

검토

# 딥 러닝 기반 영상 감시의 이상 징후 탐지: 설문 조사

후-탄 두옹<sup>ID</sup>, 비엣-투안 레<sup>ID</sup>, 빈 쩌엥 황\*<sup>ID</sup>

정보 기술 학부, 호치민시 열린 대학교, 97 Vo Van Tan, 3군, 호치민시 700000, 베트남;  
thanh.dh@ou.edu.vn (H.-T.D.); tuan.lv@ou.edu.vn (V.-T.L.)  
\* 문의: vinh.th@ou.edu.vn

**요약:** 비디오 감시에서 이상 징후 탐지는 연구 커뮤니티에서 점점 더 많은 관심을 받고 있는 고도로 발전된 주제입니다. 스트리밍 비디오에서 비정상적인 이벤트를 자동으로 감지할 수 있는 지능형 시스템에 대한 수요가 매우 높습니다. 이에 따라 공공 보안을 보장하는 효과적인 모델을 구축하기 위한 다양한 접근 방식이 제안되고 있습니다. 네트워크 이상 탐지, 금융 사기 탐지, 인간 행동 분석 등 이상 징후 탐지에 대한 다양한 조사가 이루어졌습니다. 딥러닝은 컴퓨터 비전의 여러 측면에 성공적으로 적용되었습니다. 특히 생성 모델의 강력한 성장은 제안된 방법에서 사용되는 주요 기법이라는 것을 의미합니다. 이 백서에서는 비디오 이상 감지 분야에서 사용되는 딥러닝 기반 기법에 대한 포괄적인 검토를 제공하는 것을 목표로 합니다. 특히, 딥러닝 기반 접근 방식은 목적과 학습 지표에 따라 여러 가지 방법으로 분류되었습니다. 또한 비전 기반 영역에 대한 전처리 및 특징 엔지니어링 기법에 대해서도 자세히 설명합니다. 또한 비정상적인 인간 행동을 훈련하고 감지하는 데 사용되는 벤치마크 데이터베이스에 대해서도 설명합니다. 마지막으로 영상 감시의 일반적인 과제를 논의하여 향후 연구를 위한 몇 가지 가능한 해결책과 방향을 제시합니다.

**키워드:** 비정상적인 인간 활동 인식; 딥 러닝; 비디오 감시; 이상 징후 감지



**Citation:** Duong, H.-T.; Le, V.-T.;

Hoang, V.T. 비디오 감시에서 딥 러닝

기반 이상 감지: 설문 조사. *센서* **2023**,

23, 5024. [https://doi.org/10.3390/](https://doi.org/10.3390/s23115024)

s23115024

학술 편집자: 보 저우와 서성호

수신됨: 2023년 2월 27일

개정일: 2023년 4월 28일

수락되었습니다: 수락: 2023년 5월 8일

발행일: 2023년 5월 24일

**저작권:** © 2023 작성자. 라이선스 소유자 MDPI, 스위스 바젤. 이 문서는 크리에이티브 커먼즈 저작자표시(CC BY) 라이선스 조건에 따라 배포되는 오픈 액세스 문서입니다(<https://creativecommons.org/licenses/by/4.0/>).

## 1. 소개

오늘날 공항, 기차역, 슈퍼마켓, 학교, 혼잡한 거리와 같은 공공장소에서 보안에 대한 수요가 증가함에 따라 감시 카메라는 일상 활동을 모니터링하고 비정상적인 이벤트를 감지하는 데 사용됩니다. 이 작업은 비디오의 시간적 정보와 부분 정보를 모두 사용하여 이상 현상을 찾아내는 데 중점을 둡니다. 이상 행동은 정상

적인 행동에서 벗어난 이벤트[1]로 정의할 수 있는데, 예를 들어 공항에서의 싸움, 몰래카메라, 가방 방치 등이 이에 해당합니다. 감시 카메라를 사용하는 목적은 비정상적인 사람의 행동을 조기에 감지하는 것입니다. 이는 범죄 예방이나 테러 대응 등 사람의 개입이 필요한 많은 경우에 매우 중요한 작업입니다. 그러나 비정상적인 이벤트는 0.01%만 발생하고 감시 시간의 99.9%가 낭비되기 때문에 이 과정은 노동 집약적이고 지속적인 사람의 주의가 필요하며, 지루한 과정입니다[2]. 또한, 감시 시스템은 불필요한 저장 공간을 필요로 하는 중복 영상 데이터를 많이 생성합니다. 인적 오류와 저장 비용을 줄이기 위해서는 위험한 상황으로 이어질 수 있는 이상 행동을 감지할 수 있는 효율적인 감시 시스템을 구축해야 합니다. 이를 위해서는 각 행동을 대표하는 특징을 이해하기 위해 인간 행동 인식에 대한 심도 있고 포괄적인 연구가 필요합니다.

영상에서 이상 행위 감지는 교통사고 감지, 범죄 행위 감지, 불법 행위 감지 등 다양한 용도로 활용되고 있습니다. 또한 민감한 지역에서는 총이나 칼과 같은 비정상적인 물건이나 버려진 물체를 감지하는 것이 필요합니다.

이상 징후 탐지는 최근 몇 년 동안 활발하게 연구되고 있는 분야입니다. 비정상 이벤트를 탐지하기 위한 자동화된 프로세스를 수행하기 위해 많은 기업들이

방법이 제안되었습니다. 핵심 아이디어는 먼저 훈련 비디오에서 정상 패턴을 학습한 다음 정상 사례의 표현을 추출하는 것입니다. 이러한 표현에서 벗어나는 이벤트가 있으면 비정상적인 이벤트를 감지할 수 있습니다. 하지만 영상 감시와 사람 활동 인식에는 여전히 많은 어려움이 있습니다. 이러한 어려움은 특징 추출 단계와 관련이 있으며, 여기서 가려짐, 겹침, 복잡한 배경, 센서 노이즈, 낮은 조도, 동적 배경 변화 등이 시스템 성능에 영향을 미칠 수 있습니다[3]. 이러한 다양한 문제들로 인해 학습 모델을 위한 강력하고 변별력 있는 특징을 추출하는 것은 매우 어렵습니다. 또한 이상 행동 감지는 장면 컨텍스트[4]에 따라 달라지는데, 어떤 장면에서는 한 행동이 비정상적인 것으로 간주되지만 다른 장면에서는 정상으로 간주될 수 있습니다. 이를 위해서는 실제 세계에서 일어나는 모든 사례를 검증하기 위해 방대한 양의 훈련용 데이터가 필요합니다.

이 설문조사에서는 현재 영상 감시에 사용되고 있는 딥 러닝 방법을 살펴봅니다. 특히, 지능형 비디오 시스템의 핵심은 행동을 식별하고 인식하는 것이므로 비정상적인 인간 행동에 대한 이해에 중점을 둡니다. 이 설문조사의 범위에는 비전 기반 접근 방식이 행동 인식의 인기 트렌드가 되고 있기 때문에 비전 기반 접근 방식도 포함됩니다. 이 설문조사의 중요한 기여는 다음과 같습니다:

- 비디오 감시를 위한 비전 기반 인간 활동 인식에 대한 종합적인 검토.
- 이상 징후 탐지에 사용되는 인기 있는 데이터베이스에 대한 설명입니다.
- 딥러닝 모델을 위한 데이터 처리 및 기능 엔지니어링 분석.
- 최신 딥 러닝 모델에 대한 논의와 함께 장점과 단점을 살펴봅니다.
- 비디오 감시에서 이상 징후 탐지를 위한 기존 과제와 향후 연구 현황을 파악합니다.

설문조사를 준비하기 위해 가장 먼저 해야 할 일은 딥러닝 기술을 활용한 영상 감시와 관련된 최근 논문을 검색하고 필터링하여 사람의 활동을 이해하는 것이었습니다. *영상 감시*, *딥러닝 활동 인식*, *인간 행동 표현*, *이상 행동 감지* 등의 관련 색인 용어를 사용하여 연구 논문을 다운로드했습니다. 이 논문들에서 참고 문헌을 추출하고 검증하여 설문조사에 추가했습니다. 그 후 영상 감시와 관련된 연구도 포함되었습니다. 마지막으로 각 학습 모델에 대한 인사이트를 제공하기 위해 일반적인 구조와 대표적인 방법을 자세히 설명했습니다.

이번 설문조사의 구성은 다음과 같습니다: 섹션 2에서는 설문조사의 배경과 관련 설문조사를 제공합니다. 섹션 3에서는 영상 감시에 사용되는 일반적인 데이터베이스에 대한 논의가 제공됩니다. 섹션 4에서는 비전 기반 영역의 데이터 처리 및 기능 엔지니어링에 대해 설명합니다. 섹션 5에서는 최첨단 딥 러닝 모델에 대한 자세한 분석이 제공됩니다. 섹션 6에서는 이상 징후 탐지의 기존 과제에 대해 논의하고 향후 연구 방향을 제시합니다. 마지막으로 섹션 7에서는 설문조사를 마무리합니다.

## 2. 배경 지식 및 관련 저작물

### 2.1. 배경 지식

인간 행동 인식(HAR)은 컴퓨터 비전의 근본적인 문제이며 오랫동안 연구되어 왔습니다. HAR의 목적은 특정 이벤트를 이해하고 분석하기 위해 비디오에서 일어나는 동작을 식별하는 것입니다. 동영상의 공간적, 시간적 정보는 사람의 행동을 정확하게 식별하고 동영상을 분류하기 위해 HAR에서 중요한 역할을 합니다. 영상 감시에서 비정상적인 행동을 감지하는 것은 기본적으로 공항이나 기차역과 같은 실내 및 실외 위치에서 보안을 보장하는 데 사용됩니다. 비정상적인 사람 행동 인식(AbHAR)은 HAR의 특정 문제로 간주될 수 있습니다. 비디오에서 AbHAR의 문제는 매우 다양할 수 있으며 모든 문제 사례를 해결할 수 있는 단일 접근 방식이 없다는 것입니다. AbHAR에 대한 일반적인 접근 방식은 이미지 시퀀스에서 특징 추출에 의존합니다. 이러한 특징은 물체 감지, 포즈 추정, HAR에 유용한 고밀도 궤적 찾기 등의 작업을 위해 구축됩니다.

