경량 CNN과 센서 융합을 이용한 온디바이스 낙상 감지 시스템

안용호[1]*, 이성민[2], 최윤호[3], 백인혁[3], 김민현[3]
*숭실대학교[1], 서경대학교[2], 한양대학교[3]
*nanhosoft@soongsil.ac.kr, zoq2039@skuniv.ac.kr, yunho8418@hanyang.ac.kr
lovecat7@hanyang.ac.kr, minh3789@hanyang.ac.kr

On-Device Fall Detection via Lightweight CNN and Sensor Fusion

Ahn Yong Ho[1]*, Lee Seong Min[2], Choi Yoon Ho[3], Paik In Hyeok[3], Kim Min Hyeon[3]
*Soongsil Univ.[1], Seokyeong Univ.[2], Hanyang Univ.[3]

요 약

본 논문은 기존 중앙 서버 기반 AI 감시 시스템의 고질적인 문제인 네트워크 지연, 높은 전력 소모, 사생활 침해 문제를 해결하기 위해 엣지 컴퓨팅 아키텍처 기반의 실시간 응급상황 감지 시스템을 제안한다. 제안하는 시스템은 AI 추론에 특화된 NPU와 실시간 신호 처리를 위한 FPGA 가속기의 역할을 명확히 분담하여 단말 내에서 모든 연산을 수행한다. 특히, YOLO-Pose 기반 자세 분석, 경량 행동 인식 모델, FPGA로 가속된 mmWave 레이더의 호흡 분석의 3가지 독립적인 스트림의 결과를 융합하여 시스템의 강건성을 극대화했다. 자체 제작한 프로토타입을 통해 실험한 결과, 시스템은 평균 2.65초의 응답 속도와 91.5%의 높은 탐지 정확도를 달성하며 제안 시스템의 실용성과 우수성을 검증하였다.

Ⅰ. 서 론

고령화 사회 진입과 1인 가구의 증가는 독립적인 생활 공간에서의 안전 사고, 특히 낙상 사고에 대한 기술적 대응의 필요성을 증대시키고 있다. 기존 CCTV 기반 감시 시스템은 대부분 수동적 모니터링에 의존하며, 서버 기반 AI 솔루션은 데이터 전송 지연으로 인한 문제가 발생한다.[1] 이러한 한계를 극복하기 위해 본 연구에서는 모든 연산이 엣지 디바이스 내에서 완결되는 실시간 응급상황 감지 시스템을 제안한다.

Ⅱ. 제안하는 시스템

제안하는 시스템은 세 개의 핵심 프로세서가 유기적으로 연동하는 이기종 컴퓨팅 구조를 가진다. 먼저 Intel N100 프로세서가 전체 시스템의 운영체제를 구동하며, 각 하드웨어 제어, 데이터 취합 및 최종 판단을 수행한다. Hailo-8 NPU는 YOLOv8-Pose[2], 경량 비디오 분류 모델 등 연산량이 많은 딥러닝 모델의 추론을 전담하여 저전력, 고속으로 처리한다. Xilinx KV260 FPGA는 직접 설계 및 구현한 FFT IP를 탑재하여 mmWave 레이더의 원시 데이터를 실시간으로 전처리하고, 호흡 등 미세 움직임 패턴을 분석한다. 또한, 오탐지를 최소화하고 신뢰도를 극대화하기 위해, 아래와 같은 세 가지 분석 스트림을 병렬로 처리하고 그 결과를 최종 융합한다.

자세하게는, Hailo-8 NPU가 YOLO-Pose 모델로 영상에서 신체 주요 관절을 추정한 뒤, CPU가 신체 중심의 수직 속도, 기울기 등을 계산하는 휴리스틱 알고리즘으로 자세의 위험도를 정량화한다. 동시에, NPU는 다시 최적화된 경량 비디오 인식 모델을 사용하여 영상 클립에서 '낙상'과 같은 특정 행동의 발생 확률을 계산한다. 또한, FPGA의 FFT 가속기는 mmWave 레이더 신호를 분석하여 영상으로 탐지하기 어려운 호흡 유무 및 인체의 상태를 파악한다. 각 스트림에서 도출된 '자세 점수', '행동 확률', '호흡/상태 점수'는 최종 Decision Fusion 모듈에 입력되어 종합적인 상황 판단을 내린다. 이를 통해 단순 오탐지를 효과적으로 제거하고 시스템의 전체 신뢰도를 높인다.

