

SHAP 기반 관절 선택을 통한 포즈기반 사람 이상행동 탐지 최적화

곽민준, 정규원, 심훈보, 전창재
세종대학교

gminjun76@gmail.com, kwsosoo0815@gmail.com, hunbo00@sju.ac.kr, cchun@sejong.ac.kr

Improving Efficiency of Pose-based Human Anomaly Detection via SHAP-based Joint Selection

Minjun Kwak, Kyuwon Jung, Hunbo Shim, Chang-Jae Chun
Sejong Univ.

요 약

포즈 기반 이상행동 탐지는 배경 노이즈 제거, 개인정보 보호, 연산 효율성 면에서 이점이 크지만, 대부분의 연구가 모든 관절을 이용한다는 점에서 불필요한 노이즈와 연산량을 수반한다는 단점이 있다. 본 논문은 이러한 연산 비효율과 노이즈 문제를 해결하기 위해, 모델 판단에 있어 각 관절의 기여도를 정량적으로 측정하여 입력 데이터 자체를 최적화하는 전략을 제안한다. 기존 연구들이 모델 구조의 경량화에 집중한 것과 달리, 본 연구는 XAI(Explainable AI) 기법인 SHAP(Shapley Additive exPlanations)을 활용해 각 관절의 기여도를 정량적으로 분석하고 핵심 관절만을 선택적으로 사용한다. STG-NF 모델을 베이스라인으로 SHT(ShanghaiTech Campus) 데이터셋에서 실험한 결과, 12 개의 핵심 관절만으로도 전체 관절 사용 시보다 높은 탐지 정확도를 기록하였으며, GFLOPs 51.5% 절감 및 FPS 24.40% 향상을 달성하였다. 이는 연산 자원이 제한된 저사양 엣지 환경에서의 실시간 보안 시스템 구현 가능성을 제시한다.

I. 서론

사람 이상행동 탐지(Human Anomaly Detection)는 복잡한 인체의 움직임과 동적인 주변 환경의 영향으로 인해 컴퓨터 비전 분야에서 여전히 도전적인 과제로 남아 있다. 최근 Diffusion 모델을 활용하여 가상 이상 데이터를 생성하는 PA-VAD[1] 등이 SHT(ShanghaiTech Campus) 데이터셋[5]에서 높은 성능을 보이고 있지만, 연산 비용, 개인정보 문제 배경 노이즈로 인해 포즈(Pose) 기반 방식이 주목받고 있다.

포즈 기반 모델은 배경 노이즈로부터 자유롭고 개인정보를 보호할 수 있으며, 연산 효율성이 뛰어나다는 장점이 있다. 최근 DA-Flow[2]는 0.488K 의 적은 파라미터만으로 포즈 기반 모델 중 SHT 데이터셋 SOTA 성능을 기록하며 모델 구조 경량화의 가능성을 보여주었다. 하지만 이러한 최신 모델들도 모든 관절 정보를 동일한 비중으로 처리하며, 이는 판단과 무관한 관절의 움직임이 노이즈로 작용하여 탐지 성능을 저해하는 요인이 된다.

본 논문에서는 모델의 구조적 복잡도를 개선하는 기존 연구 방향과 달리, 입력 데이터의 자체를 최적화하는 전략을 제안한다. 이는 불필요한 관절 정보를 제거함으로써 순도 높은 정보만을 남기며, 연산량을 획기적으로 낮춰 모델 설계의 유연성 확보하는 데 목적이 있다. 이를 위해 STG-NF[3]를 실험 베이스라인으로 채택하고, XAI(eXplainable AI)의 SHAP(Shapley Additive exPlanations) 지표를 활용하여 각 관절의 기여도를 정량적으로 분석한다.

실험 결과, SHT 데이터셋에서 12 개의 핵심 관절을 선별하여 활용했을 때 전체 관절을 사용하는 것보다 높은 탐지 정확도를 기록했다. 이는 입력데이터 최적화가 노이즈 억제와 성능 향상을 동시에 달성할 수 있음을 나타낸다. 결과적으로 본 기법은 전체 관절 대비

GFLOPs 51.5% 절감 및 FPS 24.40% 향상이라는 경량화 성능을 구현하였으며, 이는 모델의 복잡도를 더 극대화하여 추가적인 성능 향상을 목표로 하거나, 저사양 엣지 환경에 배포할 수 있는 실질적인 기술적 근거를 제공한다.