AbHAR은 보안을 보장하고 지시를 내릴 목적으로 행동과 사람들의 활동을 모니터링하기 위한 비디오 감시에 사용됩니다. 이를 위해서는 많은 비디오 처리 영역에서 특정 일반적인 배경 지식이 필요합니다. 비디오 감시에는 특징 추출, 장면 이해, 객체 추적, 객체 식별 및 모델 생성에 대한 지식 구현이 포함됩니다. 특히 특징 추출 단계는 시스템 전체에 큰 영향을 미치기 때문에 신중하게 다뤄야 합니다. 기존의 머신러닝 알고리즘은 비디오 데이터에서 얇은 특징을 학습하는 방식으로 AbHAR에서 괄목할 만한 성과를 거두었습니다. 랜덤 포레스트(RF)[5], 베이지안 네트워크[6], 마르코프 모델[7], 서포트 벡터 머신(SVM)[8] 등의 방법이 사람의 행동을 이해하고 인식하는 데 사용되어 왔습니다. 이러한 방법들은 전처리와 수작업으로 만들어진 특징에 크게 의존하기 때문에 처리하는 데 많은 시간과 리소스가 필요합니다. 또한, 다양한 데이터 세트에 맞게 잘 확장되지 않으며 실제 사례에 대한 성능이 좋지 않습니다[9]. 최근 딥러닝 방법은 학습 특징을 자동으로 추출할 수 있고 객체 감지 및 인식[10], 이미지 분류[11], 자연어 처리(NLP)[12] 등 어려운 연구 주제에 대해 유망한 결과를 보여줌으로써 연구 커뮤니티의 큰 관심을 받고 있습니다. 기존의 머신러닝 방법과 비교하여 딥러닝은 여러 개의 숨겨진 레이어를 활용하여 특정 작업에 대한 대표 특징을 자동으로 추출하는 다단계 학습 프로세스입니다[13]. 이러한 특징을 심층 특징이라고 하며, 다양한 시나리오에 맞게 합리적으로 확장할 수 있습니다. 딥러닝은 최근 HAR과 AbHAR에 적용되어 영상 감시 시스템에 매우 효율적인 것으로 입증되었습니다[14].

그림 1은 영상 감시에서 AbHAR의 주요 처리 단계에 대한 다이어그램을 보여줍니다. 일반적으로 객체 세분화, 객체 분류, 객체 추적, 동작 인식의 네 가지 주요 단계가 있습니다. 객체 세분화는 장면에서 움직이는 객체(주로 사람 대상)의 위치를 파악합니다. 객체 분류는 나중에 처리할 각 관심 객체의 유형을 식별합니다. 이 두 단계는 비디오 데이터에서 특징을 추출하는 데 크게 의존합니다. 그런 다음 객체 추적 알고리즘을 사용하여 프레임 시퀀스를 통해 관심 대상을 추적합니다. 객체 추적 방법은 대상 추적을 위한 동작 추정 및 재식별을 의미합니다. 마지막으로 행동 인식은 대표 특징을 학습하여 각 추적 객체를 대표하는 비정상적인 활동 유형을 분류하고 이해합니다. 비정상적인 행동이 감지되면 관계 기관에 알람 또는 알림이 전송됩니다.

## 2.2. 관련 작품

영상 감시에서 이상 징후 탐지는 수십 년 동안 연구되어 왔습니다. 이 주제는 다양한 응용 분야로 인해 연구 커뮤니티에서 많은 관심을 받고 있습니다. AbHAR에 대한 기존 접근 방식과 딥러닝 접근 방식 모두에 대해 방대한 양의 방법이 제안되었습니다. 사용 가능한 모든 연구를 추적하기 위해 철저한 조사가 상당한 관심을 끌고 있습니다. 1997년과 1999년에 HAR에 적용된 인간 동작 분석에 대한 리뷰[15,16]가 있었습니다. 2005년에는 자동화된 비디오 감시 시스템에 대한 조사가 [17]에서 이루어졌습니다. 3년 후, 국토안보 애플리케이션을 위한 행동 분석에 대한 연구[18]가 발표되었습니다. 2012년에는

[19]의 저자들이 다양한 맥락에서 비디오 감시 애플리케이션의 이상 징후 탐지에 대한 설문 조사를 실시했습니다.

2013년에는 비전 기반 HAR에 대한 설문조사[20]를 실시하여 HAR에 대한 기존 연구를 검토했습니다. 이 설문조사는 중요한 특징을 추출하기 위한 자체 평가 방법의 사용에 중점을 두었으며 향후 연구 방향을 제시하기도 했습니다. 2014년에 [21]의 저자들은 공공장소에서 지능형 감시 시스템에 사용되는 다양한 기술을 요약했습니다. 혼잡한 장면 분석에 대한 또 다른 검토[22]는 2014년에 수행되었습니다. 이후 2년 동안 [23]의 저자들은 비디오 스트림에 대한 다양한 HAR 방법에 집중했습니다. 1년 후, 수작업 및 심층 특징 표현을 모두 다루는 또 다른 포괄적인 검토[24]가 있었습니다. 동시에 행동 표현과 행동 모델링은 모두 [25]에서 논의되었습니다.

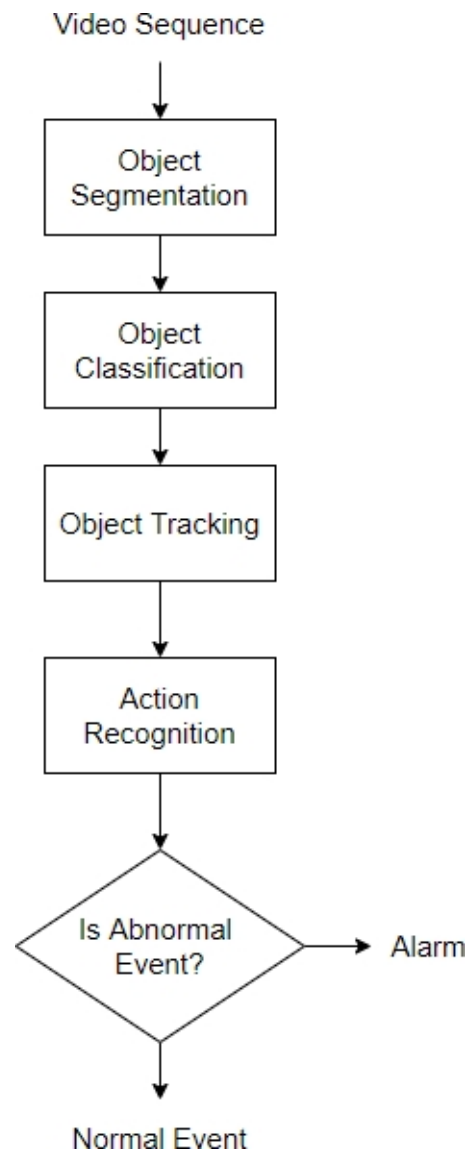


그림 1. 영상 감시의 일반적인 프로세스.

2018년에 실시한 설문조사[26]에서는 행동 감지 및 분류를 위한 RGB-D 기반 방법의 사용에 대해 다루었습니다. 저자들은 각 이상 행위 탐지 방법의 장단점에 대한 자세한 설명도 제공했습니다. 같은 해, [27]에서는 HAR과 데이터 스트림 마이닝의 관계에 대해 자세히 정리했습니다. 또한 [28]에서 비지도 및 반지도 이상 징후 탐지에 대한 논의가 있었습니다. 2019년에는 HAR에 대한 많은 조사가 있었습니다. 29]의 저자는 특징 표현, 상호 작용 인식, 사람 행동 감지 방법의 세 가지 측면에 초점을 맞춘 종합적인 검토를 수행했습니다. 30]의 저자는 비디오 감시를 위한 다양한 처리 기술을 보고했습니다. 31]의 검토에서는 다양한 유형의 2차원 및 3차원 데이터에 대한 수작업 및 심층 접근 방식에 중점을 두었습니다. 가장 최근에는 [32]의 조사에서 단일 장면 비디오 이상 탐지를 위한 다양한 기법을 자세히 설명했습니다. 33]의 저자들은 센서 기반 및 비전 기반 인간 활동 인식을 모두 논의하는 포괄적인 설문 조사를 제공했습니다.

네트워크 이상 탐지에 대한 조사[34]는 Ahmed 등이 수행했습니다. 이들은 서비스

거부, 프로브, 사용자 대 루트, 원격 대 사용자 등 네 가지 네트워크 공격 유형에 초점을 맞췄습니다. 이들은 연구 방법을 분류 기반 네트워크 이상 탐지, 통계적 이상 탐지, 정보 기반 네트워크 이상 탐지, 정보



이론과 클러스터링 기반에 대해 논의했습니다. 또한 침입 탐지 데이터 세트에 대해서도 논의했습니다. 다른 조사 [35]에서는 사이버 보안의 중요한 분야인 이상 징후 기반 네트워크 침입 탐지 시스템(NIDS)을 검토했습니다. 저자들은 [34]에서와 같이 방법을 여섯 가지 범주로 분류하고 벤치마크 데이터 세트에 대한 자세한 내용을 제공했습니다. 네트워크 이상 탐지에 대한 더 자세한 설문조사[36]는 2019년에 발표되었습니다. 이 설문조사는 크게 두 부분으로 나뉘어 진행되었습니다. 첫 번째 파트에서는 트래픽 이상 징후, 네트워크 데이터 유형, 침입 탐지 시스템을 고려한 이상 징후 탐지 배경 분석에 대해 논의했습니다. 네트워크 트래픽 이상 징후는 성격에 따른 이상 징후 분류와 인과 관계에 따른 이상 징후 분류로 구분했습니다. 또한 침입 탐지 시스템(IDS)의 두 가지 유형, 즉 모니터링 플랫폼을 사용하는 IDS 유형과 탐지 기법을 사용하는 IDS 유형에 대해 논의했습니다. 두 번째 파트에서는 이상행위 탐지 방법, 기법, 시스템에 대해 주로 논의했습니다. 이 부분에서는 통계, 클러스터링, 유한 상태 머신, 분류, 정보 이론, 진화 계산, 하이브리드/기타 방법 등이 포함되었습니다.

