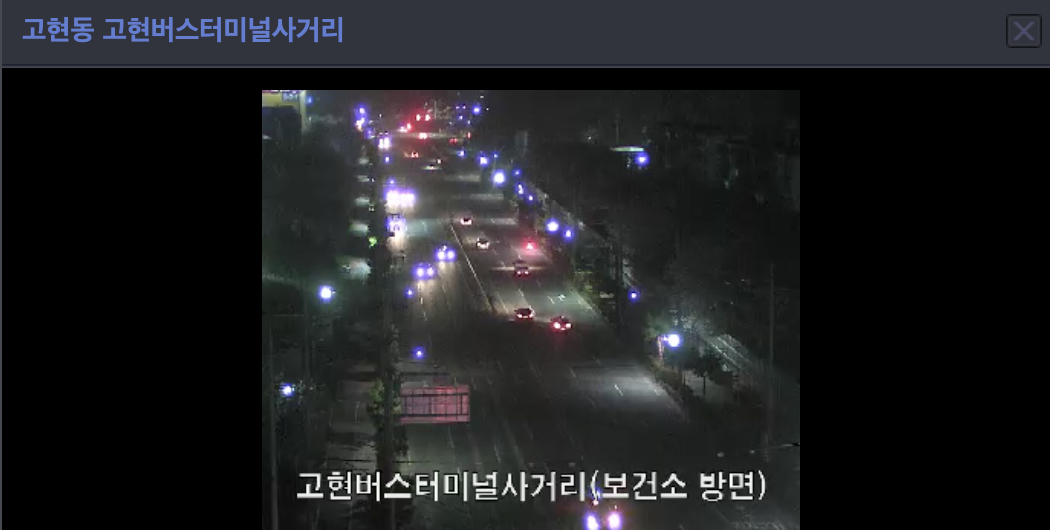
**2.3.1. DC GAN**

****

Deep Convolutional GAN(DC GAN)은 GAN의 확장 개념으로 과정을 알 수 없는 기존 GAN의 한계를 개선하고자 한다. 가장 큰 차이점으로 DC GAN은 GAN의 분석하기 어려운 fully-connected 구조 대신에 Convolutional Neural Network(CNN) 구조를 사용한다. 따라서 생성자와 구분자는 Convolutional Transpose Layer 혹은 Strided Convolutional Layer와 Batch Normalization Layer로 이루어진다.

**2.4. 영상 데이터**

**2.4.1. 도시교통정보센터 개방데이터**



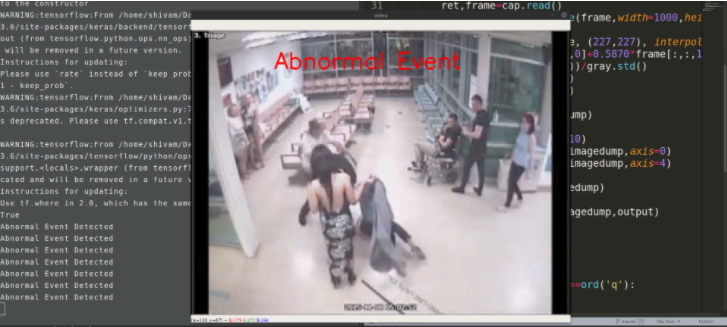
교통관제용 전국 CCTV 영상정보를 제공하고 있다. URL(키값, ip 인증)을 통해 목록 형태의 실시간 CCTV 영상정보로 실제 스트리밍 CCTV 영상을 포함하고 있다. 소통정보 데이터, 돌발정보 데이터, 교통안전 데이터, CCTV 데이터를 신청하여 제공 받았다. UTIC 개방 데이터는 경찰청 도시교통정보센터에서 제공하며 API를 통해 데이터를 URL로 전복, 화재, 기름 유출, 차도 이탈 등에 대한 정보를 제공받을 수 있다.

