2023 산학 프로젝트 챌린지

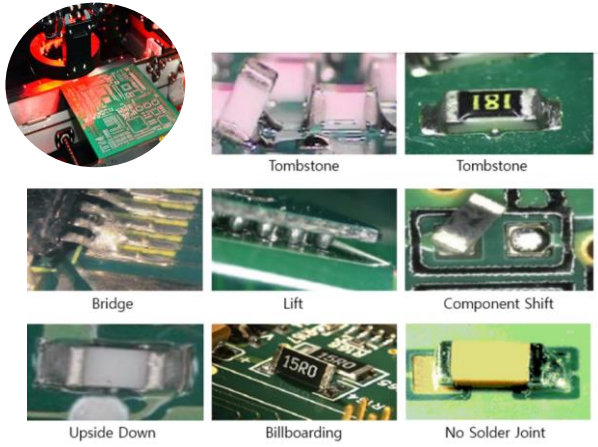
시각-언어 모델을 활용하는 제로샷 이상 감지 방법론

고영테크놀로지의 제조/검증 비용 감소를 위한 영상 인식 모델 도출

참여 기업 : Koh Young Technology
대학 : Soongsil University
팀 : SSU VIP Lab. 석사 김준용, 석사 길다영

Contents

고영테크놀러지 – 영상인식/전자부품 검사 장비 세계시장 선도



2013년 이후 **SPI,AOI(영상비전인식)장비** 세계시장 점유율 1위

700여개 관련 특허기술 보유

전세계 납품 장비 20,000건

반도체 검사 장비, 비전 인식

SPI : 납 도포검사

AOI : 부품장착, 접합 등 3차원 영상 및 비전기반의 검사

영상인식/검사 : 부품의 들뜸, 과납, 미납, 쇼트, 냉납, 소납, 틀어짐 등

사업확장(스마트팩토리) : 공정최적, 생산성, 자동화, 영상인식 알고리즘

최근 국내외 산업현장, 스마트 팩토리의 수요 증가

대형장비 시설 이용 (특히, 해외/반도체/LCD-부품 장비)

고영테크놀러지 – 제조/검증 비용 감소 및 검사 시스템 모니터링 의뢰

검사 과정에서 모든 고장 부품을 학습하고 선별하는 것이 어려움

카테고리 추가 시마다, 새로운 학습 데이터 수집과 재학습이 요구됨

→ **제조/검증비용감소를위한Zero-shot인식방법론연구논의**

고영은 비전 분야에서 이미지 분석 수행

실시간 움직임 모니터링 하는 것이 어려움

→ **제조라인검사시스템을모니터링하고실시간영상인식기술논의**

단순 검사 시스템을 넘어 모니터링 시스템 필요

장비 & 장치산업, 제조 공정상 **안전사고모니터링필요**

1. 연구 개요

Objective

수행 내용

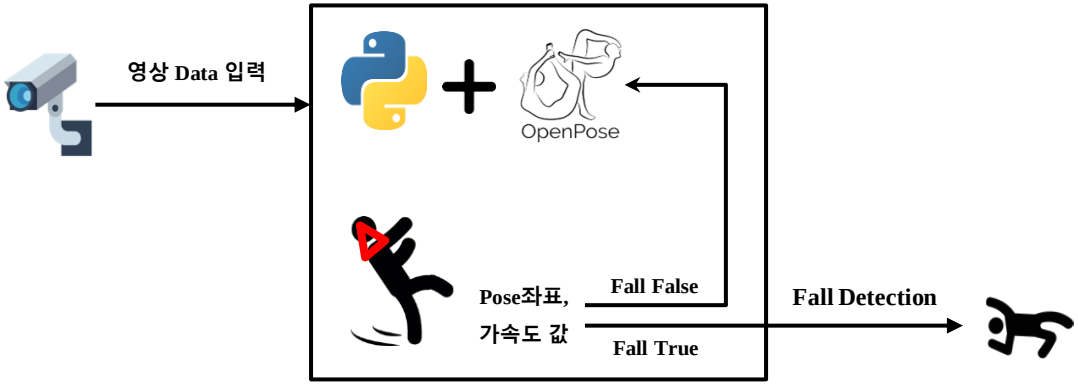
- 시각–언어 모델*을 기반의 Zero-shot*을 이용해 이상 감지라는 일반적인 프레임워크 구축
- **데이터셋 학습 없이 카테고리 확장하여 이상 상황/상태 검출**
- 사례연구로 낙상 감지를 먼저 시도했고, 이후 **다양한 상황에서의 이상 감지로 일반화 가능성을 확인**함