II. 본론

2.1 관절별 기여도 분석

SHAP(SHapley Additive exPlanations)[4]은 게임 이론의 Shapley Value 를 기반으로 복잡한 모델의 예측 결과를 해석하기 위해 제안된 프레임워크이다. SHAP 의 목표는 모델의 특정 예측값에 대해 각 입력 특징(Feature)이 기여한 영향력을 정량적인 수치로 측정하는 것이다. 이 기법은 가산적 특징 속성(Additive Feature Attribution) 방식을 따르며, 국소 정확성, 누락성, 일관성이라는 세 가지 수학적 공리를 동시에 만족하는 유일한 방안을 제공한다는 점에서 이론적 타당성을 갖는다. 본 연구에서는 SHAP 을 활용하여 STG-NF[3] 모델이 이상행동을 판별할 때 각 신체 관절 데이터가 미치는 기여도를 정량적으로 분석한다. 이를 통해 모델의 판단에 유의미한 영향을 주는 핵심 관절을 선별하고, 불필요한 관절 정보를 제거함으로써 모델의 연산 복잡도를 최적화하며 노이즈 제거에 활용한다.

2.2 사용 모델

STG-NF[3]는 포즈 기반 이상행동 탐지 모델로 사람 관절 데이터를 시공간 그래프 구조로 변환하여 이상행동을 탐지하는 Normalizing Flow 기반 모델이다. 사람 관절 움직임의 복잡한 분포를 단순한 가우시안 분포로 매핑하는 가역적 변환 함수를 학습하며, 변환 과정 중 ST-GCN 블록을 통해 관절 간의 시공간적 패턴을 학습한다. 학습 시 정상 행동 데이터만 이용하여

음의 로그 가능도(Negative Log-Likelihood)를 최소화함으로써 정상 행동 분포를 학습하며, 추론 단계에서는 학습된 분포에서 낮은 확률 밀도를 갖을 시 이상행동으로 판단한다. 우리는 SHAP 기반 특징선택 알고리즘을 바탕으로 STG-NF 입력부를 재설계하여 사용하였다.

2.3 데이터 구성

본 논문에서 제안하는 핵심 관절 선별을 통한 성능 향상 및 경량화를 검증하기 위해 사람 이상행동 탐지 분야에서 가장 널리 사용되는 벤치마크인 ShanghaiTech Campus(SHT) Dataset[5]을 활용하였다. SHT 데이터셋은 13 개의 서로 다른 장소에서 촬영된 약 317,000 개의 프레임으로 구성되어 있다. 복잡한 조명 조건과 다양한 카메라 각도를 포함하고 있으며, 총 130 개의 이상행동 상황이 포함되어 있다. 비지도 학습을 위해 정상 행동으로만 구성된 274,515 개의 훈련 프레임과 정상 및 이상행동으로 구성된 42,883 개의 평가 프레임을 사용하였다.

본 연구에서는 SHAP 분석 통해 각 관절의 기여도(SHAP value)를 정량화하고, 모델의 최대 효율성을 찾기 위해 기여도가 높은 관절 3 개, 5 개, 9 개, 12 개, 15 개를 선별하여 실험을 진행했다.

2.4 실험 결과

2.4.1 관절 기여도 분석 결과

순위	관절	기여도	순위	관절	기여도
1	우측 팔꿈치	0.001557	10	좌측 손목	0.000884
2	좌측 발목	0.001370	11	우측 골반	0.000857
3	좌측 팔꿈치	0.001363	12	우측 어깨	0.000783
4	좌측 어깨	0.001121	13	코	0.000752
5	우측 발목	0.001107	14	우측 무릎	0.000691
6	좌측 귀	0.001039	15	우측 눈	0.000632
7	우측 손목	0.001037	16	좌측 골반	0.000570
8	좌측 무릎	0.000924	17	우측 귀	0.000511
9	좌측 귀	0.000921	18	목	0.000387

표 1 관절 기여도 분석 결과

SHAP 분석을 통해 SHT 데이터셋 내 18 개 관절의 기여도를 정량화하였다. 분석 결과, 우측 팔꿈치, 좌측 발목 등 사지 관절이 상위권을 차지했으며, 이는 달리기, 싸움 등 동적 이상행동의 특성이 반영된 결과이다. 반면 목, 귀 등 신체 중심부 및 가동 범위가 작은 관절은 기여도가 낮아 모델 판단에 노이즈로 작용함을 확인하였다.

