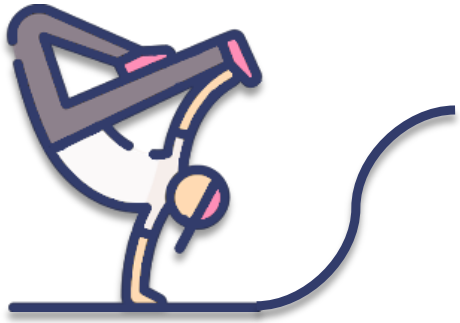


행동 추정을 통한 실시간 이상행동 탐지

마민정, 조문주, 조예서, 지윤혁

목차

행동 추정을 통한
실시간 이상행동 탐지

1. 주제 선정 배경
연구 동기 및 중요성
UCF CRIME DATASET

2. 기존 연구
선행 연구 조사

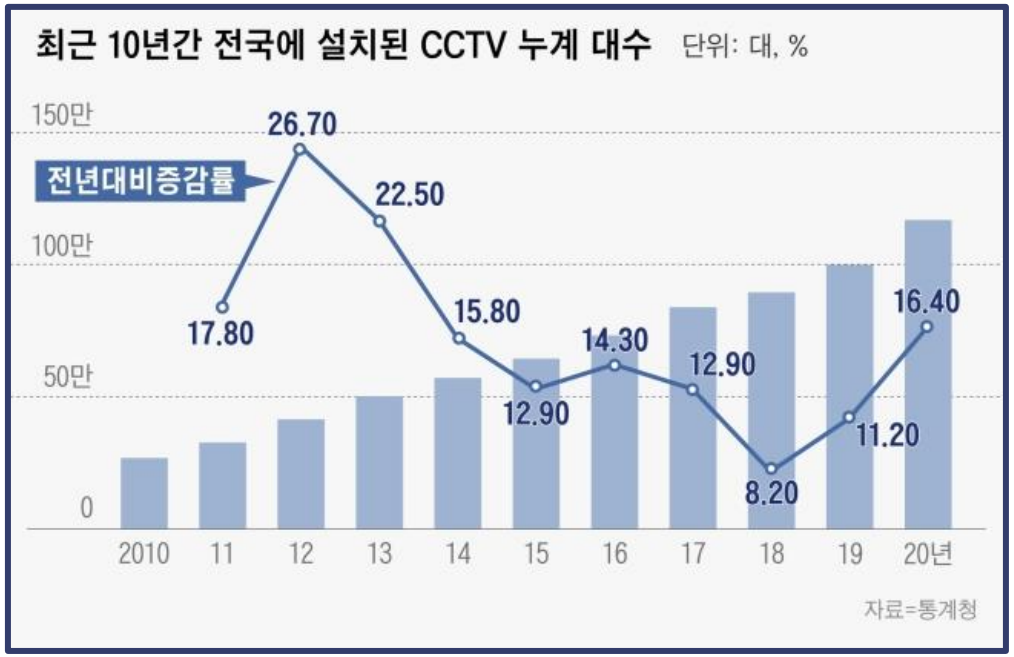
3. 전처리
데이터셋 전처리

4. 제안 방법론
문제 정의 및 가정
모델 파이프라인

5. 실험 및 결과
실험 시나리오 설계
새로운 방법론 제안

6. 결론
사용한 방법론 요약
개선 및 제언 방향

주제 선정 배경 연구 동기 및 중요성



출처 : [경찰, 과학과 만나다]② '전국 130만 CCTV' 경찰의 눈이 되다(조선비즈)

1. CCTV 설치 지속적으로 증가

24시간 동안 사람이 CCTV를 보고 모두 대응하는 것은 불가능

2. 해당 연구를 통해 비정상 케이스 탐지 자동화

사고 예방, 단시간 적절한 대응 가능

→ 효율적인 인력 관리 가능

3. 다양한 사회적 분야에도 활용 가능

단순히 범죄 상황을 넘어 노인 고독사 방지 등 다양한 분야에서 활용 가능

AI가 자동으로 움직임을 포착해서 이상행동을 감지하는 모델을 만들어보자!

주제 선정 배경 연구 동기 및 중요성

[다양한 사회적 분야의 연구 가치]