### 3. 벤치마크 데이터베이스

이상 감지 기능은 활동 인식을 학습하고 이해하기 위해 많은 분야, 특히 컴퓨터 비전 분야에서 광범위하게 연구되어 왔습니다. 이 작업은 실제 상황의 복잡성으로 인해 매우 어려울 수 있습니다. 비정상적인 이벤트는 무한히 많으며 모든 종류를 수집하는 것은 불가능합니다. 다행히도 과학자와 연구자가 이 작업을 수행하는 데 도움이 되는 많은 데이터베이스가 개발되었습니다. 이러한 데이터베이스는 주로 가시광선 스펙트럼에서 수집되었기 때문에 오클루전 및 조명 변화에 민감합니다. 이 섹션에서는 현재 연구자들이 행동 이상을 감지하는 데 사용하고 있는 인기 있는 데이터베이스를 간략하게 소개합니다. 데이터베이스를 가장 오래된 데이터베이스부터 최신 데이터베이스까지 시간순으로 정리했습니다. 각 데이터베이스에 대해 출시 연도, 데이터베이스 유형(예: 단일 장면 또는 다중 장면), 사용된 센서의 세부 정보(RGB 또는 열화상, 해상도, FPS), 이상 활동에 대한 설명, 샘플 이미지 등의 정보를 요약해 놓았습니다.

#### 3.1. CASIA 액션 데이터베이스

CASIA 액션 데이터베이스[37]는 2007년에 출시되었으며, 특정 위치에서 컬러 센서로 촬영한 RGB 단일 장면 데이터 세트의 일종입니다. 이 데이터베이스에는 다양한 시야 각에서 촬영된 사람의 야외 활동이 포함되어 있습니다. 걷기, 달리기, 구부리기, 점프, 웅크리기, 기절하기, 방황하기, 차에 주먹질하기 등 한 사람의 행동에는 24가지 피사체에서 8가지 유형이 있습니다. 또한 강도, 싸움, 따라가기, 따라가서 모으기, 만나서 헤어지기, 만나서 모으기, 추월하기 등 7가지 유형의 2인 상호 작용도 있습니다. 데이터베이스에는 1인 행동과 상호작용이라는 두 가지 주요 부분이 포함되어 있습니다(그림 2 참조). 아래에 설명이 나와 있습니다:

- 1인 액션:
  - 걷기: 길을 따라 걷는 피사체 한 명.
  - 달리기: 도로를 따라 달리는 피사체 한 개.
  - 구부리기: 피사체가 걷는 동안 몸을 구부리는 동작입니다.
  - 점프: 한 피사체가 도로를 따라 점프합니다.
  - 웅크리다: 한 피사체가 길을 걷다가 한 번 웅크린 모습.
  - Faint: 길을 걷던 피사체 한 명이 바닥에 쓰러지는 장면.
  - 방황하다: 방황하는 피사체.
  - 자동차를 치고 있습니다: 한 피사체가 자동차를 치는 장면입니다.
- 상호 작용:
  - Rob: 한 피사체가 다른 피사체를 강탈합니다.
  - 싸움: 두 피사체가 서로 싸우는 장면입니다.
  - 팔로우: 한 주제가 다른 주제를 끝까지 팔로우합니다.
  - 따라다니며 모입니다: 한 피사체가 다른 피사체를 따라가다가 다른 피사체와 함께 끝까지 함께 걷기.
  - 만남과 헤어짐: 두 피사체가 서로 만나고 헤어짐.

- 만나고 모이다: 두 피사체가 서로 만나서 끝까지 함께 걷기.
- 추월: 한 피사체가 다른 피사체를 추월하는 현상입니다.

동영상을 캡처하기 위해 보정되지 않은 세 대의 카메라를 수평 보기, 앵글 보기, 하향 보기 등 다양한 시야각에서 동시에 사용했습니다. 이 비디오는 25 FPS의 프레임 속도로 인코딩되었으며, avi 형식의 huffyuv 코덱으로 압축되었습니다. 비디오 프레임의 원래 공간 해상도는  $320 \times 240$ 으로 축소되었습니다. 각 비디오는 다양한 동작으로 인해 5초에서 30초까지 지속됩니다. 그림 2는 데이터베이스의 일부 이미지를 보여줍니다.

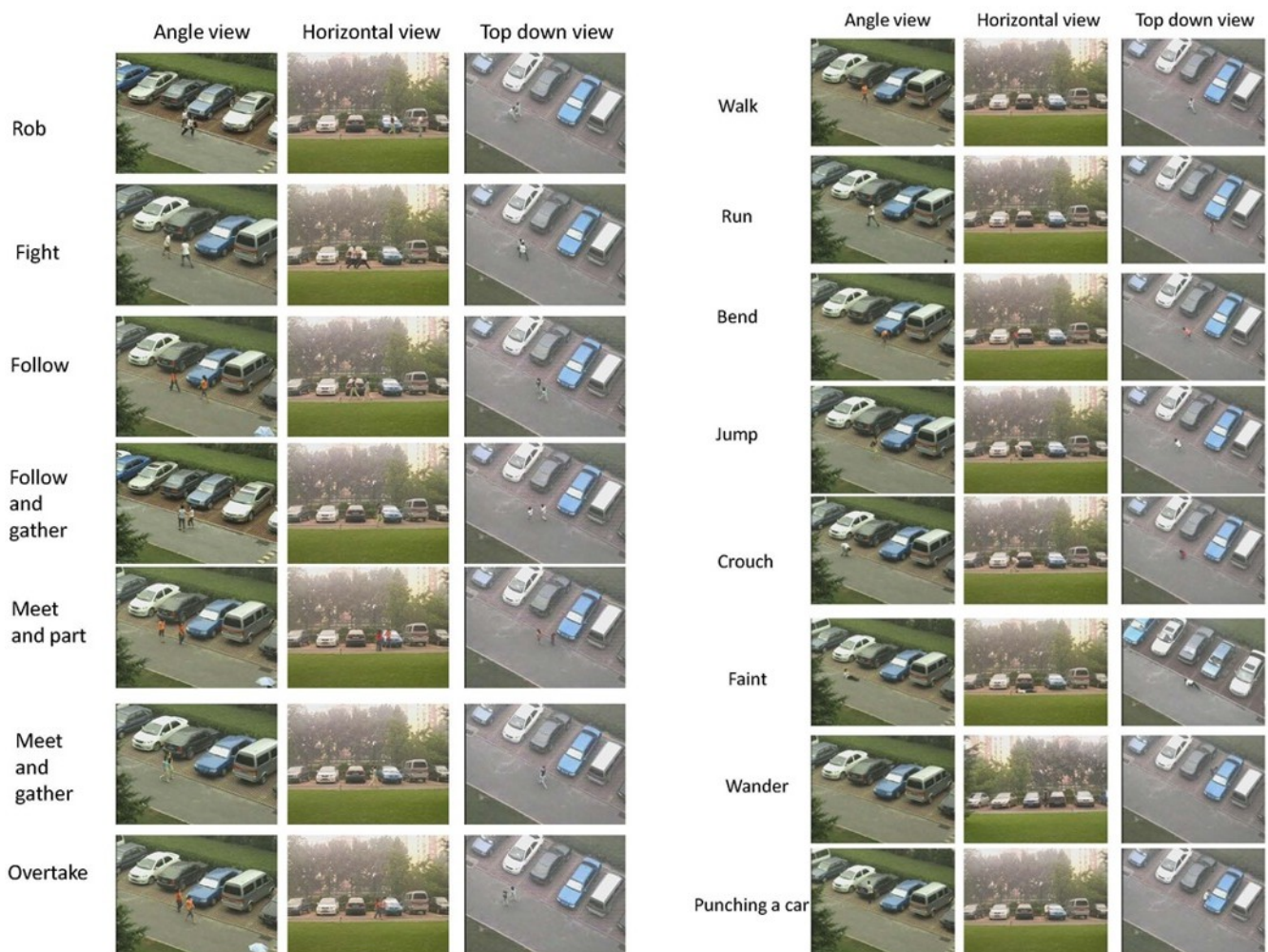


그림 2. CASIA 데이터베이스의 샘플 이미지. 왼쪽에서 오른쪽으로: 1인 동작, 2인 상호작용 [37].

### 3.2. 지하철 데이터베이스

지하철 데이터베이스는 Adam 등이 2008년 논문 [38]에서 소개했습니다. 이 역시 단일 장면 데이터베이스의 한 유형입니다. 이 데이터베이스는 지하철 입구와 출구에서 사람들을 모니터링하는 두 개의 긴 비디오 녹화로 구성됩니다(그림 3 참조). 사용 가능한 공간 실측 데이터는 없습니다. 비디오는  $512 \times 384$  해상도의 15 FPS로 그레이스케일 형식으로 녹화되었으며 총 125,475개의 프레임으로 구성되어 있습니다. 비정상적인 이벤트는 주로 방향을 잘못 잡거나, 배회하거나, 계산을 하지 않거나, 개찰구를 뛰어넘거나 비집

---

고 들어가는 사람, 벽을 청소하는 관리인 등이었습니다.



그림 3. 지하철 데이터베이스의 샘플 이미지. 왼쪽에서 오른쪽으로: 지하철 입구, 지하철 출구 [38].