기대 효과

- Unseen 영상 및 고장 카테고리 변경 시 재학습이 필요하지 않기 때문에 **비용 감축**
- 제조 공정에서 **이상 원인*, 우연 원인* 판정이 가능**해짐
- 즉각적인 사고 대응으로 인한 **산업환경 안전** 유지

* **시각-언어 모델 :** 시각 (이미지 또는 비디오)와 언어 (텍스트) 데이터를 모두 이해하고 처리하는 인공 지능 모델
* **Zero-Shot** : 모델이 학습 과정에서 배우지 않은 작업을 수행하는 것 / 학습 데이터에 없는 새로운 클래스를 인식하고 분류
* **이상 원인** : 통계로 예측 가능한 원인 (마모)
* **우연 원인** : 통계로 예측 불가능한 원인 (안전 사고)

Rule-based Fall Detection Algorithm [9]



The architecture of Rule-based Fall Detector.

- **매 프레임마다 OpenPose*를 사용**하여 인간 좌표를 감지해야 함
- 동일한 영상에서도 **카메라의 각도에 따라** 가속도 값이 달라질 수 있음
- 각각의 데이터셋에 맞게 **가속도 임계값을 직접 설정**해야 함
- **이미지나 저해상도 데이터셋**에서는 작동하지 않음

*** OpenPose** : 인간 자세 예측 (Human Pose Estimation)의 한 분야로 신체 관절을 25개의 keypoints로 예측하여 실시간으로 skeleton을 생성함

- [9] 김준용 외 "Fall Detection and Driver Carelessness Detection using Static Image-based Action Recognition," 28(3), Journal of Institute of Control, Robotics and Systems 2022
- 은상, 제 11회 숭실 캡스톤 디자인 경진대회

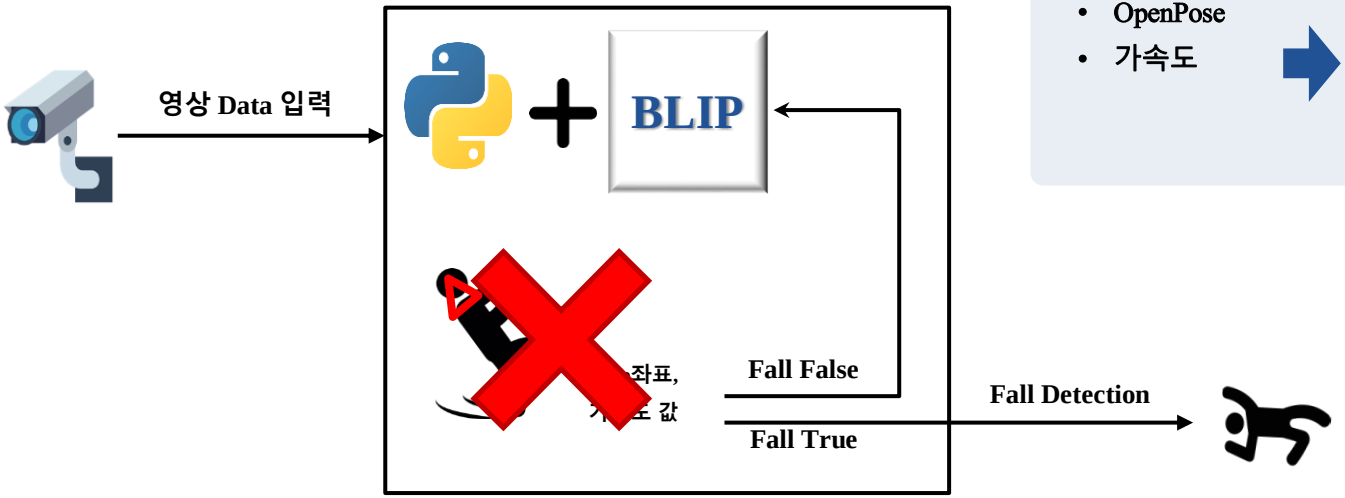
Upgrade Rule-based Fall Detection to Zero-shot Fall Detection