1. 치매, 독거노인 등 사회적 약자 감시 기능

사회적으로 돌봄이 필요한 사회적 약자의 감시를 통해 응급상황 해결 가능

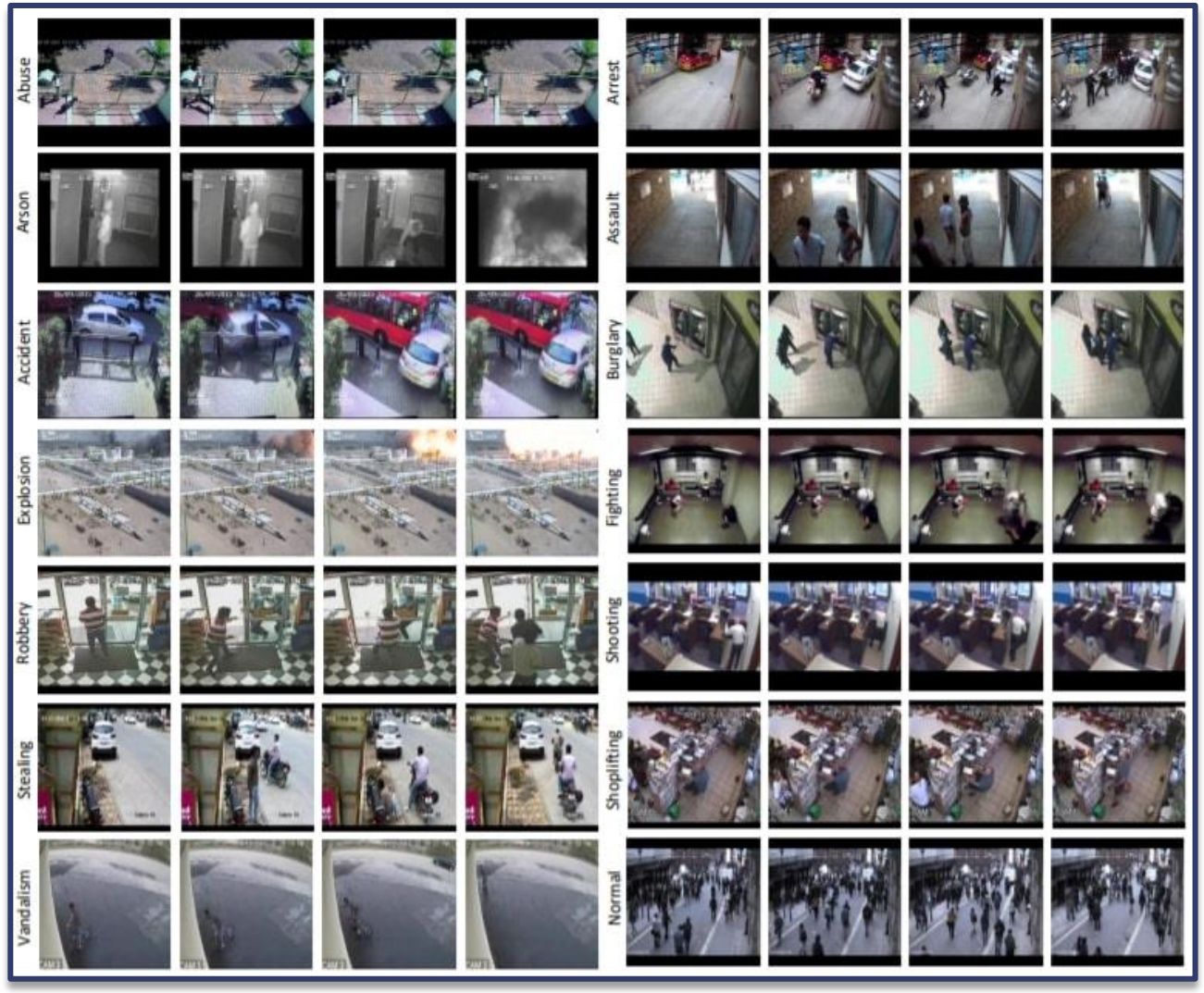
2. CCTV가 닿는 곳이면 어디든 범죄 예방과 억제

실시간으로 이상행동 탐지가 가능해지면 범죄 예방 및 억제의 가능

3. 실시간 알람을 통한 경찰 인력관리 효율화

범죄에 실시간으로 대응 및 인력을 효율적으로 관리 가능

주제 선정 배경 UCF CRIME DATASET



[UCF CRIME DATASET]

- 정제되지 않은 실제 세계의 감시 비디오

128시간, 1900개

- 13개의 이상 징후와 1개의 정상 징후

학대, 체포, 방화, 폭행, 도로 사고, 강도(Burglary), 폭발, 격투, 강도(Robbery), 총격, 절도, 들치기, 반달리즘

- 이상치 판단을 위한 우리의 목적에 적합

학습하기에 데이터의 사이즈가 충분히 크고 라벨링이 되어 있어 학습에 용이

기존 연구 선행 연구 조사

- Abnormal Situation Detection on Surveillance Video Using Object Detection and Action Recognition

인용 분야	인용한 기법	한계점 및 차이점
Preprocessing	Yolov3(객체 탐지 모델), Openpose(2D 자세 추정), p-LSTMs(2D→3D 자세추정)	객체탐지 분야에서 SOTA 모델인 PP-YOLO 적용 Openpose에 PoseFix(refinement:후처리)기법을 적용
Modeling (Pose Estimation)	스켈레톤 관절들 간의 내부적인 관계와 객체 및 인접 사람들과의 외부적인 관계를 학습, 신체부위를 그룹화하여 독립적으로 학습	전체적인 그림을 보기에 부족 → 신체 부위 전반을 함께 학습할 수 있는 AGCN 적용
Modeling (Anomaly Detection)	4가지 Mapping 기법 사용 (Action, Relation, Distance, Object Map) CONV 연산을 통한 이진분류 (Conv, ROIPooling, Flatten, FC Layer)	기존 연구는 사물정보가 없을 경우 탐지를 잘 못함 → Action/Object Map에 조건에 따른 가중치 추가 기존 연구는 DownSampling 과정에서 공간정보 손실 → ROIAlign, FCN(Fully Convolutional Network)를 통해 위치 정보를 보존하면서 차원 축소

기존 연구 선행 연구 조사

- Ensemble Deep Learning for Skeleton-Based Action Recognition Using Temporal Sliding LSTM Networks

인용 분야	인용한 기법	한계점 및 차이점
Modeling : TS-LSTM (Pose Estimation)	스켈레톤 일반화. 속도를 추출한 모션정보를 다양한 시간 간격으로 학습	위치에 대해 동적인 특성으로 인해 관절 위치 변화만으로 정확한 인식 힘듦

- Spatial Temporal Graph Convolutional Network for Skeleton based Action Recognition

인용 분야	인용한 기법	한계점 및 차이점
Modeling : ST-GCN (Pose Estimation)	스켈레톤 하나를 그래프 구조로 간주, 그래프 구조가 시간적 정보 포함, 스켈레톤 관절들 간 관계성을 식별하고 행동 인식	인식 가능한 행동 수 제한, 관절 간 관계성 로컬 영역에서만 찾을 수 있음

- Actional-structural Graph Convolutional Networks for Skeleton-based Action Recognition

인용 분야	인용한 기법	한계점 및 차이점
Modeling : AS-GCN (Pose Estimation)	두 계층의 신경망을 구성하여 모든 관절들 간 관계성 식별	멀리 떨어진 관절들 사이의 관계성을 담기는 어려울 수 있음

→ 모든 관절들 간 관계성을 식별할 수 있으며 attention의 효과를 볼 수 있는 AGCN 적용

기존 연구 선행 연구 조사

- Real-World Anomaly Detection in Surveillance Videos

인용 분야	인용한 기법	한계점 및 차이점
Modeling : C3D (Anomaly Detection)	입력된 영상에서 시간 축에 따른 픽셀의 변화량을 기반으로 모션정보와 모양정보를 추출	프레임의 전체 픽셀에 대한 모션정보를 추출하므로 모션정보가 비정상적인 상황을 유발한 주체의 모션정보가 아닐 수 있음

- Motion-Attentive Network for Detecting Abnormal Situations in Surveillance Video

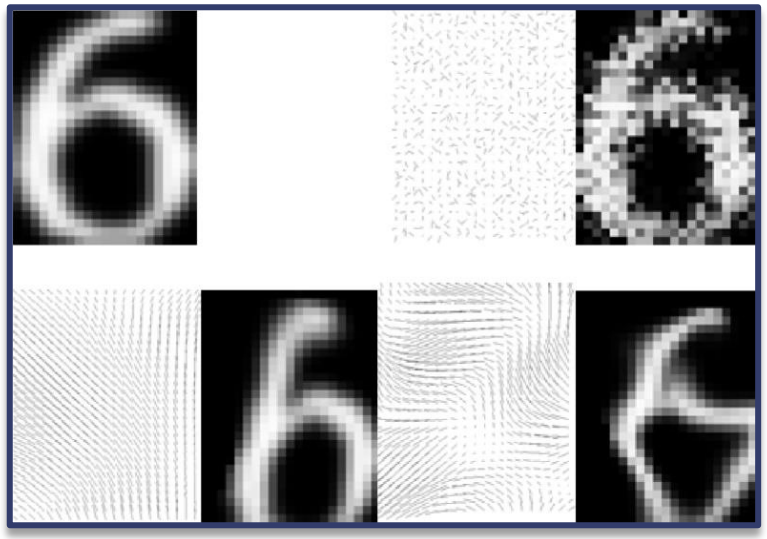
인용 분야	인용한 기법	한계점 및 차이점
Modeling (Anomaly Detection)	모션 정보를 추출하기 전 영상 속에 사람이 위치한 관심영역(ROI)에 해당하는 모션정보만을 추출	관심영역 외의 정보를 고려하지 않아 외부요인과의 관계를 고려할 수 없음

- Human Activity Recognition for Surveillance Applications

인용 분야	인용한 기법	한계점 및 차이점
Modeling : HMM (Anomaly Detection)	오토인코더를 사용하여 사람 스켈레톤의 궤적에 대한 시공간 패턴을 나타내는 특징을 추출하여 패턴의 규칙성에 부합하지 않는 스켈레톤 탐지	스켈레톤 이외의 정보를 활용하지 않으므로 시간 축에 따른 스켈레톤의 변화가 정상인 경우와 유사하거나 그 반대의 경우에 비정상적인 상황 탐지의 오탐율이 증가