### 3.3. UMN 군중 이상 데이터베이스

UMN 데이터베이스는 2009년에 공개되었으며, R. Mehran 외[39]의 "사회적 힘 모델을 이용한 비정상 군중 행동 탐지"라는 제목의 논문에서 처음 설명되었습니다. 이 비디오 시나리오의 인구가 밀집된 지역에서 배우들이 특정 위치를 돌아다니며 비정상적인 행동으로 탈출 모드로 도망치는 상황을 시뮬레이션했습니다. 따라서 이 데이터베이스는 여러 위치에서 녹화된 다중 장면 데이터베이스라고 볼 수 있습니다.

데이터베이스에는 총 11개의 짧은 동영상があり, 이 동영상은 7739프레임의 4분 17초 길이의 긴 동영상 하나로 집계됩니다. 짧은 동영상은 정상적인 동작으로 시작하여 비정상적인 동작으로 바뀝니다. 실내 장면에 대한 시나리오가 하나 있고 실외 장면이 두 개 있습니다. 모든 비디오의 프레임 속도는 30 FPS로 동일하며 고정 카메라를 사용하여 640 × 480의 해상도로 녹화되었습니다. 지상 실측 데이터는 시간적 주석입니다. 그림 4는 비디오의 일부 샘플 이미지를 제공합니다.



그림 4. UMN 데이터베이스의 샘플 이미지 [39].

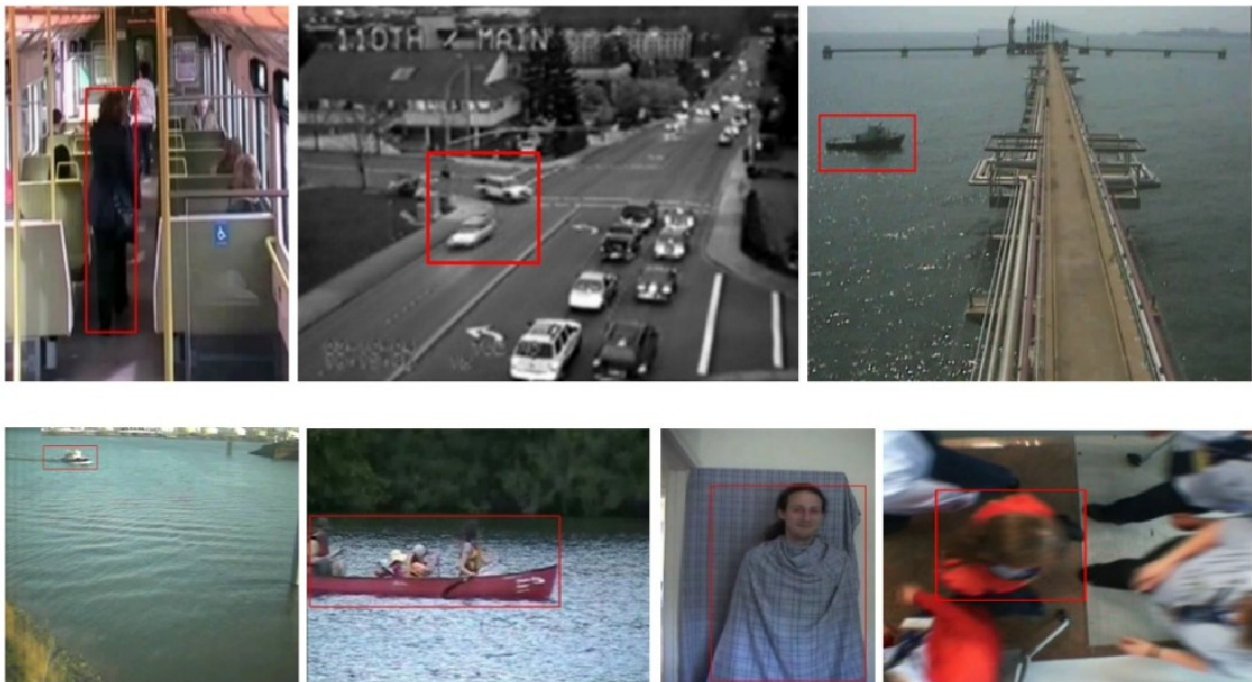
### 3.4. 비정상 행동 데이터베이스

이상 행동 데이터베이스[40]는 2010년에 요크 대학교에서 발표했습니다. 이 데이터베이스는 조명 효과, 장면의 혼란, 다양한 표적 모양, 빠른 움직임, 카메라 흔들림 등 다양한 까다로운 조건에서 녹화된 8개의 비디오(멀티씬)로 구성되어 있습니다. 데이터베이스에는 각 비디오의 특정 부분에서 비정상적인 이벤트를 감지하는 소프트웨어와 함께 시공간적 실측 자료도 함께 제공되었습니다. 이 데이터베이스의 이미지 시퀀스는 주로 공항, 강, 바다, 기차 등 특정 공공장소에서 사람과 차량의 활동에 초점을 맞추고 있습니다(그림 5 참조):

- **교통-기차:** 이 비디오는 기차에서의 일상적인 활동을 기록합니다. 조명 조건이 급격하게 변하고 카메라 흔들림이 있기 때문에 매우 까다로운 비디오입니다. 이 비디오에는 19,218 프레임의 RGB 이미지가  $288 \times 386$ 의 해상도와 25 FPS의 프레임 속도로 포함되어 있습니다. 변칙적인 이벤트는 승객의 움직임입니다.
- **벨뷰:** 이 비디오에는 교차로를 통과하는 자동차가 포함되어 있습니다. 비디오는  $320 \times 240$  해상도의 그레이스케일 형식으로 프레임 속도 10 FPS로 녹화되었으며 총 2918개의 프레임으로 구성되어 있습니다. 비정상적인 이벤트는 왼쪽 또는 오른쪽에서 도로로 진입하는 차량입니다.



- **보트-바다:** 이 비디오는 지나가는 배를 비정상적인 이벤트로 묘사합니다. 이 비디오는  $720 \times 576$  해상도의 RGB 형식으로 19 FPS의 프레임 속도로 녹화되었으며, 총 450개의 프레임으로 구성되어 있습니다.
- **보트-강:** 이 비디오는 강 위를 지나가는 보트를 비정상적인 이벤트로 보여줍니다. 이 동영상은  $720 \times 576$ 의 해상도, 5 FPS의 프레임 속도로 RGB 형식으로 녹화되었으며 총 250개의 프레임으로 구성되어 있습니다.
- **카누:** 이 동영상은 카누가 강 위를 지나가는 비정상적인 상황을 묘사합니다. 이 동영상은  $320 \times 240$  해상도, 30 FPS의 프레임 속도로 RGB 형식으로 녹화되었으며, 총 1050개의 프레임으로 구성되어 있습니다.
- **위장:** 이 비디오는 위장을 하고 걷는 사람의 모습을 보여줍니다. 올바른 동작은 정상적인 동작으로 학습되고 그 반대는 비정상적인 동작으로 학습됩니다. 이 비디오는  $320 \times 240$  해상도의 RGB 형식으로 30 FPS의 프레임 속도로 녹화되었으며 총 1629개의 프레임으로 구성되어 있습니다.
- **Airport-WrongDir:** 이 비디오는 공항에서 사람들이 줄을 서서 걷는 모습을 녹화합니다. 이 비디오는  $300 \times 300$  해상도의 RGB 형식으로 25 FPS의 프레임 속도로 녹화되었으며 총 2200개의 프레임으로 구성되어 있습니다. 비정상적인 이벤트는 사람들이 엉뚱한 방향으로 움직이는 것입니다.



**그림 5.** 비정상 행동 데이터베이스의 샘플 이미지. 왼쪽에서 오른쪽으로, 위에서 아래로: 교통-기차, 교통-벨뷰, 보트-바다, 보트-강, 카누, 위장, 공항- WrongDir [40].

### 3.5. 애비뉴 데이터베이스

이 데이터베이스[41]는 2013년에 공개되었으며, 훈련용 정상 비디오 16개와 테스트용 비

정상 비디오 21개로 구분된 37개의 비디오가 포함되어 있습니다. 데이터베이스는 RGB 단일 장면 유형입니다. 총 47개의 비정상 이벤트가 있으며, 이상한 행동, 잘못된 방향, 비정상적인 물체 등 세 가지 주요 주제로 분류됩니다. 이 비디오는 홍콩중문대 캠퍼스 도로에서 촬영되었으며, 총 30,652개 프레임(훈련 15,328개, 테스트 15,324개)이 포함되어 있습니다. 각 이미지 시퀀스의 해상도는  $640 \times 360$ , 프레임 속도는 25 FPS입니다. 작성자는 시간 및 공간 주석을 모두 제공했습니다. 세 가지 주요 비정상 이벤트 유형(그림 6 참조)이 아래에 정의되어 있습니다:

- 이상한 행동: 달리기, 물건 던지기, 배회하기 등의 행동.
- 잘못된 방향: 사람들이 잘못된 방향으로 이동합니다.



- 비정상적인 물체: 자전거와 같은 이상한 물체를 들고 다니는 사람.



그림 6. Avenue 데이터베이스의 샘플 이미지. 왼쪽에서 오른쪽으로: 달리기, 물건 던지기, 배회 [41].

### 3.6. UCSD 이상 징후 탐지 데이터베이스

이 데이터베이스[42]는 2013년에 출시되었으며 두 개의 하위 데이터셋을 포함합니다: 보행자 1과 보행자 2. 둘 다 보행자 1의 경우  $238 \times 158$ , 보행자 2의 경우  $360 \times 240$  해상도로 10 FPS로 기록된 회색조 이미지 시퀀스를 포함합니다. 각 데이터 세트는 단일 장면 데이터 세트입니다. 두 데이터셋 모두 정상적인 행동만 포함된 훈련 비디오와 비정상적인 이벤트가 포함된 테스트 비디오가 있습니다. 데이터 세트를 녹화하기 위해 정적 카메라를 사용했으며 보행자 통로가 내려다보이는 엘리베이터에 설치했습니다. 비정상 이벤트에는 다음이 포함됩니다:

- 자전거 타는 사람, 스케이트 타는 사람, 소형 카트 등 보행자가 아닌 물체의 보행로 통행이 금지됩니다.
- 산책로를 가로질러 건너거나 산책로를 둘러싼 풀밭을 걷는 사람과 같은 비정상적인 보행자 움직임 패턴.