**2.4.2. AI HUB 공항 이상행동 CCTV 영상**

****

공항 출입국관리 구역 내에서 발생할 수 있는 다수의 움직이는 사람의 신원을 자동으로 식별하고, 위험 상황을 실시간으로 탐지하는 인공지능 시스템 구축, 실증, 검증 지원 및 육성 과제로 인공지능 식별, 추적 시스템의 공항 출입국 관리 실증 적용을 위한 자체 학습 데이터 가공 및 제작하여 제공하고 있다. 제공 데이터량은 10,000컷 비디오 데이터셋 촬영을 하였고, alchera 이상행동 데이터, nabi 이상행동데이터의 4가지 이상행동(돌진하는 행위, 역방향 이동, 물건을 놓고 사라지는 유기, 2인 감지), cubox 이상행동데이터의 8가지 이상행동(돌진, 방치, 2인감지, 역방향, 정상, 실신, 파손, 폭행)에 대한 데이터를 제공한다.

**2.5. 상황 분석 CCTV(선행 연구)**

****

CCTV를 계속 보고 감시하는 일은 매우 소모적이고 지루한 일임에도 불구하고, CCTV 카메라의 수와 중요성은 날이 갈 수록 증가하고 있다. 따라서 사람이 CCTV를 계속 보고 있지 않아도 되도록 딥 러닝과 관련된 연구가 감시카메라 영역에도 발전되어 졌다. Avenue Dataset을 사용한 이상 행동 관측 연구에서는 인코더와 디코더로 이루어진 3D Neural Network을 사용해 reconstruction loss를 구하고 이상 행동 여부를 구분했다. 학습한 모델을 바탕으로 실시간으로 이상 행동이 발생하면 ‘Abnormal Event’라는 알림을 보낸다.

**2.6. 상황 분석 CCTV의 문제점 및 해결 방안**

**2.6.1. 문제점**

이상 행동 분석에 필요한 데이터는 다른 Classification Problem들과 비교하여 데이터가 매우 부족하고 카메라의 위치와 각도, 주변 환경, 등의 요소에 따른 변인이 너무 많이 발생한다. 또한 이상 행동을 인식해도 ‘Abnormal Event’라는 알림을 보낼 뿐 어떠한 상황인지는 알지 못한다.

**2.6.2. 해결 방안**

Pose Estimation을 사용해서 다른 요소를 배제하고 사람의 행동 데이터 만을 추출하기 때문에 여러 환경에서 적용 가능한 유의미한 학습 데이터와 정확성을 얻을 수 있다. 또한 DC GAN을 사용함으로써 실제와 비슷한 학습 데이터를 증대시킬 수 있으므로 학습 데이터 부족 문제를 해결할 수 있고 상황에 대한 분석(Classification)이 가능해질 것이다.

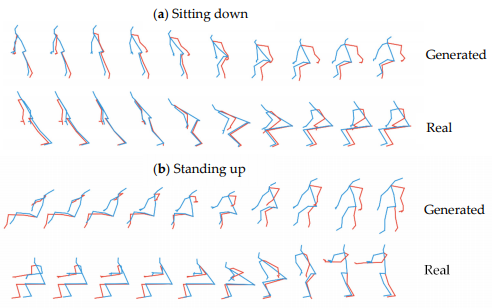
**3. 프로젝트 내용**

**3.1. 시나리오**

**3.1.1. Pose-Estimation를 통한 행동 추출**

Pose-Estimation 모델을 사용해서 Avenue Dataset, AI HUB Dataset, 등의 사고 상황 Dataset들로부터 연속적인 사람 행동 데이터(key-point data)를 추출하고 분류한다.

**3.1.2. DC GAN을 통한 데이터 증강 및 학습**

****

DC GAN으로 행동 데이터를 증강시키고 이를 학습 데이터로써 딥 러닝 모델에 학습시킨다.

**3.1.3. 실시간 사고/행동 인식**

학습 데이터로 부터 훈련된 모델에 테스트 데이터를 넣어서 실시간으로 정확한 분석이 가능한지 확인하고, 가능하다면 정확률을 높이기 위한 분석을 하고 위 과정을 다시 반복한다.