→ 객체가 없어도 오탐율이 낮고 픽셀별로 위치 정보를 잘 탐지할 수 있는 AT-Net 적용(자체적으로 구조를 변경)

전처리 데이터셋 전처리



- Data Augmentation

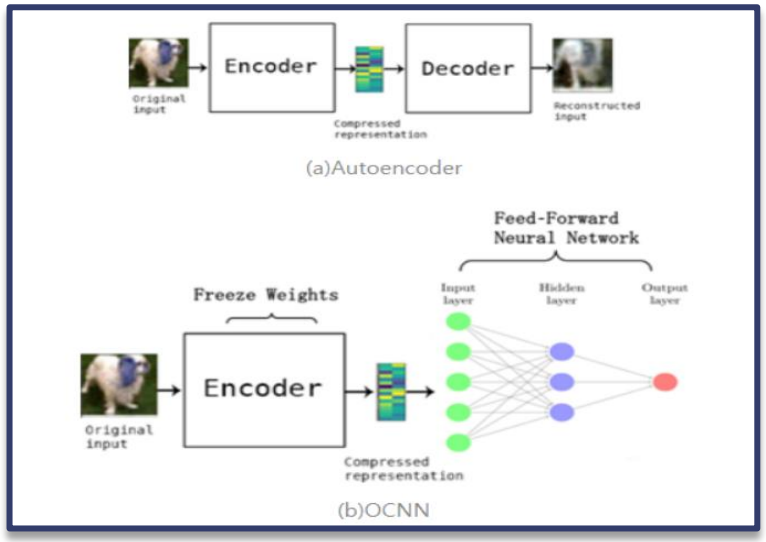
Class Imbalance 문제 해결

(Addition of Random Noise, Elastic Deformation)

- Background Subtraction

관심 영역만 더 높은 해상도로 명확하게 보여주는 기법

각 segment에서 세부적인 관심 영역을 찾는 것이 이상 행동 탐지에 도움이 됨



- OCNN

특정 데이터의 이상치 탐지에 적절한 특징을 학습

제안 방법론 문제 정의 및 가정

[용어 정리]

- GCAN

Graphical Convolutional Anomaly Network의 약자로 우리가 정의한 모델명

[목적함수]

- AGCN) Cross Entropy Loss
- AT-Net) Cross Entropy Loss

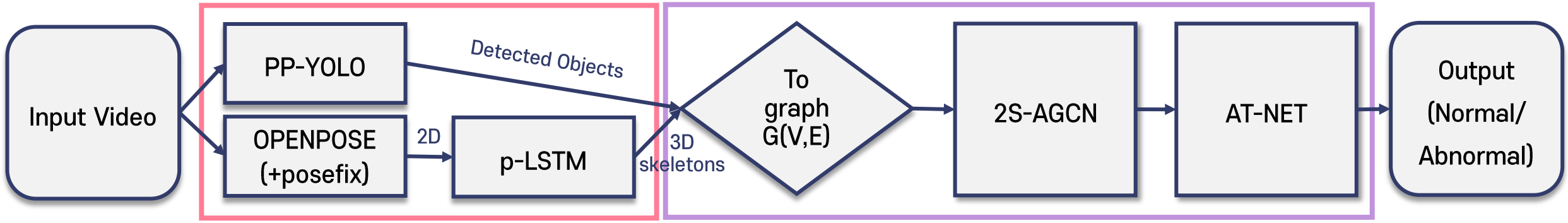
[평가지표]

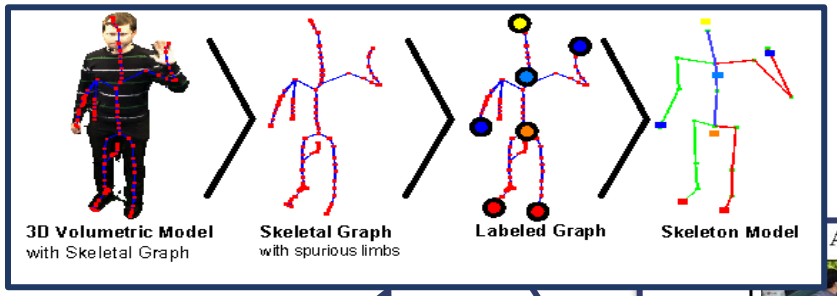
F1.5 Score

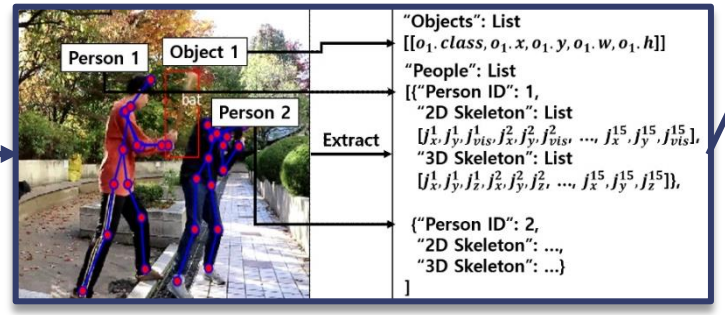
[문제 가정]

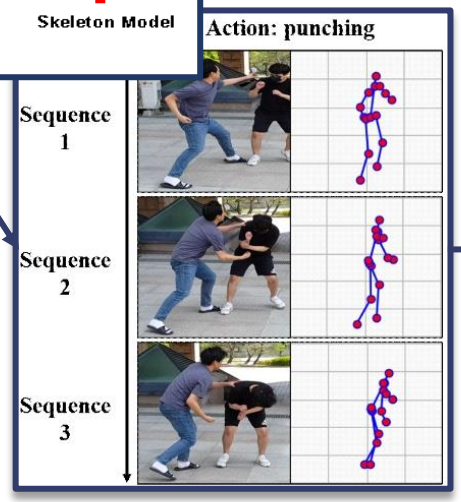
1. AGCN을 활용할 경우 보다 많은 공간상 관계성과 정보를 담을 수 있을 것이라 가정
2. ROIAlign을 하면 픽셀을 좀 더 세분화해서 정확하게 객체를 추출할 것이라 가정
3. FCN을 통해 DownSampling시 위치정보를 보존하면서 차원 축소가 되고, 연산량도 줄어들 것이라 가정

제안 방법론 모델 파이프라인









제안 방법론 모델 파이프라인

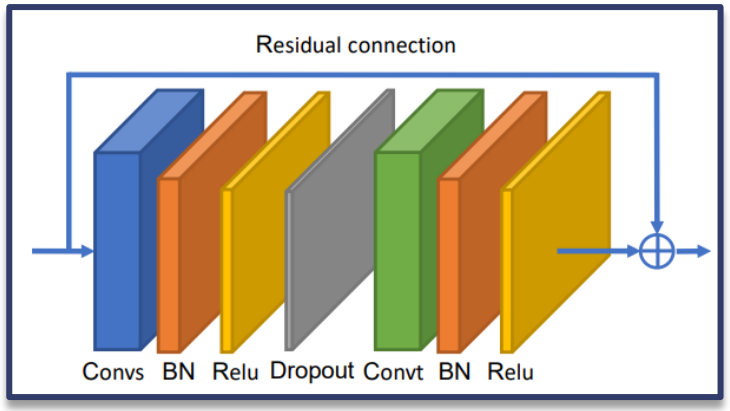
[AGCN 주요 Contribution]

$$\mathbf{f}_{out} = \sum_{k}^{K_v} \mathbf{W}_k \mathbf{f}_{in} (\mathbf{A}_k + \mathbf{B}_k + \mathbf{C}_k)$$