보행자 1(Ped1) 데이터 세트에는 34개의 정상 훈련 비디오와 36개의 비정상 테스트 비디오가 포함되어 있으며, 이 비디오에는 카메라를 향해 건너거나 카메라에서 멀어지는 사람들의 그룹이 포함되어 있습니다. 이러한 비정상 사례는 주로 자전거와 자동차가 군중 속으로 들어오는 등 비정상적인 차량과 관련이 있습니다. 일부 샘플 이미지는 그림 7에 나와 있습니다.



그림 7. 비보행자 엔티티가 있는 Ped1 하위 데이터 집합 [42].

보행자 2(Ped2) 데이터 세트에는 16개의 훈련 비디오와 12개의 비정상 이벤트가 포

함된 12개의 테스트 비디오가 포함되어 있습니다. Ped2에는 카메라 평면과 평행하게 보행자가 움직이는 장면이 포함됩니다. Ped2에 대한 이상 이벤트의 정의는 Ped1과 동일합니다. Ped2 데이터 세트의 몇 가지 예시 이미지가 그림 8에 나와 있습니다.



그림 8. 비보행자 엔티티가 있는 Ped2 하위 데이터셋 [42].

### 3.7. 상하이텍 캠퍼스 데이터베이스

이 데이터베이스[43]는 2016년에 공개되었습니다. 여기에는 정상 이벤트만 포함된 330개의 훈련 비디오와 130개의 비정상 이벤트가 포함된 107개의 테스트 비디오가 포함되어 있습니다. 총 프레임은 317,398개이며, 불규칙 프레임은 17,090개입니다. 데이터베이스는 보행자 통로를 내려다보는 해상도  $856 \times 480$ , 24 FPS의 RGB 카메라를 사용하여 수집했습니다. 주로 이상한 물체, 잘못된 방향, 이상한 동작과 관련된 복잡한 조명 조건, 카메라 각도, 다양한 이상 유형이 포함된 13개의 장면(멀티씬)으로 구성되어 있습니다(그림 9 참조).

- 이상한 행동: 달리기, 훔치기, 밀기, 점프하기, 울타리 뛰어넘기, 물건 떨어뜨리기, 던지기, 싸우기 등의 행동. 아래는 몇 가지 샘플 이미지입니다.
- 잘못된 방향: 일반적으로 사람들은 정상적인 방향을 따르지만 누군가는 그 반대의 방향을 따르는 경우가 있습니다.
- 비정상적인 물체: 자전거나 유모차 등 이상한 물체를 가지고 다니는 경우입니다.



그림 9. 애비뉴 데이터베이스를 구성하는 샘플 이미지. 왼쪽에서 오른쪽으로: 강도, 물건 던지기, 싸움 [43].

### 3.8. UCF-범죄 데이터베이스

UCF-Crime 데이터베이스는 2018년에 출시되었습니다[2]. 이 데이터베이스는 여러 위치(멀티씬)에서 여러 대의 RGB 카메라로 촬영한 128시간 분량의 1900개 인터넷 동영상을 편집한 것입니다. 비정상적인 사건에는 학대, 체포, 방화, 폭행, 교통 사고, 강도, 폭발, 싸움, 강도, 총격, 절도, 절도, 기물 파손 등이 포함됩니다. 이 비디오는 13가지 실제 상황을 다루며, 13가지 그룹 활동의 이벤트 인식과 각 특정 그룹의 이상 감지라는 두 가지

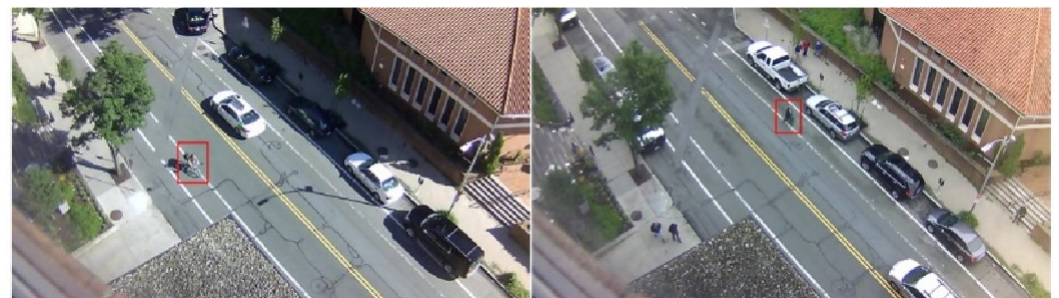
주요 작업에 사용할 수 있습니다. 저자들은 시간적 주석만 제공했습니다. 그림 10은 데이터베이스의 일부 샘플 이미지를 보여줍니다.



**그림 10.** UCF 범죄 데이터베이스의 샘플 이미지. 왼쪽에서 오른쪽으로: 학대, 싸움, 절도 [2].

### 3.9. 거리 장면 데이터베이스

스트리트 씬 데이터베이스는 2020년에 출시된 가장 최신 데이터베이스입니다[44]. 이 데이터 세트는 RGB 단일 장면 유형입니다. 총 203,257개의 이미지 시퀀스가 있으며, 원본 비디오에서 15 FPS의 프레임 속도로 추출되었습니다. 데이터베이스는 무단횡단, 자전거 차선 이탈, 배회 등 205개의 비정상적인 이벤트로 구성되어 있습니다.  $1280 \times 720$ 의 해상도로 녹화되었습니다(그림 11 참조). 이 데이터베이스는 다양한 활동, 오클루전, 그림자, 움직이는 배경 등으로 인해 까다로운 작업입니다.



**그림 11.** 거리 장면 데이터베이스의 샘플 이미지. 왼쪽에서 오른쪽으로: 차선 밖 자전거 타는 사람, 무단횡단 [44].

## 4. 비디오 이상 징후 탐지를 위한 데이터 사전 처리

논의된 벤치마크 데이터베이스는 비디오와 이미지의 원시 데이터 모음일 뿐입니다. 이러한 데이터는 ML 알고리즘에 입력하기 전에 사전 처리를 거쳐야 합니다. 데이터 준비와 기능 엔지니어링은 영상 감시 시스템의 성능에 큰 영향을 미치는 두 가지 중요한 프로세스로 간주됩니다. 비전 기반 도메인의 경우 이러한 프로세스에는 배경 구성, 전경 추출, 특징 표현 추출과 같은 여러 단계가 포함됩니다. 특징 엔지니어링 프로세스의 주요 목적은 노이즈를 줄이고, 중요한 표현 특징을 선택하고, 중요한 정보를 잃지 않고 고차원 특징을 하위 공간 도메인으로 변환하고, 과적합 문제를 줄이는 것입니다. 그러나 조명 조건의 변화, 복잡한 배경, 오클루전 또는 피사체 간의 가짜 상호 작용과 같은 많은 문제가 있습니다.

### 4.1. 세분화

데이터 전처리의 첫 번째 단계는 세분화입니다. 세그멘테이션은 이미지나 동영상에서 대상 피사체를 추출하는 데 사용됩니다. 세분화에는 배경 구성과 전경 추출 기술이 포함됩니다. 배경 구성 알고리즘은 장면의 글로벌 대표 특징을 파악하여 배경 정보를 모델

링하고, 식별된 피사체는 현재 프레임과 구성된 배경의 차이에 따라 분석됩니다[45]. 배경 구성 기법은 장면에서 빠른 물체를 추적하는 데 사용되며 고정 카메라와 함께 사용할 때 컴퓨터 효율적입니다. 46-48]과 같은 일부 통계적 배경 구성 방법은 적응형 파라미터가 있는 다중 모드 영역에서 잘 작동할 수 있지만, 환경 소음이나 열악한 조명 조건으로 인해 성능이 크게 저하됩니다. 45,49,50]에서 보고된 신경망 기반 기법은 이러한 한계를 극복할 수 있지만, 쉽게 오버핏됩니다.

데이터. 기술 선택은 주로 애플리케이션의 목적에 따라 달라집니다. 움직이는 카메라로 동적 배경을 녹화하려면 전경 추출 기반 세그멘테이션 알고리즘을 사용해야 합니다. 비디오 시퀀스에서 공간 및 시간 정보를 모두 분석하여 배경에서 대상 피사체를 추출합니다. [51,52]의 저자들은 광학 흐름 기법을 사용하여 폐색과 왜곡을 처리하고 카메라를 이동하여 비디오 녹화에서 대상을 추출하는 데 성공했지만, 그 방법은 복잡하고 시간이 많이 걸립니다. 또한 [53,54]에서는 노이즈에 민감하지만 낮은 컴퓨팅 성능을 요구하는 전경 분할을 수행하기 위해 시간 정보를 사용하기도 했습니다. 참고 문헌 [55,56]에서는 마르코프 랜덤 필드를 사용하여 경계를 보존하고 복잡한 배경을 처리했지만 이러한 방법은 컴퓨터 효율적이지 않습니다.

비디오 이상 탐지를 위해 제안된 많은 방법들은 배경과 전경 정보를 모두 활용하려고 시도했습니다. Lai 등[57]은 미래 프레임과 RGB 차이를 생성하기 위해 두 개의 디코더를 포함하는 네트워크를 도입했습니다. Doshi 등[58]은 사전 학습된 객체 감지 모델(YOLO)을 사용하여 비디오에서 객체를 감지하여 위치 및 외관 특징을 캡처했습니다. 반면, Cai 등[59]은 이미지 클립과 그 광학 흐름 클립을 입력으로 사용하여 구조와 움직임 정보를 캡처했습니다. 그런 다음 두 개의 디코더가 두 개의 인코더에서 추출한 융합된 특징을 사용하여 미래 프레임과 광학 흐름 이미지를 생성했습니다.