Rule-based

- OpenPose
- 가속도

Zero-shot

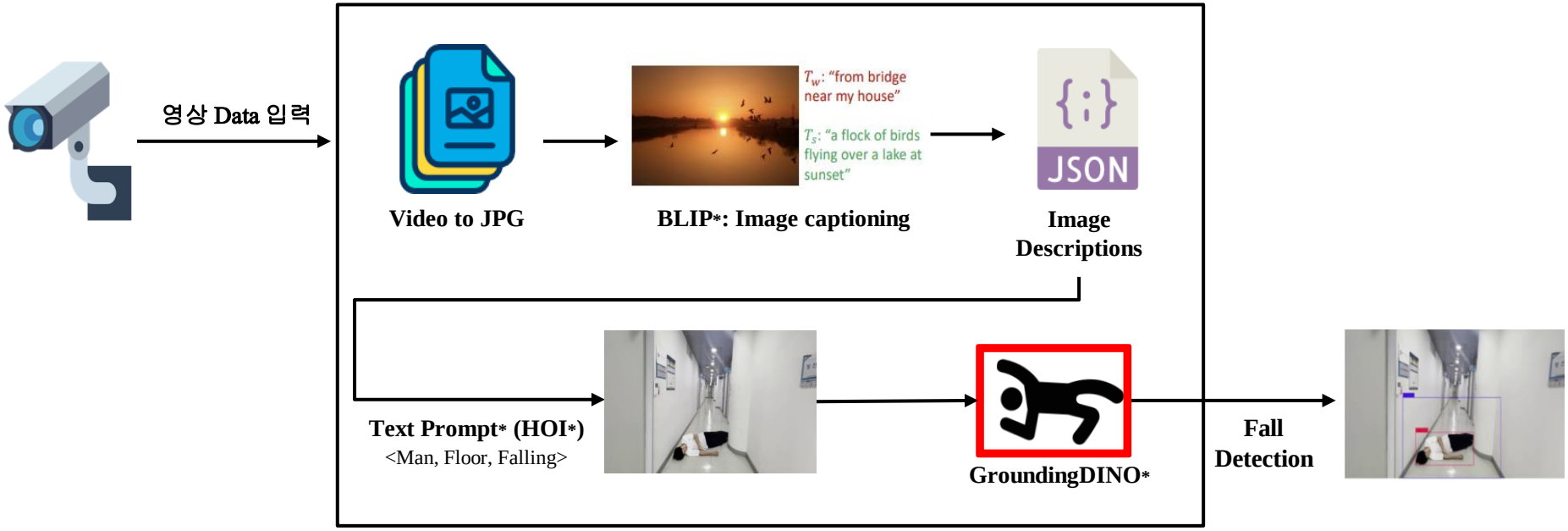
- Blip
- GroundingDINO
- Human-Object Interaction(HOI)



The Architecture of Rule-based Fall Detector.