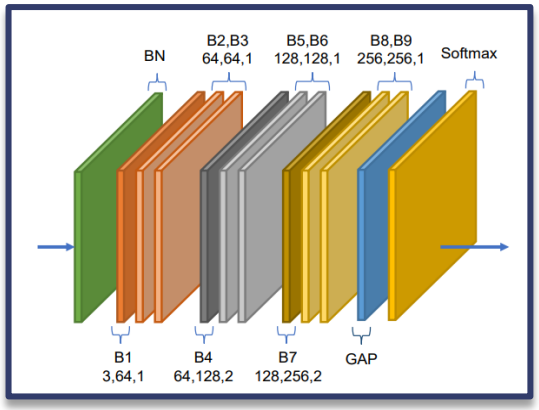
Ak : 기존 GCN의 인접행렬과 동일, 바로 옆 관절
Bk : 학습을 통해 인접한 관절 결정, attention 효과
Ck : 데이터에 의존적, 각 sample에 대한 독특한 그래프 학습

- Adaptive Graph Convolutional Network
- 사용자나 원자의 속성, 연결의 종류 등을 고려해야 하는 경우에는 그래프 데이터로 나타내면 공간상 관계성과 정보를 담을 수 있음

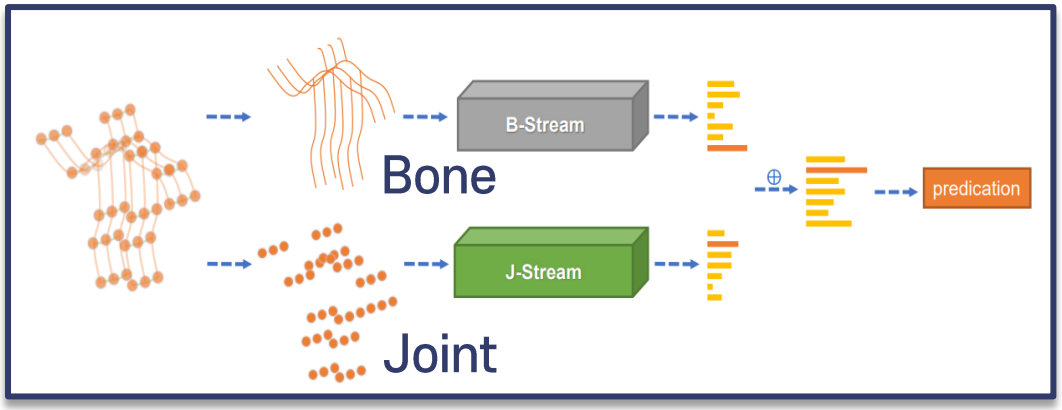
< AGC block >



< AGCN >

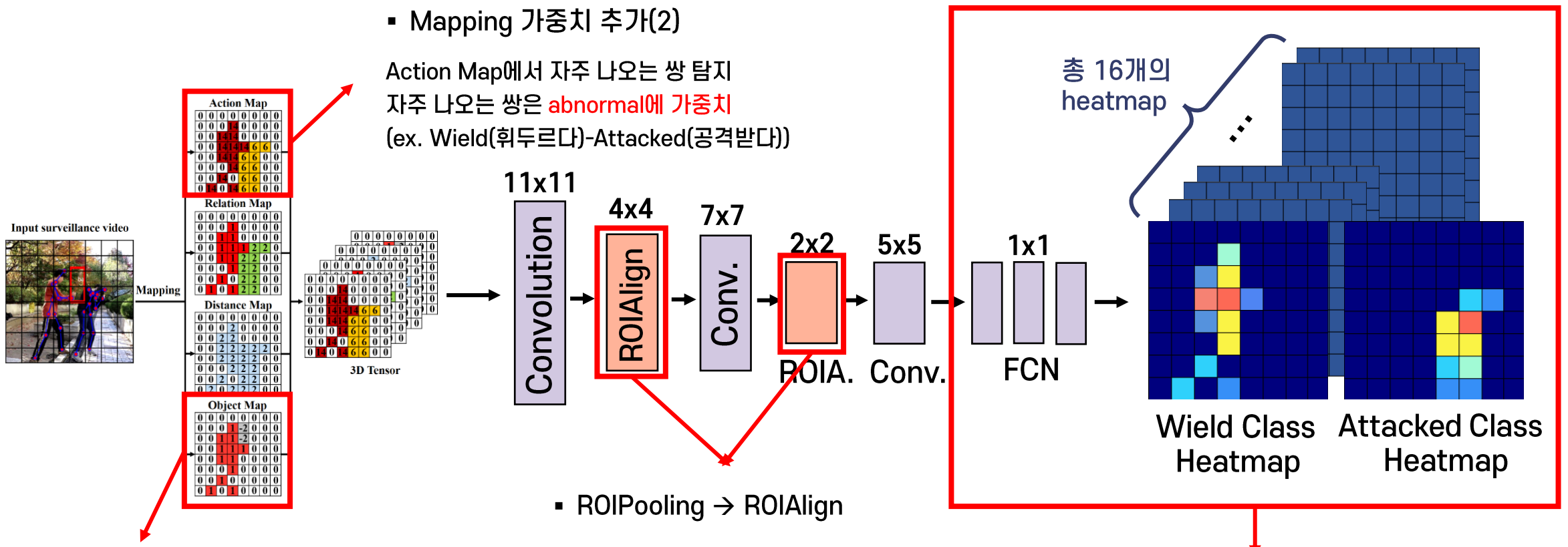


< 2S-AGCN >



제안 방법론 모델 파이프라인

[AT-Net 주요 Contribution]



▪ Mapping 가중치 추가(2)

Action Map에서 자주 나오는 쌍 탐지
자주 나오는 쌍은 abnormal에 가중치
(ex. Wield(휘두르다)-Attacked(공격받다))

▪ Mapping 가중치 추가(1)

Object Map에서 사물이 감지되는 경우
abnormal에 가중치 (ex. Theft, Stabbing)

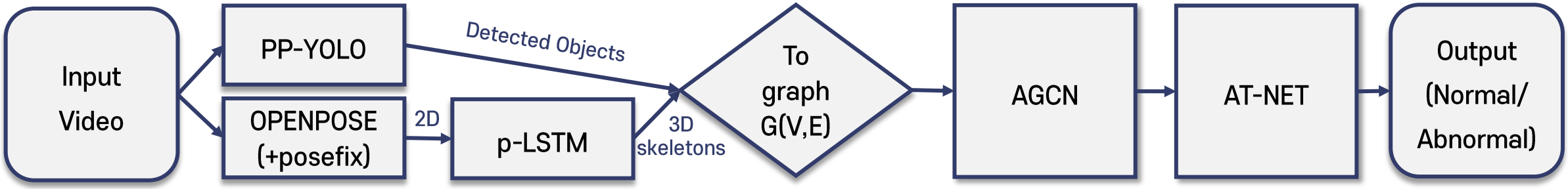
▪ ROIPooling → ROIAlign

ROI가 소수점 좌표를 가지고 있을 가능성
→ Max Pooling 과정에서 픽셀 단위로
좀 더 정확한 위치정보를 가지며
downsampling

▪ FC Layer → FCN(Fully Convolutional Network)