#### 4.2. 특징 추출 및 선택

수작업 특징 기반 추출[60,61]은 인간 행동을 이해하는 데 유용한 특징을 추출하는데 사용되었습니다. 이러한 방법은 특정 조건으로 제한되며 새로운 환경에 적응할 수 있는 유연성이 부족합니다. 또한 시간이 많이 걸리고 컴퓨터 비효율적입니다. 연구자들은 로컬, 글로벌, 시맨틱 피처의 세 가지 유형으로 분류되는 새로운 대표 피처로 눈을 돌렸습니다. 이러한 특징들은 노이즈와 동적 환경에 대한 장점과 견고함을 명확하게 보여주었습니다.

- 지역 대표 기능은 지역 설명자 알고리즘을 사용하여 이미지의 입력 영역이 지역적으로 정량화되는 방식을 관리합니다. 이 알고리즘은 이미지에서 영역의 로컬리티를 고려하여 개별적으로 설명합니다. [60,62,63]에서 보고된 바와 같이 HOG(기울기 히스토그램)는 이미지의 기울기 크기와 방향에 대한 로컬 설명을 추출하는 기본 기법입니다. HOG는 광도 변환에 변하지 않지만 고정된 크기의 사람 감지에만 사용할 수 있습니다. 스케일 불변 특징 변환(SIFT)은 [64,65]에서 사용되었으며 3D 투영에서도 기하학적 및 광도 변환에 불변성을 보였지만 고차원 특징을 포함하고 있어 컴퓨터가 비효율적이고 실시간 애플리케이션에 적합하지 않습니다. 속도 향상 강인한 특징(SURF) 알고리즘[66,67]은 더 빠르고 감지 포인트의 품질을 유지하는 SIFT의 대안입니다. 마지막으로, [68,69]의 형상 기반 로컬 특징 기술자는 대상 피사체의 가장자리 구조를 보존함으로써 노이즈에 대한 견고성을 입증했습니다. 그러나 전처리 단계



에서 실루엣 분할에 크게 의존한다는 단점이 있습니다.

- 글로벌 대표 특징은 입력 이미지가 전역적으로 정량화되는 방식을 관리하는 이미지 설명자를 사용하여 이미지 내용을 추상적으로 나타내는 특징 벡터를 반환합니다. [31,70]의 전역 기술자는 모서리, 가장자리, 용기, 광학적 흐름의 세부 정보를 필수 특징으로 인코딩했습니다. 이러한 특징들은 카메라 깊이에서 쉽게 얻을 수 있지만 장면에 따라 다르며 일반적인 정보가 부족합니다. [71-73]의 일부 연구자들은 3D 공간 볼륨을 사용하여 배경 차감과 무관한 3D 전역 특징 벡터를 추출했습니다. 그러나 이러한 3D 특징은 노이즈와 오클루전에 매우 민감했습니다. [74,75]에서는 공간 특징을 주파수 특징으로 변환하기 위해 이산 푸리에 변환(DFT)도 사용되었지만, 역 프로세스는 비정상적인 대상 피사체를 식별하는 데 필요한 공간 및 시간 정보를 잃을 수 있습니다.
- 의미적 특징은 비디오 시퀀스에서 인체의 자세를 분석하여 얻을 수 있습니다. 이러한 특징은 클래스 간 변화에 강한 인간의 포즈 정보[76,77]로 변환할 수 있지만, 포즈 정확도를 달성하기는 어렵습니다. 연구자들은



[78,79]에서는 텍스처와 색상과 같은 외관 기반 특징을 사용하여 컨텍스트 정보를 얻었지만, 이는 클래스 내 변화에 민감했습니다. 3차원 시맨틱 특징은 기하학적 정보와 시각적 정보를 모두 제공하는 RGD-D 카메라[68,80]에서 얻을 수 있지만, 노이즈와 오클루전 문제의 영향을 많이 받습니다.

## 5. 딥러닝 방법

이 섹션에서는 AbHAR에서 비디오 감시를 위해 사용하는 딥 러닝 방법에 대해 설명합니다. 딥 러닝 방법은 행동 인식과 영상 이해 측면에서 상당한 성과를 거두었습니다. 딥 러닝 시스템은 공간 및 시간 정보를 모두 포함하는 이미지 시퀀스에서 AbHAR 프로세스를 사용하여 자동으로 학습하고 대표적인 특징을 추출할 수 있습니다. 이러한 딥러닝 방법은 다양한 전략, 프로세스 기법, 네트워크 아키텍처의 최종 목표에 따라 재구성 기반 방법, 다중 클래스 분류 방법, 미래 프레임 예측 방법, 스코어링 방법의 네 가지 주요 그룹으로 분류할 수 있습니다. 표 1은 이상 징후 탐지에 사용되는 딥 러닝 기반 기법을 카테고리별로 요약한 것입니다.

표 1. 이상 징후 탐지를 위한 딥 러닝 기반 기술 요약.

방법	유형	연도	참조
컨볼루션 장단기 메모리	RB + 미래 프레임 예측	2016	[4]
2D 컨볼루션 자동 인코더	RB	2016	[81]
스파스 자동 인코더	재구성 기반	2016	[82]
느린 특징 분석 + 심층 신경망	특점	2016	[9]
스파스 디노이즈 자동 인코더	다중 클래스 분류	2017	[83]
자동 인코더 + 캐스케이드 딥 CNN	다중 클래스 분류	2017	[3]
시공간 자동 인코더	RB + 미래 프레임 예측	2017	[84]
사전 학습된 DNN + 가우시안 분류기	다중 클래스 분류	2018	[85]
자동 인코더 + 로우 레벨 기능	재구성 기반	2018	[86]
다중 인스턴스 학습	특점	2018	[2]
로우 레벨 기능 + 자동 인코더	재구성 기반	2018	[86]
GAN을 사용한 프레임 예측	미래 프레임 예측	2018	[43]
기존 기능과 심층 기능의 조합	특점	2019	[87]
현지화 특징 추출	특점	2019	[88]
AnomalyNet	재구성 기반	2019	[89]
광학 흐름 + 다중 인스턴스 학습	특점	2019	[90]
소셜 포스 맵 + 다중 인스턴스 학습	특점	2019	[91]
주의 모듈 + 자동 인코더	재구성 기반	2019	[92]
구성 요소 분석 + 전이 학습	다중 클래스 분류	2019	[93]
SSD + 자동 인코더를 사용한 물체 감지	다중 클래스 분류	2019	[94]
스파스 코딩 심층 신경망	특점	2019	[95]
적응형 프레임 내 분류 네트워크	분류	2019	[96]
자동 인코더 + 가우시안 혼합 모델	특점	2020	[97]

### 5.1. 재구성 기반 방법

이것은 AbHAR에서 가장 일반적으로 사용되는 방법입니다 [82,92]. 공식적으로  $x$ 는 입력 이미지 시퀀스이고  $f$ 는  $x$ 를 재구성하는 신경망이라고 가정합니다. 재구성 비용 함수  $\Theta$ 는 원래 입력  $x$ 와  $f(x)$  사이의 오차  $e$ 를 계산하는 함수로 정의할 수 있습니다(방정식 (1) 참조).

$$e = \Theta(x, f(x)) \quad (1)$$

재구성 오류 함수  $\Theta$ 에 사용되는 널리 사용되는 신경망은 오토엔코더 네트워크입니다. 자동 인코더 네트워크는 인코더와 디코더의 두 부분으로 구성되며, 그 개략적인 구조

는 그림 12를 참조하세요. 인코더는 입력  $X$ 를 간결하고 변별력 있는 대표 피쳐  $z$ 를 포함하는 잠재 피쳐로 인코딩할 수 있는 능력을 가진 신경망이며, 디코더는 이러한 대표 피쳐를 원래 형태인  $X'$ 로 디코딩할 수 있는 능력을 가진 신경망이기도 합니다. 이 네트워크는  $X$ 와  $X'$  사이의 비용 함수  $\Theta$ 에 의해 계산된 오류 값  $e$ 를 줄이도록 훈련됩니다.

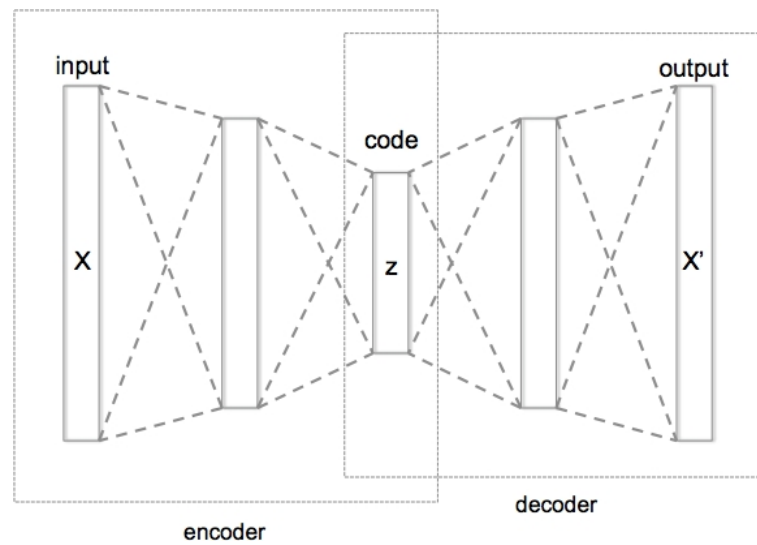


그림 12. 자동 인코더 회로도 구조 [98].

