

# GAN을 이용한 Pose Detection Data Augmentation

이광원, 장규범, 배성호  
경희대학교 컴퓨터공학과

[kwlee3@khu.ac.kr](mailto:kwlee3@khu.ac.kr), [jkb2221@khu.ac.kr](mailto:jkb2221@khu.ac.kr), [shbae@khu.ac.kr](mailto:shbae@khu.ac.kr)

## Data Augmentation using GAN for Improving Pose Detection

Kwang-Won Lee, Kyu-Bum Jang, Sung-Ho Bae  
Department of Computer Science and Engineering, KyungHee University

### 요 약

최근 4차 산업혁명의 핵심기술인 인공지능(AI) 기반의 영상분석 기술을 활용해 CCTV에 찍히는 사람, 차량 등의 움직임을 자동으로 포착해 위험 상황을 감지하여 대응하는 지능형 관제 시스템을 확대 구축해나가고 있다. 지능형 CCTV의 핵심기술인 딥러닝은 많은 양의 데이터와 이에 대한 정답을 동시에 학습시키는 방식으로 학습이 이루어지기 때문에 CCTV의 활용도 및 정확도를 높이기 위해서는 다양한 행동 기반의 영상 이미지 구축이 선결되어야 한다. 하지만 이상행동에 대한 데이터는 아직 부족한 실정이다. 이에 대해 본 논문에서는 GAN(Generative Adversarial Network)을 사용하여 CCTV 속 이상행동 데이터를 증강하고, 이를 다시 딥러닝 모델에 학습시켜 상황에 대한 구체적 분석이 가능한 개선된 성능의 AI 설계를 제안한다.

### 1. 서 론

최근 딥러닝 기술의 발전으로 영상분석 기술을 활용한 CCTV의 지능화가 진행되고 있다. 코로나 19 현행 방역 지침인 1~2m 거리에 가까워지면 ‘방역 사항 준수 안내’ 알람이 울리고, 마스크 미착용은 물론 ‘텍스트’까지 감지하여 경고 문구가 나오기까지 하는 수준에 이르렀다. 사람을 감지하면 옷의 색상, 성별, 나이, 안경 착용 등을 분석하고, 차량의 경우 종류, 색상, 번호 등 구체적인 정보를 감지한다. 또한, 이상행동 및 위험 상황을 감지하여 관련 영상을 관제요원에게 우선 표출하는 시스템을 구축해나가고 있다. 이때 사용하는 딥러닝 기술은 다양한 CCTV 영상을 인공지능이 스스로 학습하는 기법으로 다양한 영상을 누적시켜 객체 인식률을 높이고 있다. 지능형 CCTV의 핵심기술인 딥러닝 기술을 적용하기 위해서는 많은 수의 데이터 확보가 필수적이다. 쉽게 모을 수 있는 데이터가 있지만 모으기가 굉장히 어려운 데이터도 존재한다. 적은 양의 데이터는 딥러닝 모델의 정확도를 떨어뜨리는 원인이 된다. 따라서 딥러닝 연구 분야 중 Data Augmentation 분야는 중요한 주제이다.

Data Augmentation 분야는 다른 분야에 비해 소홀한 경향이 있었다. 그동안 새로운 자료를 수집하지 않고 다양성을 늘리는 Data Augmentation 방법을 이용해왔다. 원 이미지를 변형하여 이미지 복사본을 만들어 활용했는데 이런 방식은 GAN이 제대로 된 합성 이미지를 만들어 내지 못하고 이미지 변형을 모방 학습하는 상황이 초래될 수 있다. 본 논문에서는 이러한 문제를 해결하기 위해서 Crossing GAN을 활용하여 적은 양의 이상행동 데이터를 Data Augmentation 하여 상황 분석 CCTV 모델의

인식 정확도를 높이고자 한다. 기존 데이터 증강 전 데이터를 학습시킨 모델과 GAN을 이용해 Data Augmentation을 진행한 모델을 비교 분석한다.

### 2. 관련 연구

#### 2.1 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.

##### 2.1.1 Open Pose

Open Pose는 real-time multi-person 2D pose estimation 시스템으로, 주어진 이미지 속 여러 사람의 신체 정보를 특정하는 135개의 키 포인트를 검출한다. Human detection을 한 후 regression을 통해 관절을 찾는 top-down 방식의 Deep Pose와는 반대로 Open Pose는 찾은 관절을 이어 사람을 만드는 bottom-up 접근법을 사용한다. 다시 말해, Open Pose는 다중 CNN 네트워크를 거쳐 생성한 Confidence Map과 Part Affinity Fields로 팔꿈치, 목, 등의 신체 부위를 검출한 후 greedy bipartite matching 알고리즘으로 pose estimation을 한다. 이러한 bottom-up 접근 방식은 이미지 안에 존재하는 사람의 수와 상관없이 높은 estimation 정확성과 real-time 성능을 향상할 수 있다.

### 2.1.2 Alpha Pose

Alpha Pose는 Open Pose와는 반대로 이미지 속 사람을 먼저 찾은 후 pose estimation을 하는 top-down 혹은 Two-step Framework를 고수한다. Top-down 방식이라 하더라도 Object Detection과 Single Person Pose Estimator(SPPE)의 성능에 따라 좋은 결과를 얻을 수 있다. Alpha Pose에서는 Symmetric Spatial Transformer Network(SSTN)로 여러 bounding box에서 정확한 single person 영역을 추출한 후 SPPE가 estimate한 사람 동작을 Spatial De-Transformer Network (SDTN)으로 원본 이미지의 좌표 공간으로 remap한다. Alpha Pose의 주요 장점은 person detection 알고리즘을 교체할 수 있다는 점이다. 주어진 환경과 data에 따라 Human-ReID, PoseFlow, 등의 다양한 tracker를 사용할 수 있다.