- OpenPose와 가속도 제거
- OpenPose 대신 **BLIP[12]으로 사람의 행동 예측**
- 가속도 대신 **GroundingDINO[13]로 낙상 감지**

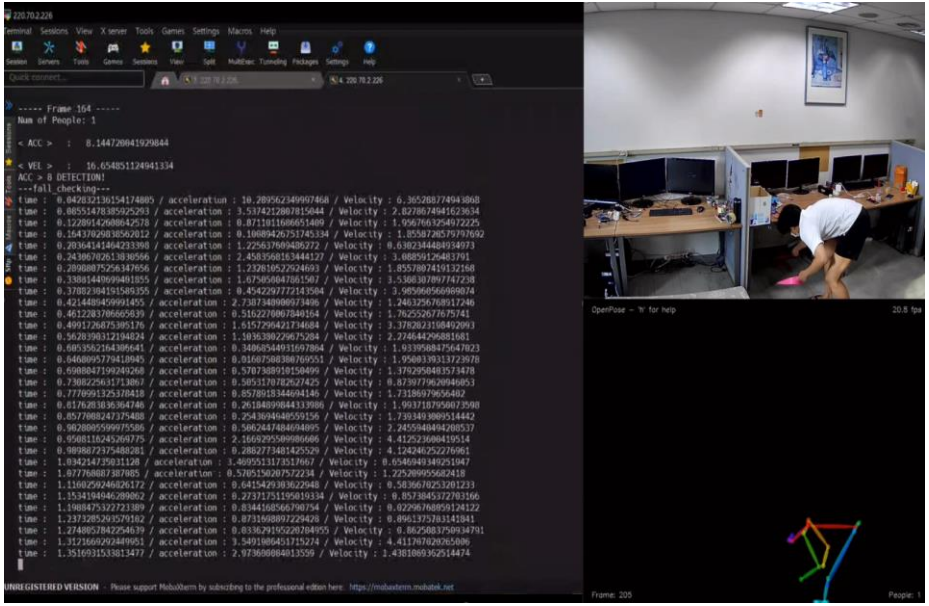
Zero-shot Fall Detection using Multi-modal Descriptions



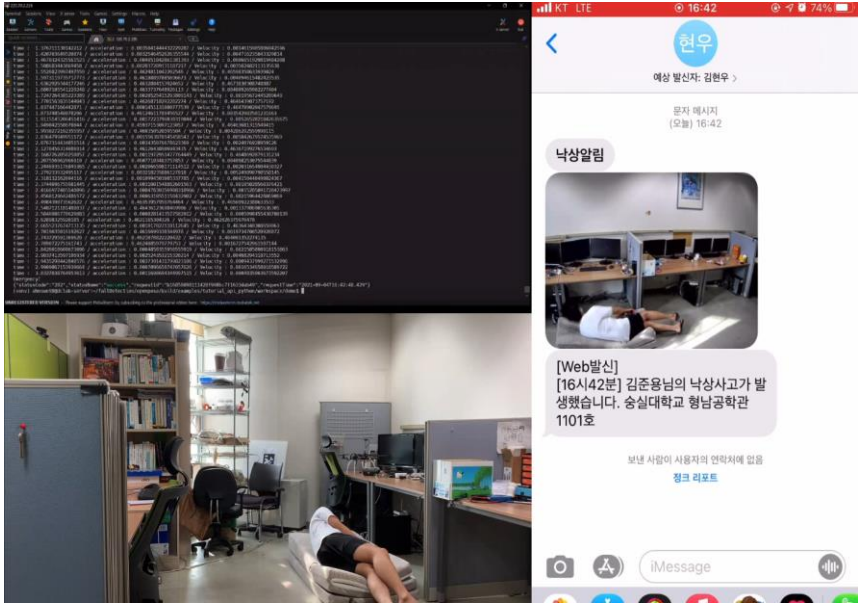
The Architecture of Zero-Shot Fall Detector.