많은 접근 방식이 이상 징후를 식별하기 위해 자동 인코더 구조를 활용했습니다. [81]의 연구에서는 입력과 출력 사이의 시공간 정보를 보존하기 위해 스파스 코딩 자동 인코더를 사용했습니다. 2D 컨볼루션 네트워크를 사용하여 비디오의 각 세그먼트에서 그레이 스케일 2D 이미지 시퀀스를 인코딩했습니다. 저자들은 공간 및 시간 영역 모두에서 더 나은 재구성을 위해 시간적 차원에서는 1채널 입력 이미지 스택킹만 사용했습니다. [4]의 연구도 재구성 기반 방법을 AbHAR에 대한 접근 방식으로 사용했습니다. 그러나 이들은 컨볼루션 장단기 메모리(Conv-LSTM) 네트워크를 주요 아키텍처로 사용하여 시공간적 정보를 강화했습니다. LSTM 네트워크 [99]는 영상 데이터의 장기 정보를 포착할 수 있어 예측 정확도를 높일 수 있습니다. [86]의 연구자들은 에지 및 광학 흐름과 같은 낮은 수준의 특징을 사용하여 원시 프레임에 추가 정보를 추가했습니다. 자동 인코더의 또 다른 용도는 [82]에서 제안되었는데, 저자들은 작업에 두 가지 다른 자동 인코더를 사용했습니다. 첫 번째는 일반 자동 인코더이고 두 번째는 차원 특징을 제한하고 잠재 계층에서 가장 유용한 활성 뉴런을 유지하는 스파스 자동 인코더였습니다(그림 13 참조).

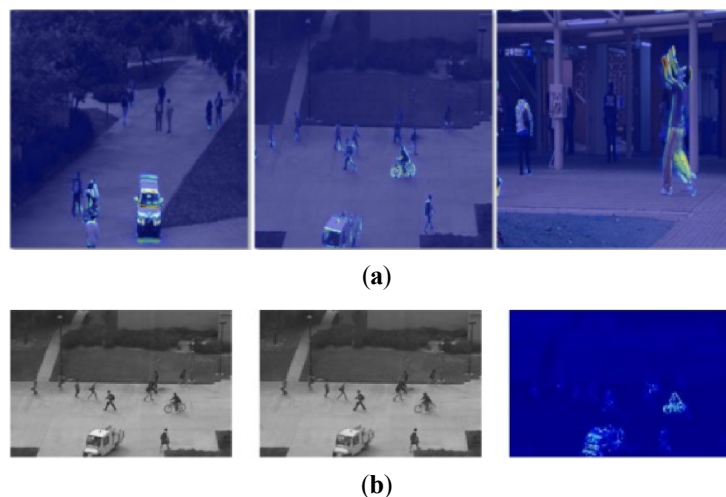
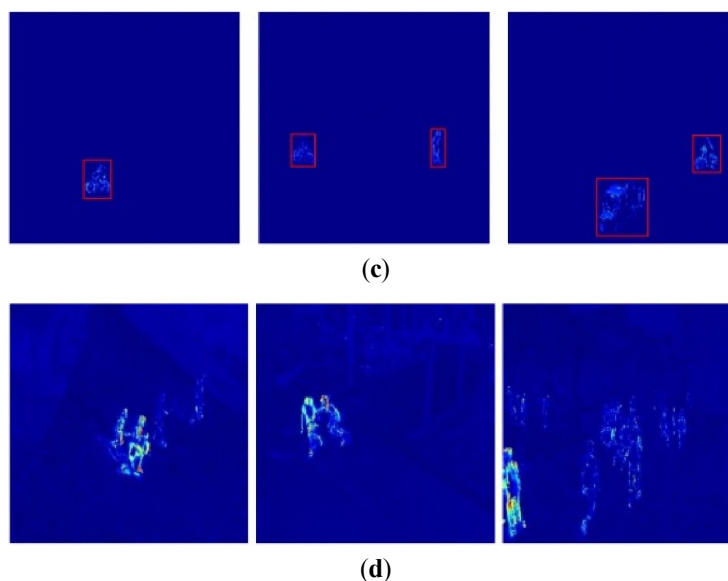


그림 13. 계속.



**그림 13.** 최근 방법의 오류 맵. (a) 제로샷 크로스 도메인 비디오 이상 탐지(zxVAD) - Aich 외. [100] (2023), (b) 동적 로컬 집계 네트워크 - Yang 외. [101] (2022); (c) 시공간적 기억-Liu 등 [102] (2022); (d) 하이브리드 주의 및 동작 제약-Zhang 등 [103] (2022).

[84]에서 저자들은 3D 컨볼루션 자동 인코더를 사용하여 시간 정보를 보존하고 시간 차원의 공간 특징을 추적했습니다. 또한 훈련 샘플 수를 늘리기 위해 데이터 증강을 적용했습니다. 이후 [89]의 저자들은 이미지의 움직임과 외형 특징을 모두 결합한 특징 학습 프레임워크를 위한 새로운 네트워크, 즉 SC2NET을 제안했습니다. 이 네트워크는 희소성 손실을 계산할 수 있으며 구성 오류를 통해 훈련되어 유용한 시공간적 특징을 구성하는 방법을 학습할 수 있었습니다.

자동 인코더를 사용하는 접근 방식은 네트워크가 비정상 인스턴스에 대해 높은 구성 오류 점수를 반환할 것이라는 가정을 기반으로 합니다. 그러나 경우에 따라 자동 인코더가 정상 인스턴스뿐만 아니라 비정상 인스턴스도 일반화할 수 있기 때문에 이 가정이 반드시 맞는 것은 아닙니다. 이는 재구성 오류 점수가 예상보다 낮다는 것을 의미합니다. [92]의 연구자들은 이 문제를 해결하기 위해 새로운 접근 방식을 제안했습니다. 그들은 각 인코딩 특징을 디코더 네트워크에 대한 쿼리로 취급했습니다. 모든 정상 인코딩 특징은 메모리에 저장되었습니다. 디코더는 각 쿼리 인스턴스에 대해 메모리에서 가장 가까운 정상 인코딩을 반환했습니다. 따라서 비정상 인스턴스는 가장 가까운 정상 인코딩을 매핑할 수 없으며 높은 재구성 오류를 반환합니다(표 2 참조).

**표 2.** 최근 비디오 이상 징후 탐지에 사용되는 방법 비교.

응답		고이 1	고이 2	응이이네그	년도	참고
센서 2023. 23. 564	인지 기억 증강 네트워크(CMAN)	-	96.2	-	2021	[104]
	단일 및 다중 프레임 이상 징후 탐지	-	97.5	87.2	2021	[105]
	오토인코더의 멀티 레벨 메모리 모듈					
	연결 건너뛰기(ML-MemAE-SC) 포함	-	99.3	91.1	2021	[106]
	메모리 모듈이 있는 자동 인코더(AMM)	-	97.2	87.9	2021	[107]
	이상 징후 탐지에 대한 설명	73.1	80.1	-	2021	[108]
	주의 기반 적대적 자동 인코더(A3N)	90.7	97.7	89.4	2022	[109]
	AD를 위한 그룹 활동	84.4	95.0	82.3	2022	[110]
	변형 이상 탐지 네트워크(VADNet)	-	96.8	87.3	2022	[111]

표 2. 계속

방법	Ped 1	Ped 2	Avenue	상하이테크	Year	참조
컨텍스트 관련 비디오 이상 감지	-	96.3	87.1	73.6	2022	[112]
현지화 기반 재구성(LBR)	81.1	97.2	92.8	72.6	2022	[113]
전경-배경 상호 분리						
생성적 적대 네트워크 (FSM-GAN)	-	98.1	80.1	73.5	2022	[114]
듀얼 스트림 메모리 네트워크	-	98.3	88.6	75.7	2023	[115]
주의 기반 잔여 자동 인코더	-	97.4	86.7	73.6	2023	[116]
양방향 프레임 보간	-	98.9	89.7	75.0	2023	[117]
제로 샷 크로스 도메인 비디오 이상 징후 감지(zxVAD)	78.6	95.8	83.2	-	2023	[100]

### 5.2. 다중 클래스 분류기 방법

이상 징후 탐지를 위한 주요 방법으로 오류 재구성을 사용하는 것이 널리 사용되고 있음에도 불구하고, 다양한 접근 방식이 이상 감지 문제를 다중 클래스 분류로 구성했습니다. 이 분류 방법은 비디오 세그먼트를 입력  $x$ 로 수신하고 재정의된 카테고리의 클래스 레이블을 나타내는 출력  $y$ 를 반환합니다(방정식 (2) 참조).

$$y = f(x), y \in R \quad (2)$$

이상 징후 탐지의 주요 문제 중 하나는 데이터 세트의 불균형이며, [3]의 접근 방식은 캐스케이드 프레임워크를 사용하여 간결하고 강력한 특징을 학습함으로써 이 문제를 해결하려고 시도했습니다. 이 접근 방식에는 두 단계가 있는데, 첫 번째 단계는 자동 인코더 스택을 사용하여 정상 비디오 패치에서만 학습합니다. 두 번째 단계는 컨볼루션 신경망으로, 첫 번째 단계에서 모델링할 수 없거나 추가 조사가 필요한 비디오 패치를 처리합니다. 최종 특징들은 가우시안 분류기에 입력되어 분류를 수행합니다. [83]에서 볼 수 있는 또 다른 접근 방식은 비디오 세그먼트의 대표 특징을 학습하기 위해 로컬 및 글로벌 설명자를 모두 사용했습니다. 이러한 특징들은 각 비디오 패치의 공간적 정보와 시간적 정보로 구성되었습니다. 로컬 디스크립터의 경우 유사성 메트릭을 사용하여 시공간적 특징을 추출했습니다. 한편, 글로벌 디스크립터는 사전 학습된 자동 인코더 네트워크를 활용하여 잠재적 특징을 학습했습니다. 그런 다음 이러한 특징을 다른 자동 인코더에 넣어 가장 대표적인 특징을 선택하고, 선택된 특징을 가우시안 분류기에 입력해 이상 징후를 감지했습니다. [85]의 저자들도 가우시안 분류기를 사용하여 예측을 수행했지만, 사전 학습된 모델을 활용하고 중간 특징을 추출하여 분류기가 이상 징후를 탐지할 수 있는 강력한 패턴을 찾았습니다. 분류기가 결정을 내릴 수 없는 경우에는 사전 학습된 모델의 최상위 계층으로 특징을 전송하여 더 많은 판별 특징을 학습했습니다.