Ⅲ. 실험 및 결과

3.1. 실험 환경

제안 시스템의 프로토타입은 Intel N100 기반 임베디드 보드, Hailo-8 NPU, Xilinx KV260 FPGA 보드, 그리고 RGB 카메라와 mmWave 레이더 센서를 통합하여 제작되었다. AI 모델 학습에는 공개 데이터셋[3][4][5][6][7]과 자체 제작한 데이터셋 및 Unity 3D 엔진으로 자체 구축한 합성 데이터 파이프라인을 통해 생성된 데이터를 활용하였다. 호흡 감지 및 인체 탐지 기법은 Acconeer사의 알고리즘[8]을 활용하였다.

3.2. FPGA 기반 FFT 가속 성능 검증

mmWave 레이더에서 입력되는 신호 처리를 CPU에서 수행할 때와 제안하는 FPGA 가속기를 사용했을 때의 성능을 비교하였다. 표 1과 같이, 처리할 데이터 크기가 16,384을 넘어서면서 하드웨어 가속의 효과가 명확히 나타났으며, 65,536 포인트에서는 소프트웨어 대비 약 7.59배의 성능 향상을 보였다. 이는 CPU의 연산 부담을 FPGA로 효과적으로 오프로드했음을 증명한다.

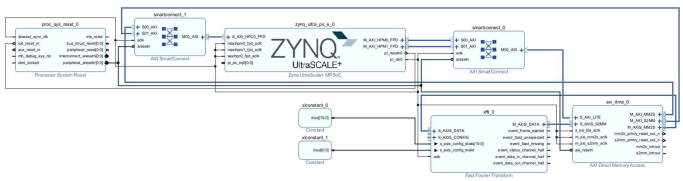

그림 1. FPGA 배선도 다이어그램

FFT 크기	4,096	16,384	65,536
CPU	0.569 ms	2.897 ms	17.640 ms
FPGA	2.320 ms	2.325 ms	2.325 ms
향상폭	0.25배	1.25배	7.59배

표 1. FFT 연산 성능 비교표
* 소프트웨어(Python) 연산은 Numpy 라이브러리의 구현체(rfft)를 사용함.

3.3. 비디오 모델 성능 검증

기존 선행연구에 따르면, 시계열 및 문맥을 검증할 수 있는 Transformer 기반 모델을 Hailo-8 NPU에서 구동하는 것이 가능하였다.[9] 그러나, 전체적인 시스템 구성 및 통합을 위해 요구되는 성능을 만족시키지 못하는 문제가 발생하여, 본 시스템에서는 부득이하게 COCO Dataset으로 사전 학습된 MobileNet-v3[10]의 가중치를 활용하고, 최종 계층에 FC-Layer와 드롭아웃을 조정한 이진 분류기를 연결하여 상기한 데이터셋 기반으로 전이학습한 모델을 사용하였다. 해당 모델은 선행연구 대비 보다 경량의 모델임에도 불구하고, 적은 데이터셋 모수에서 보다 강한 성능을 발휘하는 특성이 있어 본 시스템 적용에 적합하였다. 최종적으로 학습된 모델은 Hailo-8 NPU를 위한 양자화 및 컴파일 이후 테스트에서 Unseen 데이터셋 대상으로 90.4%의 정확도를 달성하였다.

모델	Accuracy	Precision	Recall	F1 Score
MobileNet v3 전이학습체	0.9046	0.8360	0.9256	0.8678
MISAKA NETWORK	0.7644	0.8037	0.7818	0.7926

표 2. MobileNet 전이학습 모델과 MISAKA-NETWORK-H 모델 간 성능 측정표

3.4. 최종 시스템 동작 검증

실제 낙상 시나리오 및 일상생활 동작을 포함한 자체 테스트셋으로 최종 시스템의 성능을 평가하였다. 비교군으로는 일반적인 중저가형 워크스테이션 장비를 사용하였다. 결과 평균 2.65초의 빠른 응답 속도를 기록했으며, 200회의 실험에서 91%의 정확도를 달성하였다. 또한 워크스테이션 대비 뛰어난 전력 효율과 더불어 유사한 정확도를 확보하는 데 성공하였다. 이는 데이터 융합 로직이 오탐지를 보완하고, 프로세서 간 동시 처리를 통해 실시간성을 보장한 결과이다.