* **BLIP** : 이미지 장면을 텍스트로 설명가능한 시각-언어 모델(Image Captioning)
* **Prompt** : 대규모 언어 모델에게 높은 품질의 응답을 얻어낼 수 있도록 입력 텍스트 값들을 조합하는 방식
* **Human-Object Interaction (HOI)** : 이미지 내 Human과 Object 사이의 관계를 파악해 Interaction을 예측하는 방법
* **GroundingDINO** : 입력 받은 텍스트와 일치하는 객체를 이미지에서 검출하여 나타내는 오픈소스

Real-World Example : Rule-based Fall Detection



Rule-Based Fall Detection Process on Real-Time



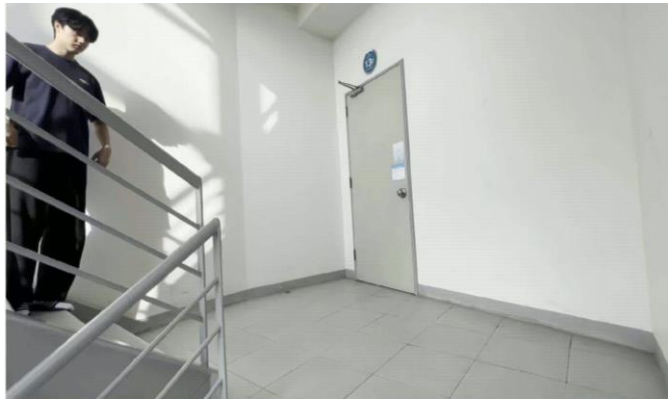
Rule-base Fall Detection Result on Real-Time

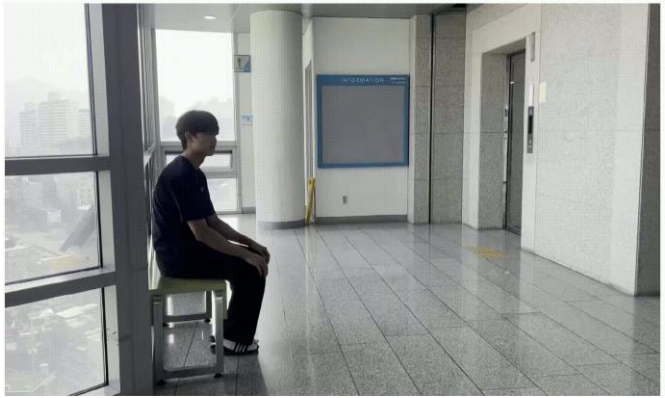
Real-World Example : Zero-shot Fall Detection

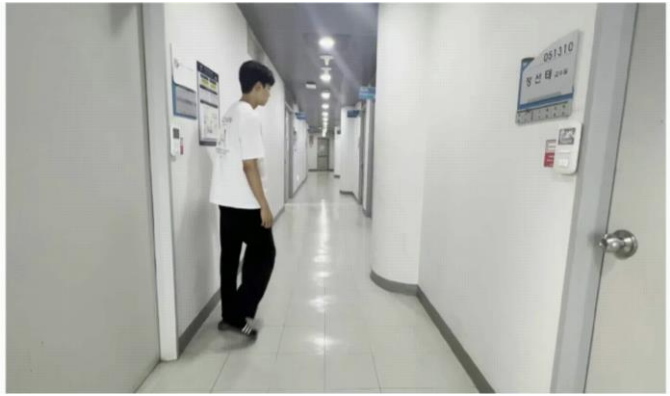
Experimental Setup

The Imaging Source
DFK 33UX264
+ 6mm Lens

MSI GS76 Stealth







Quantitative Evaluation

- **낙상 데이터 학습 없이** 기존의 Rule 기반의 낙상 감지와 **유사하거나더 우수한 정확도**를 보임
- HOI 사용 시 **여러 Object, Human를 한 이미지**에 나타낼 수 있어 **Interaction 파악**이 가능함

Method		Kaggle Dataset		Multiple Dataset [2]		UR Dataset [3]	
		AP	F1 score	AP	F1 Score	AP	F1 Score
기존 방식	Rule-based	-	-	0.8776	0.9080	0.4348	0.5647
제안 방식 (Zero-shot)	GroundingDINO [13]	0.6634	0.6816	0.8961	0.6937	0.4246	0.4262
	BLIP + GroundingDINO	0.7650	0.6442	0.9138	0.5240	0.5429	0.3333
	Ours	**0.7546**	**0.6158**	**0.9185**	**0.6066**	**0.4960**	**0.2778**