2.4.2 관절 별 탐지 성능 및 연산량 분석

입력 관절	관절 개수	인접 행렬	파라미터 수	AUC-ROC	AUC-PR	EER	FPS	GFLOPs
ALL	18	OpenPose	616	85.76%	83.59%	26.64%	7359.3	0.95
TOP-3	3	FC	616	83.74%	82.41%	27.82%	11962.8	0.05
TOP-5	5	FC	616	84.04%	82.59%	27.70%	9999.2	0.11
TOP-9	9	FC	616	85.32%	83.39%	26.84%	9134.4	0.28
TOP-12	12	FC	616	85.96%	83.88%	26.24%	9155.1	0.46
TOP-15	15	FC	616	85.74%	83.81%	25.76%	8623.0	0.60

표 2 입력 관절별 성능 비교표

실험은 STG-NF[3]의 공식 코드를 기반으로 고정된 시드(Seed) 환경에서 진행했다. 입력 데이터는 전체 관절과 SHAP 기여도 상위 K 개(Top-3, 5, 9, 12, 15)로 구성했다. Top-K 실험 시, 선택된 관절 간의 특징 추출을 위해 기존의 골격 구조(OpenPose) 대신 완전 연결(Fully Connected) 인접행렬을 적용했다. 표 2 와 같이 성능 지표로는 AUC-ROC, AUC-PR, EER, 경량화 지

표로 GFLOPs, FPS 을 사용했다. TOP-12 조합의 성능이 베이스라인 대비 AUC-ROC 는 0.2%p 향상되었으며, GFLOPs 51.5% 절감 및 FPS 24.40% 향상을 달성하였다. 이는 불필요한 관절 정보 제거가 모델의 연산 복잡도는 획기적으로 낮추면서도 노이즈 억제를 통해 탐지 정확도를 개선할 수 있음을 입증한다. 본 결과는 파라미터 수의 변화 없이 입력 데이터 최적화만으로 저사양 엣지 환경에서의 실시간 보안 시스템 구현 가능성을 제시한다.

III. 결론

본 논문은 포즈 기반 이상행동 탐지 모델에서 모델 구조 변경 없이 입력 데이터 최적화만으로도 모델의 경량화와 탐지 성능 향상을 동시에 달성할 수 있음을 입증했다. 대부분의 기존 연구가 모델 아키텍처의 경량화에 초점을 맞춘 것과 달리, 본 논문은 XAI 기법인 SHAP 을 활용하여 각 관절 데이터의 기여도를 정량적으로 분석하고 이상행동 판단에 유의미한 핵심 관절만을 선별하는 전략을 제안했다.

실험 결과, SHT 데이터셋에서 모델 파라미터 수 0.616K 동일하게 유지한 상태에서 기여도 상위 12 개 관절만을 입력값으로 사용했음에도 불구하고, 전체 관절을 사용한 베이스라인 대비 AUC-ROC 85.96%라는 더 높은 정확도를 기록했다. 동시에 GFLOPs 51.5% 절감 및 FPS 24.4% 향상이라는 연산 효율성 증가를 입증했다. 이는 불필요한 관절 정보가 모델의 판단에 노이즈로 작용하는 것을 나타내며, 입력 데이터 최적화를 통해 저사양 엣지 환경에서도 실시간 보안 시스템 구현할 수 있는 가능성을 제시한다.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 정보통신방송혁신인재양성(메타버스융합대학원)사업 연구 결과로 수행되었습니다(IITP-2026-RS-2023-00254529).

참 고 문 헌

- [1] [S. Hashimoto, H. Nishimura, Y. Wang, and M. Kurokawa, "Pseudo Anomalies Are All You Need: Diffusion-Based Generation for Weakly-Supervised Video Anomaly Detection," arXiv preprint arXiv:2512.06845, 2025.
- [2] R. Wu, Y. Chen, J. Xiao, B. Li, J. Fan, F. Dufaux, C. Zhu, and Y. Liu, "DA-Flow: Dual Attention Normalizing Flow for Skeleton-based Video Anomaly Detection," IEEE Transactions on Multimedia, vol. 27, pp. 8847–8858, 2025.
- [3] O. Hirschorn and S. Avidan, "Normalizing Flows for Human Pose Anomaly Detection," in Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2023, pp. 13545–13554.
- [4] S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," in Advances in Neural Information Processing Systems (NIPS/NeurIPS), vol. 30, 2017.
- [5] W. Liu, W. Luo, D. Lian, and S. Gao, "Future Frame Prediction for Anomaly Detection – A New Baseline," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 6536–6545.