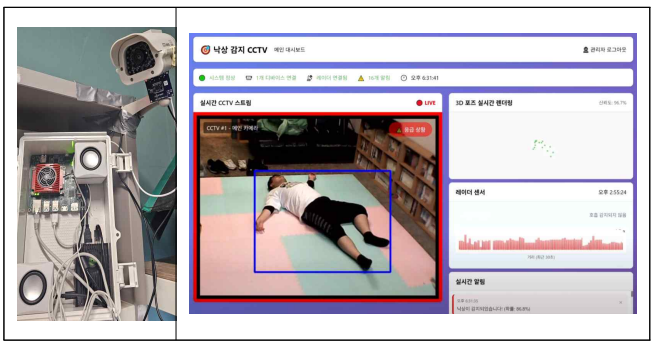

그림 2. 프로토타입 장비 및 실험 관리용 대시보드

항목	엣지 디바이스 측정값	워크스테이션 측정값
정확도	91.5%	94.5%
평균 반응 시간	2.65초	2.95초
소비전력	~25W	~350W

표 3. 본 시스템 및 워크스테이션간 성능 측정표
* 워크스테이션은 Ryzen 7500F, RTX 3060Ti, DDR5 32GB 시스템에서 측정
** 워크스테이션은 1Gbps 네트워크를 통해 비디오 프레임을 전송받은 환경임.

Ⅳ. 결론

본 논문에서는 NPU와 FPGA의 역할을 분담하여 저전력·고효율의 실시간 응급상황 감지 시스템을 성공적으로 구현하였다. 실험 결과를 통해 제안 시스템이 기존 서버 기반 시스템의 한계인 지연과 고비용 문제를 극복하고, 높은 정확도와 실시간성을 동시에 달성할 수 있음을 보였다. 본 연구는 향후 독거노인 비대면 케어, 산업 현장 안전 관리 등 다양한 분야에 실질적으로 기여할 수 있는 기반 기술이 될 것으로 기대된다.

ACKNOWLEDGMENT

본 논문은 과학기술정보통신부 정보통신창의인재양성사업의 지원을 통해 수행한 한이음 드림업 프로젝트 결과물입니다.

참 고 문 헌

[1] 박용석(Yong-Suk Park), 김현식(Hyun-Sik Kim), 최규만(Kyuman Choi), "영상기반 독거노인 낙상감지 시스템의 구현," 한국통신학회 학술대회논문집, vol. 2019.11, pp. 304-305, Nov. 2019.

[2] R. Varghese and S. M., "YOLOv8: A Novel Object Detection Algorithm with Enhanced Performance and Robustness," 2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS), Chennai, India, 2024, pp. 1-6, doi: 10.1109/ADICS58448.2024.10533619.

[3] E. Alam, A. Sufian, P. Dutta, M. Leo, and I. A. Hameed, "GMDCSA24: A Dataset for Human Fall Detection in Videos," Zenodo, 2024. [Online]. Available: https://doi.org/10.5281/zenodo.12921216

[4] I. Charfi, J. Mitran, J. Dubois, M. Atri, and R. Tourki, "Optimised spatio-temporal descriptors for real-time fall detection: comparison of SVM and Adaboost based classification," Journal of Electronic Imaging (JEI), vol. 22, no. 4, pp. 17, Oct. 2013.

[5] Jose Camilo Eraso, Elena Muñoz, Mariela Muñoz, and Jesus Pinto, "Dataset CAUCAFall," Mendeley Data, V4, 2022. doi: 10.17632/7w7fccy7ky.4

[6] T. Grutschus, O. Karrar, E. Esenov, and E. Vats, "Cutup and Detect: Human Fall Detection on Cutup Untrimmed Videos Using a Large Foundational Video Understanding Model," arXiv preprint, 2024. [Online]. Available: https://arxiv.org/abs/2401.16280

[7] B. Kwolek and M. Kepski, "Human fall detection on embedded platform using depth maps and wireless accelerometer," Computer Methods and Programs in Biomedicine, vol. 117, no. 3, pp. 489-501, Dec. 2014. doi: 10.1016/j.cmpb.2014.09.003

[8] Acconeer, "acconeer-python-exploration," GitHub Repository, 2025. https://github.com/acconeer/acconeer-python-exploration.

[9] 안용호, 이성민, "MISAKA-NETWORK-H: 실시간 긴급 탐지를 위한 Hailo-8 NPU 최적화 비디오 트랜스포머," 한국통신학회 학술대회논문집, pp. 2012-2013, 제주, 2025-06-18.

[10] A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, et al., "Searching for MobileNetV3," arXiv preprint arXiv:1905.02244, 2019.