Comparison of our proposed Zero-Shot Fall Detection with Rule-based Fall Detection

Rule-based	Zero-shot
카테고리 확장 불가능	카테고리 확장 가능
제한 조건 많음	모든 CCTV에서 동작
이미지에서 동작 ×	이미지에서 동작 O
로직 복잡함	로직 단순함

Comparison Table


Zero-shot + X

BLIP + DINO (한 이미지에 한 가지만 표현 가능)

Person
(human)

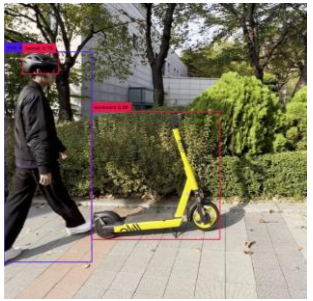
Helmet
(Object)

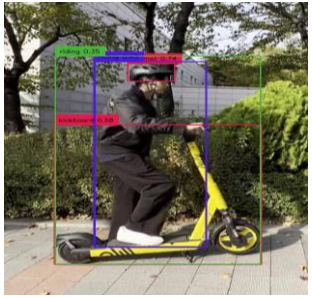
Kickboard
(Object)

Riding
(Interaction)

→ 킥보드를 안타는 경우에도 Riding 검출

BLIP + DINO + HOI (Ours)

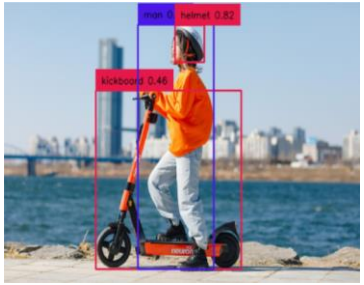




Zero-shot + Protective Equipment Detection

→ 킥보드를 안타는 경우, Riding 검출 ×

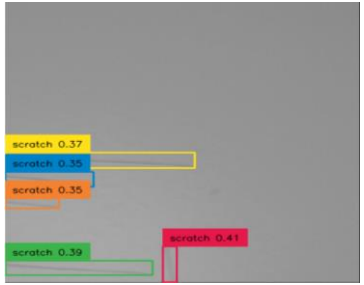
→ 킥보드를 타는 경우, Riding 검출

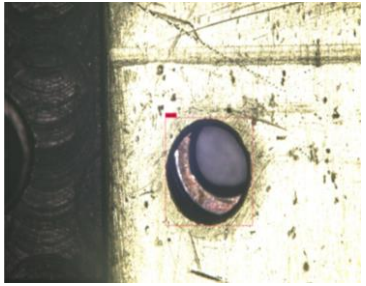


Protective Equipment
Detection

Fire
Detection

Dangerous Situation
Detection





Abnormal Detection
(Quality Control for Industrial Components)

4. 결론

Conclusion & Future Works

역할

- 제안하는 Zero-shot 이벤트 검출 방법론이 기업의 실질적 매출에 의미 있는 기여가 되도록 노력함
- 신설된 로봇 비전 수업에서 산업 수요가 반영된 최신 이론/데이터 기반 방법론을 주요 내용으로 교육
- 학위 과정 종료 및 채용 단계까지 본 산학프로젝트가 진행 및 후속성과가 도출되도록 노력함

결론

- [정확성] **낙상 데이터 학습 없이** Large Language Model*과 Text Prompting 기술만을 활용해 Zero-shot 낙상 감지를 가능하게 하여 기존의 Rule 기반의 낙상 감지와 **유사하거나더 우수한 정확도** 보임
- [효율성] **수집하고 학습하지 못한 문제 상황**에도 **Zero-shot 기반의 일반적인 방법론**을 통해 데이터 수집 및 학습 **비용 절감**의 가능성을 보임
- [유연성] 낙상 이외에 **다른 이상 상황 및 상태도 감지**할 수 있도록 **확장 가능한 모델**로 설계됨