Downsampling 시 위치정보를 보존하면서 차원 축소, 연산량 감소
다양한 크기의 Input Image Size가 들어갈 수 있음

실험 및 결과 실험 시나리오 설계



분야	제안 기법	실험 시나리오
Preprocessing	Back-Subtraction	(1) 적용하지 않은 경우 (2) Back-Subtraction을 적용한 경우
Preprocessing	OCNN	(1) 적용하지 않은 경우 (2) OCNN을 적용한 경우
Preprocessing	PoseFix	(1) 적용하지 않은 경우 (2) PoseFix를 적용한 경우
Modeling	PP-YOLO	(1) YOLO V3 적용한 경우 (2) YOLO V4 적용한 경우 (3) PP-YOLO 적용한 경우

실험 및 결과 새로운 방법론 제안

[Preprocessing : Back-Subtraction 적용]

출처 : 'Deep anomaly detection through visual attention in surveillance videos'

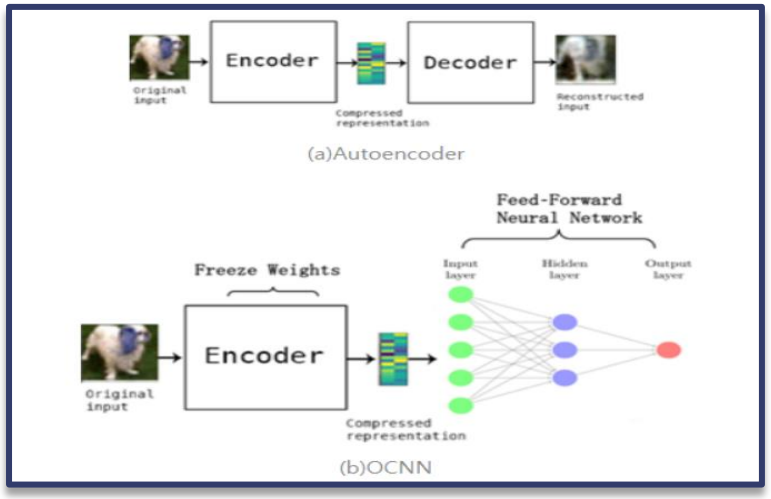
Method	Dataset		
	Robbery_106	Fighting_042	Road accident_022
Sultani et al. [6]	*99.9*	77.7	35.63
Proposed method	96.5	*98.2*	*92.53*

< UCF-Crime 데이터셋의 정확도 비교>

배경을 제거하고 이상 지역에 집중하는 방식

강도, 싸움, 교통사고와 같은 다양한 종류의 사건에서 높은 정확도를 확인

[Preprocessing : OCNN 적용]



Object-based Convolutional Neural Network

1. 학습된 Auto-Encoder로 특징을 추출
2. 전이 학습(Transfer-Learning)을 이용해 재학습

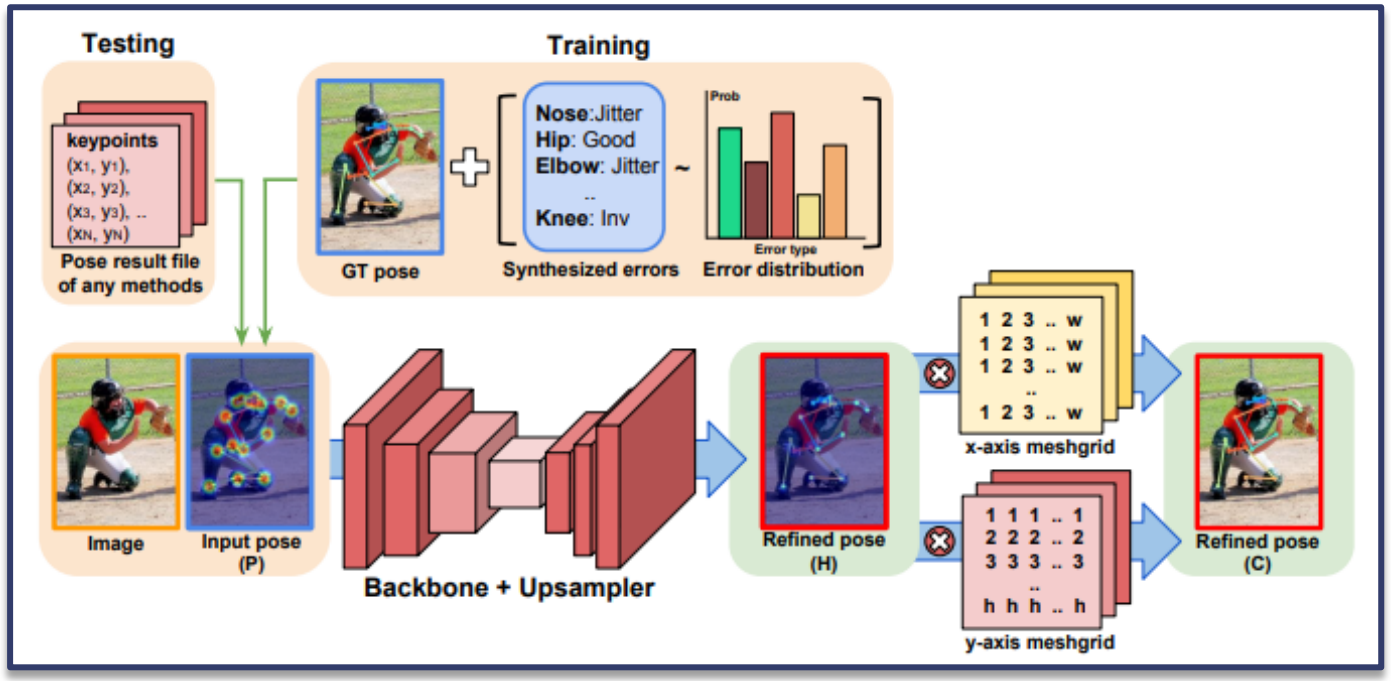
→ 특정 데이터의 이상치 탐지에 적합한 특징을 학습 가능

실험 및 결과 새로운 방법론 제안

[Preprocessing : PoseFix 추가]