**3.2. 요구사항**

**3.2.1. Pose Estimation에 대한 요구사항**

* Tensorflow 등의 라이브러리와의 버전 호환성을 확인한다.
* GPU(CUDA, cuDNN)를 사용한 성능 향상을 확인한다.

**3.2.2. DC GAN에 대한 요구사항**

* DC GAN으로 생성된 학습 데이터가 적합하지 않다면 PAC GAN, HP GAN등의 다른 모델의 사용을 검토한다.
* 적절한 Data Augmentation 방법을 사용해야한다. 기본적인 증강 기술과 새로운 방법의 데이터 증강 기술을 가정하고 수치화 해야한다.

**3.2.3. 실시간 사고/행동 인식에 대한 요구사항**

* 데이터 증강 하기 전과 증강 후 성능 비교 분석이 필요하다.
* 이상 행동 데이터가 의미있는 데이터로 증강되었는지 분석해야한다.

**4. 향후 일정 및 역할 분담**

**4.1. 향후 일정**

| 4주차 | * 주제관련 자료조사 * 기초조사서 제출 * 멘토 회의(장윤호 멘토님) | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| 5주차 | * 객체 추적 시스템 개발 진행 * 주간보고서 제출 * 3월 면담보고서 제출 * 04-05(월) 16:15~16:30 배성호 지도교수님과 3차 회의 | | | | | |
| 6주차 | * 객체 인식 개발 진행 * 멘토 회의(장윤호 멘토님) * 주간 보고서 제출 * 멘토 면담확인서 제출 | | | | | |
| 7주차 | * GAN을 활용한 Data Augmentation 개발 진행 * 중간보고서 서론 작성, * 04-19(월) 16:15~16:30 배성호 지도교수님과 4차 회의 * 주간보고서 제출 | | | | | |
| 8주차 | * 원본 데이터 학습과 Data Augmentation 결과로 학습시킨 데이터 성능 비교 진행 * 가설 검증, 수치화, 문서화 진행 * 중간보고서 본문 작성 * 주간보고서 제출 | | | | | |
| 9주차 | * 중간보고서 제출 준비 및 검토(참고 문헌, 부록, 시연동영상 URL) * 05-03(월) 16:15~16:30 배성호 지도교수님과 5차 회의 * 중간보고서 제출 | | | | | |
| 10주차 | * 최종보고서 준비(요약 및 서론 변경사항 정리, 관련 연구 구체화, 프로젝트 내용 업데이트, 요구사항 업데이트, 시스템 구성도 및 UML Diagram 설계한 것 정리) * 추가 개발 진행 * 주간보고서 제출 * 4월 면담보고서 제출 | | | | | |
| 11주차 | * 최종보고서 준비(구현내용 정리, 프로젝트 연구 결과 정리 및 성능평가) * 추가 개발 진행 * 05-17(월) 16:15~16:30 배성호 지도교수님과 6차 회의 * 주간 보고서 제출 * 멘토 면담확인서 제출 | | | | | |
| 12주차 | * 최종보고서 준비(결론, 기대효과, 추후 연구방향, 참고문헌) * 최종 발표 준비(시나리오, PPT, 시연) * 주간 보고서 제출 | | | | | |
| 13주차 | * 05-25(화) 최종 발표(온라인으로 평가 진행) * 최종 보고서 검토 | | | | | |
| 14주차 | * 최종 보고서 검토 | | | | | |
| 15주차 | * 최종 보고서 제출 | | | | | |

**4.2. 역할 분담**

* Data Detection: 이광원, 장규범
* Pose Estimation: 이광원, 장규범
* Data Augmentation: 이광원, 장규범
* Code Compatibility: 이광원
* Document Management: 장규범

**5. 결론 및 기대효과**

해당 연구를 통해 CCTV 이상행동 데이터로부터 추출한 데이터를 다양한 기법으로 데이터 증강시켜 적합한 이상 행동 데이터를 얻을 수 있다. 기존의 데이터로 학습시킨 딥러닝 모델의 성능을 평가하고 Data Augmentation한 데이터로 학습시킨 모델을 비교 분석함으로써 얼마나 모델 성능이 좋아졌는지 의미있는 데이터를 만들었는지 알 수 있다. 각 가설을 진행하고 수치화 함으로써 이상 행동 데이터 중 가장 적합한 Data Augmentation 방법을 제시할 수 있다.