*** Large Language Model**: 대규모의 텍스트 데이터로 사전 훈련되어 다양한 자연어 처리 작업에 활용할 수 있는 인공 지능 모델

4. 결론

Conclusion & Future Works

성과

- 김준용, 김성흠, "Fall Detection and Driver Carelessness Detection using Static Image-based Action Recognition," 28(3), Journal of Institute of Control, Robotics and Systems 2022
- 장재훈, 김준용, 김성흠, "A Large-scale 3D Object Dataset for 6-DoF Pose Estimation," Journal of Institute of Control, Robotics and Systems, Under review
- 김준용, 김성흠, "Deep Learning based Object Detection Method and its Application for Intelligent Transport Systems," 27(12), Journal of Institute of Control, Robotics and Systems 2021
- 길다영, 김성흠, "Lightweight Deep Learning for Room Layout Estimation with a Single Panoramic Image," Journal of Institute of Control, Robotics and Systems 2022
- Dayoung Kil and Seong-heum Kim, "Lightweight Room Layout Estimation using a Single Panoramic Image," ICCAS 2022
- Deepak Ghimire, Dayoung Kil, Seong-heum Kim, "A Survey on Efficient Convolutional Neural Networks and Hardware Acceleration," Electronics 2022 (피인용 횟수 60건 이상)
- 은상, 제 11회 숭실 캡스톤 디자인 경진대회
- 최우수상, AI융합학부 경진대회
- 길다영, 김성흠, "단일 파노라마 입력의 실내 공간 레이아웃 복원 모델 경량화," C-2022-036602 (소프트웨어등록)

계획

- Zero-shot을 통해 공정(라인)상에서 작업자, 장비, 공정의 **실시간 모니터링 시스템** 구현
- AOI(Automated Optical Inspection) 장비의 픽업 불량, 장비 이상 원인 감지 시스템을 **Zero-shot 방법론으로 고도화**
- 추가 장비 필요 없이 **모든 CCTV에 간단하게 적용**

Reference

[1] Daniil Osokin, “Real-time 2D Multi-Person Pose Estimation on CPU:Lightweight OpenPose”, arXiv 2018

[2] Dao Huu Hung, et al., Fall Detection with Two Cameras based on Occupied Area, Proc. of 18th Japan-Korea Joint Workshop on Frontier in Computer Vision. 2012

[3] Bogdan Kwolek, et al., Human fall detection on embedded platform using depth maps and wireless accelerometer, Computer Methods and Programs in Biomedicine, 2014

[4] 임동하, 박철호, 김상훈, 유윤섭, “3축 가속도 센서를 이용한 낙상 감지 시스템″, 한국정보처리학회 ,2013 May 10 , 2013년, pp.356 – 358

[5] 한도협, 지석훈, 배연두, 김한슬, 김영종, “ 가속도 센서 기반의 낙상 감지 알고리즘에 대한 연구″, 한국IT정책경영학회 논문지, ’20-12 Vol.12 No.06

[6] 강윤규, 이지나, 신용태, “순환신경망, GRU를 이용한 자세 추정 기반 낙상 감지 기법″, 한국IT정책경영학회 논문지, ’20-10 Vol.12 No.06

[7] 강윤규, 강희용, 원달수, “PoseNet과 GRU를 이용한 Skeleton Keypoints 기반 낙상 감지″, 한국IT정책경영학회 논문지, Vol. 22, No. 2 pp. 127-133, 2021

[8] 정필성, 조양현, “사물인터넷 기반의 낙상 감지 시스템″, 한국정보통신학회논문지, Vol. 19, No. 11 : 2546~2553 Nov. 2015

[9] 김준용, 김성흠, "Fall Detection and Driver Carelessness Detection using Static Image-based Action Recognition," 28(3), Journal of Institute of Control, Robotics and Systems, 2022

[10] Suchen Wang, et al., Learning Transferable Human-Object Interaction Detector With Natural Language Supervision. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. In CVPR, 2022, pp. 939-948