기존 시스템의 에러를 줄이는 후처리시스템

Openpose의 input으로 활용하기 전 정제과정을 통해 성능 향상



Methods	AP	AP_{50}	AP_{75}	AP_M	AP_L	AR	AR_{50}	AR_{75}	AR_M	AR_L
AE [20]	56.6	81.7	62.1	48.1	69.4	62.5	84.9	67.2	52.2	76.5
+ PoseFix (Ours)	**63.9**	**83.6**	**70.0**	**56.9**	**73.7**	**69.1**	**86.6**	**74.2**	**61.1**	**79.9**
PAFs [5]	61.7	84.9	67.4	57.1	68.1	66.5	87.2	71.7	60.5	74.6
+ PoseFix (Ours)	**66.7**	**85.7**	**72.9**	**62.9**	**72.3**	**71.3**	**88.0**	**76.7**	**66.3**	**78.1**
Mask R-CNN (ResNet-50) [11]	62.9	87.1	68.9	57.6	71.3	69.7	91.3	75.1	63.9	77.6
+ PoseFix (Ours)	**67.2**	**88.0**	**73.5**	**62.5**	**75.1**	**74.0**	**92.2**	**79.6**	**68.8**	**81.1**
Mask R-CNN (ResNet-101)	63.4	87.5	69.4	57.8	72.0	70.2	91.8	75.6	64.3	78.2
+ PoseFix (Ours)	**67.5**	**88.4**	**73.8**	**62.6**	**75.5**	**74.3**	**92.6**	**79.9**	**69.1**	**81.4**
Mask R-CNN (ResNeXt-101-64)	64.9	88.6	71.0	59.6	73.3	71.4	92.4	76.8	65.9	78.9
+ PoseFix (Ours)	**68.7**	**89.3**	**75.2**	**64.1**	**76.4**	**75.2**	**93.1**	**80.9**	**70.3**	**81.9**
Mask R-CNN (ResNeXt-101-32)	64.9	88.4	70.9	59.5	73.2	71.3	92.2	76.7	65.8	78.9
+ PoseFix (Ours)	**68.5**	**88.9**	**75.0**	**64.0**	**76.2**	**75.0**	**92.9**	**80.7**	**70.1**	**81.8**
IntegralPose	66.3	87.6	72.9	62.7	72.7	73.2	91.8	79.1	68.3	79.8
+ PoseFix (Ours)	**69.5**	**88.3**	**75.9**	**65.7**	**76.1**	**75.9**	**92.4**	**81.8**	**71.1**	**82.5**
CPN (ResNet-50) [7]	68.6	89.6	76.7	65.3	74.6	75.6	93.7	82.6	70.8	82.0
+ PoseFix (Ours)	**71.8**	**89.8**	**78.9**	**68.3**	**78.1**	**78.2**	**93.9**	**84.3**	**73.5**	**84.6**
CPN (ResNet-101)	69.6	89.9	77.6	66.3	75.6	76.6	93.9	83.5	72.0	82.9
+ PoseFix (Ours)	**72.6**	**90.2**	**79.7**	**69.0**	**78.9**	**78.9**	**94.1**	**85.0**	**74.2**	**85.1**
Simple (ResNet-50) [30]	69.4	90.1	77.4	66.2	75.5	75.1	93.9	82.4	70.8	81.0
+ PoseFix (Ours)	**72.5**	**90.5**	**79.6**	**68.9**	**79.0**	**78.0**	**94.1**	**84.4**	**73.4**	**84.1**
Simple (ResNet-101)	70.5	90.7	78.8	67.5	76.3	76.2	94.3	83.7	72.1	81.9
+ PoseFix (Ours)	**73.3**	**90.8**	**80.7**	**69.8**	**79.8**	**78.7**	**94.4**	**85.3**	**74.3**	**84.8**
Simple (ResNet-152)	71.1	90.7	79.4	68.0	76.9	76.8	**94.4**	84.3	72.6	82.4
+ PoseFix (Ours)	**73.6**	**90.8**	**81.0**	**70.3**	**79.8**	**79.0**	**94.4**	**85.7**	**74.8**	**84.9**

Table 2: Improvement of APs when the PoseFix is applied to the state-of-the-art methods. The APs are calculated on the test-dev set.

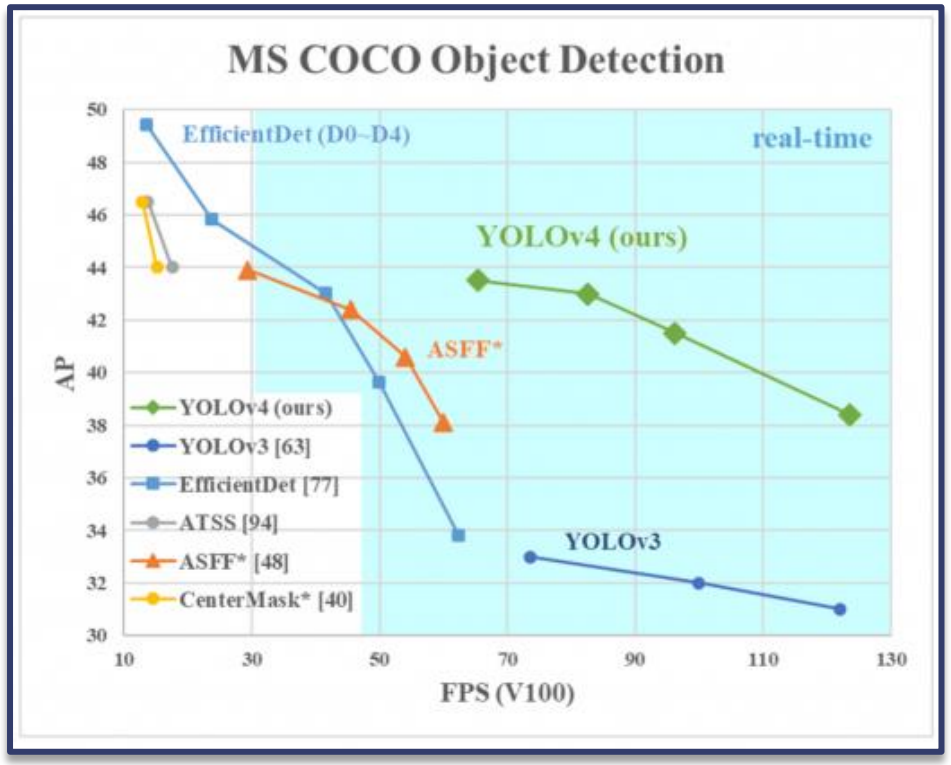
입력 이미지에 있는 모든 사람의 human body keypoints의 2D skeleton을 수정하여 정확성을 높임

실험 및 결과 새로운 방법론 제안

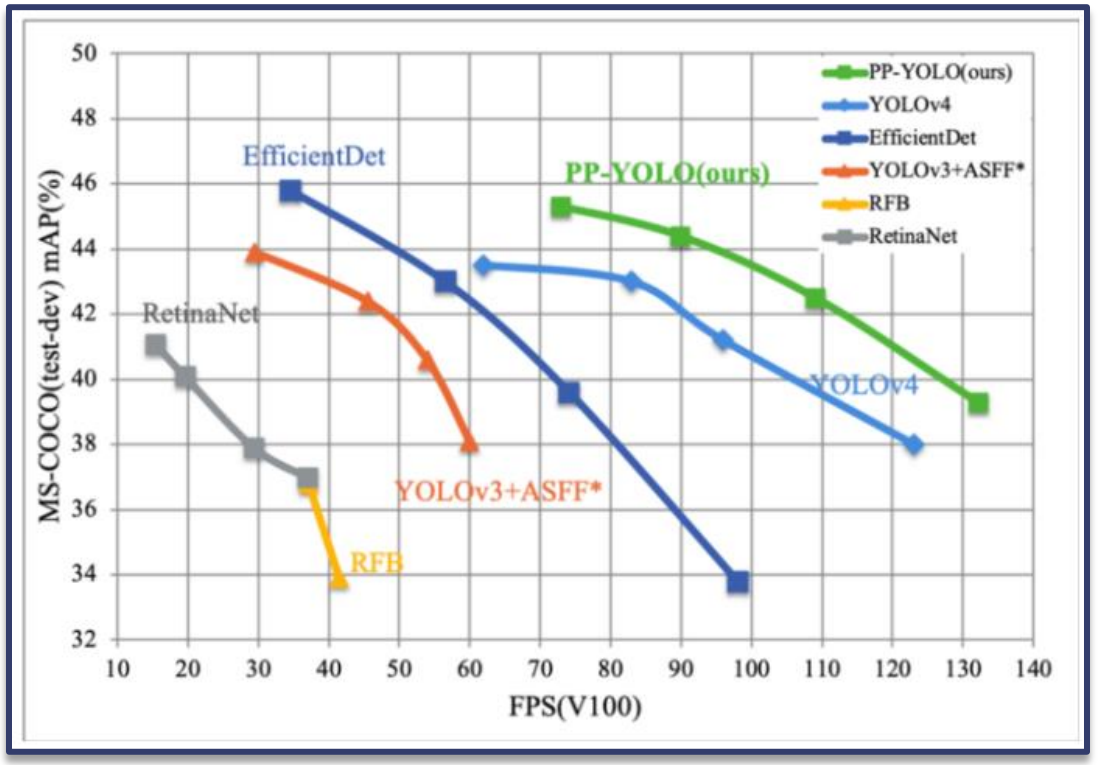
[Modeling : PP-YOLO 대체]