또한 이상 행동 데이터를 바탕으로 고성능 상황 인식 CCTV를 만들 수 있다. 이상 행동이 발생할 시 행동에 대해 인식할 수 있는 상황 인식 CCTV의 성능을 개선할 수 있고, 이를 통해 효율적인 CCTV 통합관제센터를 운영할 수 있는 상황 인식 CCTV를 제공할 수 있다.

**6. 참고문헌**

[1] 박호식, 배철수. Real-time recognition and Tracking System of Multiple Moving Objects. 다중 이동 객체의 실시간 인식 및 추적 시스템. 한국통신학회논문지 ‘11-07 Vol 36 NO.7

[2] 서울대학교 산학협력단. Development of Big data-based Intelligent CCTV Search Technique. 빅데이터기반 지능형 CCTV 영상 검색 기술 개발. 경찰청. 2018.04.13

[3] SAMUEL MURRAY. Real-Time Multiple Object Tracking. STOCKHOLM, SEWDEN 2017.09.28

[4] Alex Bewley , Zongyuan Ge , Lionel Ott, Fabio Ramos, Ben Upcroft. SIMPLE ONLINE AND REALTIME TRACKING. Queensland University of Technology, University of Sydney. 2017.07

[5] [hu64/SpotNet: Repository for the paper SpotNet: Self-Attention Multi-Task Network for Object Detection](https://github.com/hu64/SpotNet)

[6] [Zhongdao/Towards-Realtime-MOT: Joint Detection and Embedding for fast multi-object tracking](https://github.com/Zhongdao/Towards-Realtime-MOT)

[7] [Build your own Vehicle Detection Model using OpenCV and Python](https://www.analyticsvidhya.com/blog/2020/04/vehicle-detection-opencv-python/)

[8] [Deep Surveillance with Deep Learning – Intelligent Video Surveillance](https://data-flair.training/blogs/deep-surveillance-with-deep-learning-intelligent-video-surveillance-project/)

[9] Xingyi Zhou, Vladlen Koltun, Philipp Krahenbuhl. Probabilistic two-stage detection. 2021.03.12

[10] Chengyuan Zhang, Lei Zhu, ShiChao Zhang. PAC-GAN: An Effective Pose Augmentation Scheme for Unsupervised Cross-View Person Re-identification. 2019.06.05

[11] Zhongyue Huang, Jingwei Xu and Bingbing Ni. Human Motion Generation via Cross-Space Constrained Sampling. Shanghai Jiao Tong University, China

[12] 국회입법조사처. CCTV 통합관제센터 운영실태 및 개선방안. NARS 입법,정책보고서 제29호 2019.11.01

[13] [적은 데이터로 AI 가르친다…엔비디아, GAN 모델 훈련 혁신](http://www.aitimes.com/news/articleView.html?idxno=134695)

[14] ['딥러닝'으로 위험도 예측](http://www.engjournal.co.kr/news/articleView.html?idxno=1382)

[15] [[전문가 기고] 학습 능력 높이는 데이터 증강 기술](http://www.hellot.net/new_hellot/magazine/magazine_read.html?code=202&idx=47643&public_date=2019-08)

[16] 이광형, 김용균, 지정규, 오해석. Object Tracking and Face extract by Real-time Image 실시간 영상에서 객체 추적 및 얼굴 추출. 제 19회 한국정보처리학회 춘계학술발표대회 논문집 제10권 제 1호