[11] Alec Radford, et al., Learning transferable visual models from natural language supervision. arXiv preprint arXiv: 2103.00020, 2021. 2, 3, 4, 5, 6, 8

[12] Junnan Li, et al., BLIP: Bootstrapping Language-Image Pre-training for Unified Vision-Language Understanding and Generation. Proceedings of the 39th International Conference on Machine Learning, PMLR 162:12888-12900, 2022

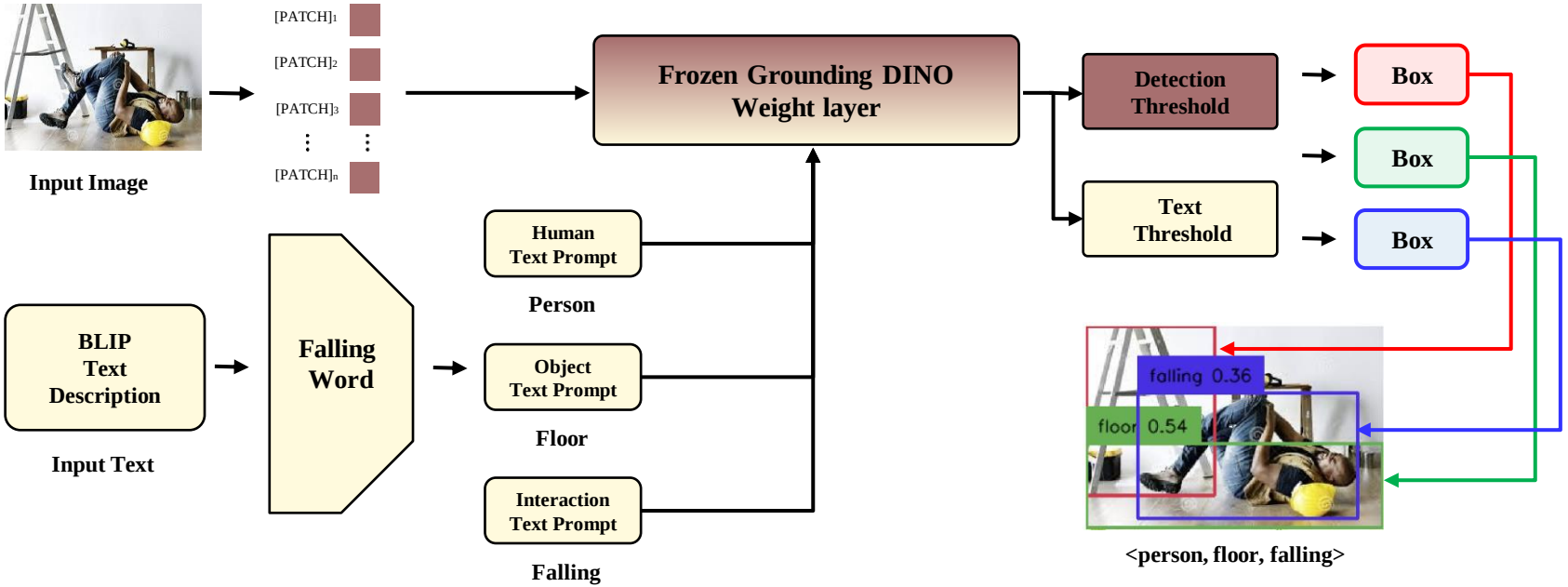
[13] Shilong Liu, et al., Grounding DINO: Marrying DINO with Grounded Pre-Training for Open-Set Object Detection. arXiv:2303.05499. 2023

[14] Alex Kendall, et al., PoseNet: A Convolutional Network for Real-Time 6-DOF Camera Relocalization. In ICCV, 2015, pp. 2938-2946

Thank You

참여 기업 : 고영테크놀로러지
대학 : Soongsil University
팀 : SSU VIP Lab. 석사 김준용, 석사 길다영

Zero-Shot Fall Detection using Multi-Modal Descriptions



The Framework of Zero-Shot Fall Detection.