PP-YOLO는 YOLO-V3에 기초한 물체 검출기

속도 변화가 거의 없도록 하면서 검출기의 정확도를 최대한 향상



< YOLO-V3 보다 좋은 YOLO-V4의 성능 >



< YOLO-V4의 보다 좋은 PP-YOLO의 성능 >

실험 및 결과 새로운 방법론 제안

[Modeling : PP-YOLO 대체]

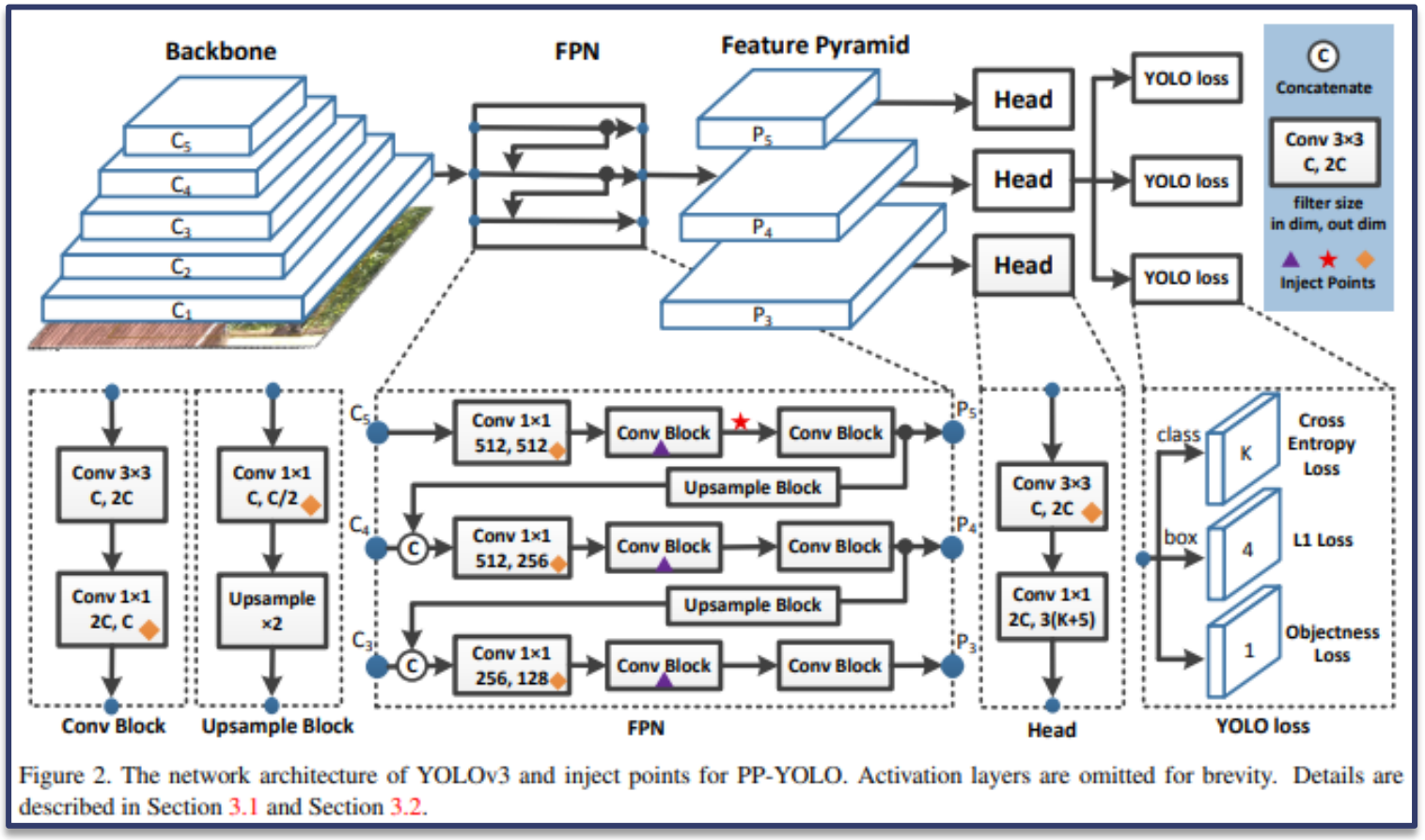


Figure 2. The network architecture of YOLOv3 and inject points for PP-YOLO. Activation layers are omitted for brevity. Details are described in Section 3.1 and Section 3.2.

- DarkNet-53 → ResNet50-vd
- batch size 64 → 192
- 이동 평균을 parameter로 사용
- DropBlock가 FPN에 적용
- IoU loss 사용
- IoU prediction branch 사용
- Grid Sensitive 사용
- 매트릭스 NMS 사용
- FPN에 CoordConv 사용
  (1X1 conv layer를 대체)
- Spatial Pyramid Pooling 사용
  (상위 피쳐맵에 사용)

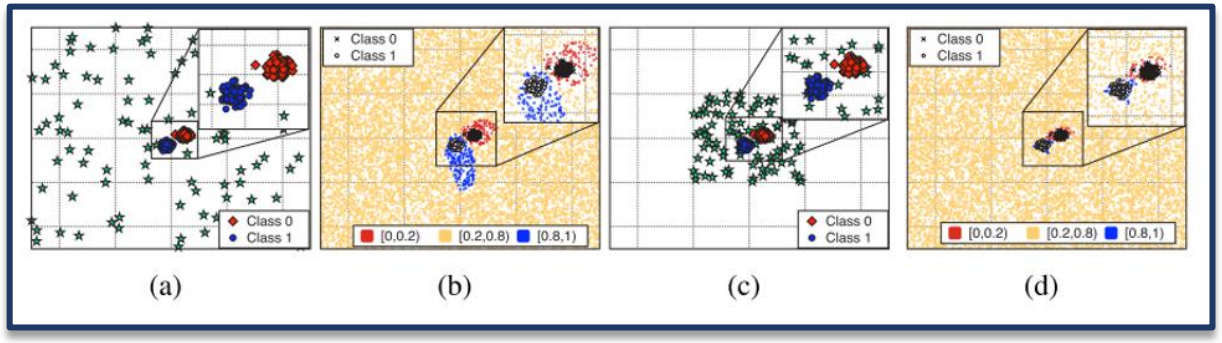
결론 사용한 방법론 요약

[GCAN 요약]