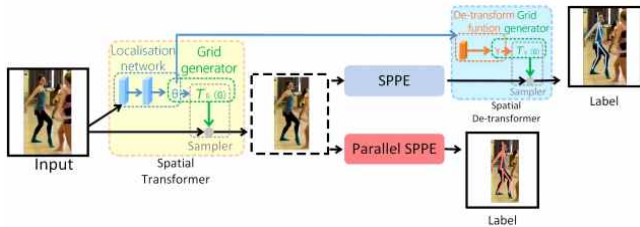


그림 1. Alpha Pose의 STN & SPPE 과정

## 2.2 GAN, Person Image Generation

GAN은 합성된 영상이나 음성 등의 data를 생성한다. GAN은 가짜 데이터를 만드는 generator와 가짜 데이터를 검출하고자 하는 discriminator로 이루어져 있다. 이들은 적대적 학습을 거쳐서 discriminator를 속일 수 있는 generator가 만들어진다.

### 2.2.1 C2 GAN

Cycle In Cycle (C2 GAN)은 key point-guide image generation을 하나의 network 안의 두 개의 다른 image generator에서 학습한다는 특징이 있다. 두 generator는 한 개의 image 기반 cycle과 두 개의 key point 기반 cycle로 연결되어 있다. Cycle들은 end-to-end 방식으로 연결되어 있으므로 generator들은 cross-modal data와 parameter를 주고받으며 같이 학습한다.

### 2.2.2 Crossing GAN

Crossing GAN은 person image generation에 특화된 GAN이다. Crossing GAN은 shape과 appearance feature들의 joint influence를 얻기 위해 두 가지 branch로 이루어진 generator를 사용한다. 이미지를 입력받아 appearance code를 생성하는 Shape-guided

Appearance-based generation (SA) branch, 그리고 pose data를 입력받아 shape code를 생성하는 Appearance-guided Shape-based generation(AS) branch 두 종류이다. 두 branch는 각자의 단계(block)를 거치며 crossing operation을 수행한다. 이때 image modality와 pose modality는 서로 joint influence를 받는다. Co-attention fusion module (CFM)은 양 branch의 최종 block에서 받은 appearance와 shape code를 합쳐 사람 이미지를 생성한다. 생성된 이미지에 대하여, input 이미지 속 사람과의 유사도, 그리고 targeted pose와의 일치 여부를 검사하기 위해 discriminator 또한 두 가지가 존재한다.

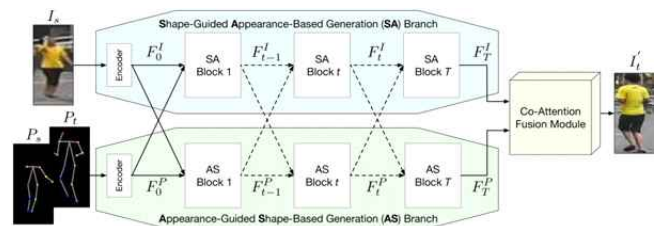


그림 2. Crossing GAN Generator 구조

## 3. 진행 과정

### 3.1 Data Preparation

GAN을 학습시키기 위한 Market1501 Dataset, Accident data를 추출하기 위한 AI HUB, Avenue Dataset 을 준비했다. Pose estimation data는 Open Pose와 Alpha Pose를 비교해서 둘 중 검출률이 높은 estimation 방식을 사용한다.



그림 3. AI HUB Fall 데이터 예시

### 3.2 Model Preparation

C2, Crossing, 등의 GAN을 사용한다.

### 3.3 Real World Application

GAN으로 augment 한 사고 data를 평가할 Evaluation System(CCTV Program with AI)을 개발한다. GAN 별 score를 비교 분석한다.

#### 4. 결론 및 향후 연구

본 논문에서는 지능형 CCTV의 이상행동 인식률을 높이기 위해 기존의 적은 양의 이상행동 데이터를 Open Pose를 사용하여 이미지 속 여러 사람의 신체 정보를 특정하여 Pose Estimation 한 후 Crossing GAN을 이용하여 Data Augmentation을 제안하였다. 본 논문에서 제안하는 방식을 활용하여 Data Augmentation을 진행하고 증강된 데이터를 지능형 CCTV 모델에 학습시키면 기존 원본 데이터를 사용했을 때보다 향상된 상황 인식률을 얻을 수 있을 것이다.

#### 5. 참고문헌

- [1] Ngoc-Trung Tran, Viet-Hung Tran, Ngoc-Bao Nguyen, Trung-Kien Nguyen, Ngai-Man Cheung. "On Data Augmentation for GAN Training", In IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 30, 2021
- [2] Ian J. Goodfellow et al. "Generative Adversarial Nets", In Advances in Neural Information Processing System. (NIPS), pp.2672-2680, 2014
- [3] Zhe Cao et al. "OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields", arXiv:1812.08008v2 [cs.CV] 30 May 2019
- [4] Hao-Shu Fang et al. "RMPE: Regional Multi-Person Pose Estimation", arXiv:1612.00137v5 [cs.CV] 4 Feb 2018
- [5] Hao Tang et al. "Cycle In Cycle Generative Adversarial Networks for Keypoint-Guided Image Generation", arXiv:1908.00999v3 [cs.CV] 16 Apr 2020
- [6] Hao Tang et al. "XingGAN for Person Image Generation", arXiv:2007.09278v1 [cs.CV] 17 Jul 2020