분야	적용/추가/대체한 기법
데이터 전처리 기법	Data Augmentation, Back-Subtraction, OCNN
전처리 모델	PP-YOLO, Openpose(+posefix), p-LSTM
모델링	AGCN, AT-Net(+Weight/ROIAlign/FCN 변경)
목적함수	Cross Entropy Loss
평가지표	F1.5 Score

결론 개선 및 제언 방향

[Modeling : OOD 적용]



< OOD >

- In distribution : 학습 데이터의 분포
- Out of distribution : 학습 데이터의 분포를 따르지 않는 모든 분포

Table 2: Comparison of AUROC values for the OOD detection task. Results as reported by the original references except (a) by Ren et al. (2019), (b) by Lee et al. (2018), and (c) by Choi et al. (2018). Results for Typicality test correspond to using batches of 2 samples of the same type.

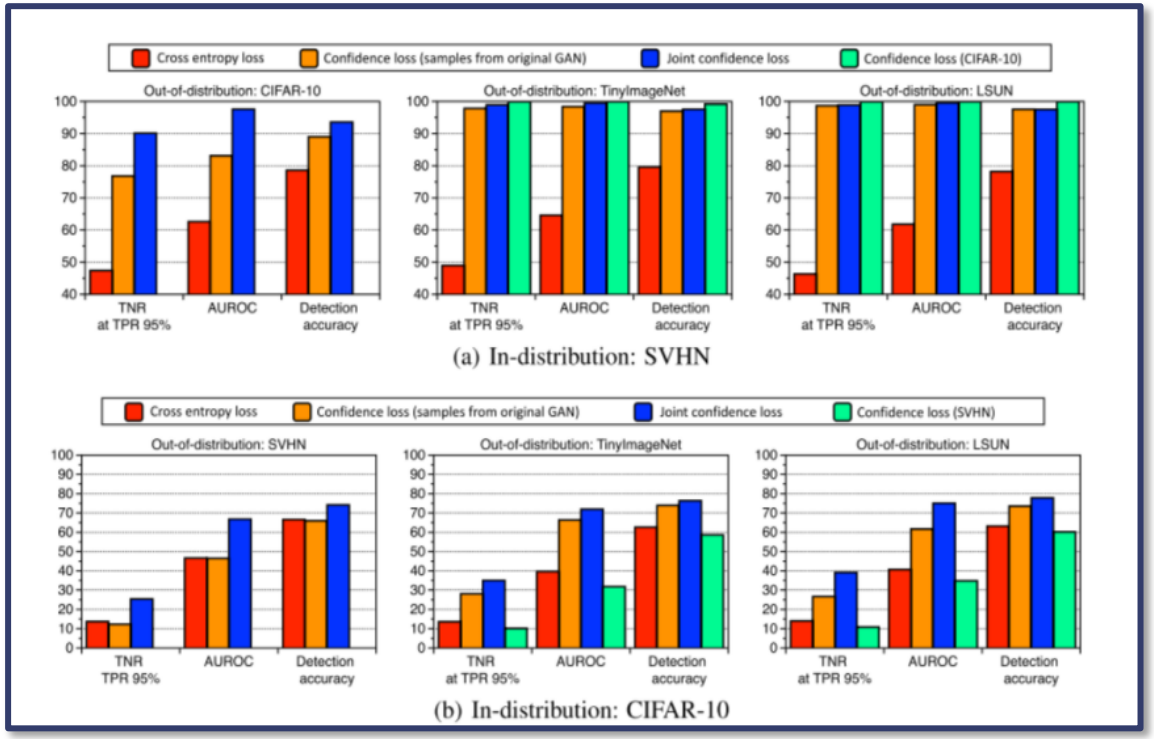
Trained on:	FashionMNIST		CIFAR10		
OOD data:	MNIST	Omniglot	SVHN	CelebA	CIFAR100
Classifier-based approaches					
ODIN (Liang et al., 2018)[a,b]	0.697	-	0.966	-	-
VIB (Alemi et al., 2018)[c]	0.941	0.943	0.528	0.735	-
Mahalanobis (Lee et al., 2018)	0.986	-	0.991	-	-
Outlier exposure (Hendrycks et al., 2019)	-	-	0.984	-	**0.933**
Generative-based approaches					
WAIC (Choi et al., 2018)	0.766	0.796	**1.000**	**0.997**	-
Outlier exposure (Hendrycks et al., 2019)	-	-	0.758	-	0.685
Typicality test (Nalisnick et al., 2019b)	0.140	-	0.420	-	-
Likelihood-ratio (Ren et al., 2019)	0.997	-	0.912	-	-
S using Glow and FLIF (ours)	**0.998**	**1.000**	0.950	0.863	0.736
S using PixelCNN++ and FLIF (ours)	0.967	1.000	0.929	0.776	0.535

< Outlier Exposure >

GAN을 이용해 Out of distribution 데이터 인공적으로 생성하는 기법
GAN의 generator가 Out of distribution sample을 생성하도록 함
test 단계에서 Out-of-distribution 데이터셋은 걸러내는 것이 목표

결론 개선 및 제언 방향

[Loss Func : Confidence Loss + Cross Entropy Loss + GAN Loss 대체]



OOD(Out Of Distribution) 기법 중 ODIN모델의 손실함수를 Cross entropy loss, Confidence loss, GAN loss를 모두 활용했을 때 성능 향상

→ Backbone 모델의 손실함수(Cross entropy loss)도 바꾼다면 검출 성능 향상 기대

결론 참고문헌

☑ 선행 연구 조사

https://url.vet/litc6

https://url.vet/v6r4z

https://url.vet/2990a

https://dl.acm.org/doi/10.1145/3388770.3407442

http://koreascience.or.kr/article/JAKO202106763002085.page?&lang=ko

https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/viewPaper/17135

https://openaccess.thecvf.com/content_iccv_2017/html/Lee_Ensemble_Deep_Learning_ICCV_2017_paper.html

https://www.researchgate.net/publication/277813677_Human_Activity_Recognition_for_Surveillance_Applications

☑ 제안 방법론

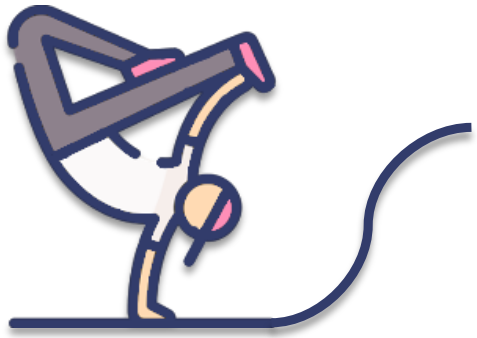
https://arxiv.org/pdf/1812.03595.pdf

https://arxiv.org/pdf/2004.10934.pdf

https://arxiv.org/pdf/2007.12099.pdf

☑ 데이터셋

https://webpages.uncc.edu/cchen62/dataset.html



감사합니다