[93]의 연구에서도 사전 학습된 모델을 사용하여 대표적인 특징을 학습했습니다. 이들은 일반화를 위해 전이 성분 분석 기법[118]을 사용하여 다양한 영역에서 컨볼루션 신경망의 특징 공간을 탐색했습니다. [94]에서 제안한 방법은 비디오의 각 프레임에 대해 단

일 샷 검출기 모델(SSD)[119]을 사용하여 장면의 각 사람 객체에 초점을 맞췄습니다. 장면에서 각 대상 피사체를 찾은 후, 두 개의 서로 다른 자동 인코더 모델을 사용하여 각 피사체의 움직임과 외모 특징을 모두 학습했습니다. 이러한 특징들은 K-평균 기법[120]을 사용하여 서로 다른 하위 집합으로 연결되고 클러스터링되었습니다. 각 클러스터는 한 종류의 정상 이벤트를 나타냅니다. 각 클러스터의 분류를 수행하기 위해 1 대 나머지 분류기를 학습시켰습니다. 분류기로부터 음수 점수를 받고 어떤 클러스터에도 속하지 않는 이벤트는 비정상 이벤트로 태그가 지정되었습니다. 참고. [96] 또한 각 비디오 패치의 광학적 흐름과 그래데이션의 모든 대표 특징을 결합하여 이상 감지 문제를 다중 클래스 분류 문제로 구성했습니다. 이를 통해 저자는 움직임과 모양 특징을 모두 학습하는 적응형 프레임 내 분류 네트워크를 구축하여 비디오 패치에서 이상 징후를 감지하고 위치를 파악할 수 있었습니다.

### 5.3. 미래 프레임 예측 방법

재구성 기반 방법의 한 가지 문제점은 자동 인코더가 정상 인스턴스뿐만 아니라 비정상 인스턴스도 실수로 재구성할 수 있다는 점입니다. 이로 인해 많은 오탐이 발생하고 성능이 저하될 수 있습니다. [29]의 연구에서는 이 문제를 잘 처리할 수 있는 미래 프레임 예측이라는 새로운 방법을 제안했습니다. 공식적으로  $x_t$ 가 시간 단계  $t$ 의 입력 비디오 세그먼트라고 가정할 때, 미래 프레임 예측 방법은  $t + 1$  시점의 다음 세그먼트 프레임을 예측하고 예측된 프레임과 현재 프레임 간의 오류 비용을 비교하는 함수  $p$ 를 제공합니다. 오류 값이 정의된 임계값보다 크면 해당 프레임은 비정상 인스턴스로 태그가 지정됩니다. 방정식 (3)은 이 접근 방식에 대한 수학적 형태를 보여줍니다.

$$x_{t+1} = p(x)_t \quad (3)$$

이 접근 방식에는 생성적 적대 신경망(GAN)[43]이라는 신경망이 사용됩니다. 기본적으로 GAN은 생성기와 판별기의 두 가지 주요 부분으로 구성된 생성 모델입니다. 제너레이터는 학습 데이터의 통계 정보를 기반으로 새로운 인스턴스 데이터를 생성하는 데 사용됩니다. 판별기의 역할은 입력이 제너레이터에서 나오는 것인지(가짜), 아니면 학습 데이터에서 나오는 것인지(진짜)를 확인하는 것입니다. 참고. [121]은 접근 방식에 GAN을 사용했습니다. 제너레이터는 미래 프레임을 예측하기 위해 U-Net 아키텍처[122]를 사용하는 신경망이었습니다. 저자들은 이미지 간 번역 작업에서 상당한 성능을 보여 주었기 때문에 U-Net 모델을 선택했습니다. 판별기는 비정상적인 이벤트가 있는 프레임을 판별하는 데 사용되는 신경망이었습니다.

재구성 기반 방법에 대한 일부 이전 연구에서도 이러한 미래 프레임 예측 접근 방식을 활용했습니다. 예를 들어, [81]의 접근 방식은 비디오 세그먼트에서 공간 및 시간 정보를 모두 인코딩하고 중심 프레임이 주어지면 미래 프레임을 추가로 예측했습니다. 참조. [84] 및 ref. [4]와 같은 일부 초기 연구에서도 미래 프레임 예측 아이디어를 활용하여 현재 프레임을 재구성했습니다. 참고 문헌 [84]에서는 두 개의 분리된 분기를 사용하여 미래 프레임을 예측하고 동시에 프레임을 재구성하는 방법을 학습했습니다. 저자는 손실 값을 사용하여 시간 정보를 추출하기 위해 네트워크를 훈련했습니다. 손실 값은 예측 손실과 재구성 손실을 모두 합한 값입니다. [4]의 접근 방식도 병렬 분기를 사용하여 프레임 예측과 재구성을 모두 학습했습니다. 그러나 이 네트워크는 비디오 내에서 관심 지점을 식별하여 이상 징후를 감지하는 데 유용한 정보를 추출할 수 있었습니다.

### 5.4. 채점 방법

점수 매기기 방법에서 네트워크는 각 비디오 세그먼트에 대한 이상 점수를 예측하려고 시도합니다. 따라서 비정상적인 인스턴스에 대해 높은 점수 값(또는 때로는 낮은 점수 값)을 할당하는 것이 목적인 회귀 문제로 간주할 수 있습니다. 수학적으로, 점수 매기기



방법은 함수  $s$ 를 사용하여 입력 비디오 세그먼트  $x$ 를 취하고 각 세그먼트의 이상 점수를 나타내는  $t$  값을 할당합니다(방정식 (4) 참조).

$$t = s(x), t \in \mathbb{R} \quad (4)$$

[9]의 접근 방식은 추출된 특징에서 이상 점수를 계산하기 위해 합제곱 미분을 사용했습니다. 이러한 특징은 이미지 시퀀스에서 의미적 의미를 학습하는 데 사용된 심층 및 저층 특징 분석[123]으로 구성되었습니다. [2]의 연구에서는 다중 인스턴스 학습 기법을 사용하여 레이블이 약하게 지정된 비디오 세그먼트에서 이상 점수를 측정했습니다. 이들은 3D 컨볼루션 신경망[124]을 활용하여 이미지 시퀀스에서 공간 및 시간 정보를 모두 학습했습니다. 시공간적 특징이 네트워크 모델에 입력되었습니다. 이 모델은 각 비디오 세그먼트의 최종 점수를 감지하기 위해 약한 레이블을 사용하여 지도 방식으로 학습되었습니다. 다중 인스턴스 학습을 사용한다는 아이디어는 [90]의 연구에서 영감을 얻었는데, 이 연구에서는 비디오 세그먼트에서 유망한 특징을 결정하기 위해 주의 메커니즘[125]과 함께 광학적 흐름 특징을 사용하여 최종 점수를 계산했습니다. 마찬가지로 [2]의 아이디어를 기반으로 한 참조 [91]에서는 원래 모델을 수정하여 모션 특징을 학습하는 새로운 이중 네트워크를 제안했습니다.

구조. 저자들은 주의 모듈을 추가하고 소셜 포어 맵[39]을 활용하여 모션 대표자를 학습했습니다.

[87]의 저자는 광학적 흐름을 사용하여 배경 비디오에서 물체를 전경으로 표시했습니다. 그런 다음 컨볼루션 신경망을 사용하여 식별된 물체의 특징을 광학 흐름의 히스토그램과 함께 추출했습니다. 모든 특징은 행렬 인수분해 방법을 사용하여 결합하고 서로 다른 하위 집합으로 클러스터링했습니다. 투표 시스템은 새로운 인스턴스가 정상인지 비정상인지 감지했습니다. [88]의 접근 방식은 비디오에서 튜브를 추출하여 비디오 세그먼트의 로컬 라이제이션에 중점을 두었습니다. 또한 이 방법은 점수 값을 기반으로 이상 징후를 정확하게 식별할 수 있음을 보여주었습니다.

[95]의 연구자들은 이상 징후 탐지를 위해 스파스 코딩을 사용했습니다. 스파스 코딩 [41]은 먼저 모든 정상 인스턴스로부터 사전 매핑을 학습했습니다. [95]의 저자들은 스택형 순환 신경망 자동 인코더에서 추출한 공간 및 시간 정보를 모두 사용하여 시간적 일관성 스파스 코딩을 제안했습니다. [97]의 연구에서는 이상 탐지를 위해 가우시안 혼합 모델을 사용했습니다. 가우시안 혼합 완전 컨볼루션 가변 자동 인코더라는 새로운 네트워크를 제안하여 모션 및 외관 특징을 모두 사용하여 비디오 패치를 모델링했습니다. 이 모델은 간단한 에너지 기반 방법을 통해 각 비디오 패치에 대한 확률 점수를 예측했습니다.

### 5.5. 이상 점수

예측 프레임/재구성된 프레임  $\hat{I}$ 와 기준 프레임  $I$  사이의 차이는 피크 신호 대 잡음비(PSNR)를 사용하여 계산됩니다:

$$PSNR(\hat{I}, I) = 10 \log_{10} \frac{[\max_i I]^2}{\sum_{i=1}^N (I - \hat{I})^2} \quad (5)$$

동영상의 모든 프레임에 대한 PSNR 점수는 [0, 1] 범위로 정규화됩니다. 그런 다음 다음 공식을 사용하여 이상 점수를 구합니다:

$$S(t) = \frac{PSNR_t - \min(PSNR)}{\max(PSNR) - \min(PSNR)} \quad (6)$$

이상 점수  $S(t)$ 는 프레임이 정상 이벤트인지 비정상 이벤트인지 판단하는 데 사용됩니다.

## 6. 연구 격차, 과제 및 향후 연구

### 6.1. 연구 격차

최근의 방법[43,126,127]은 구조적 특징과 광학적 흐름을 추출하여 비디오의 외형과 움직임 정보를 모두 활용하는 데 중점을 두고 있습니다. 광학적 흐름을 추출하려면 높은 계산 비용이 필요합니다. 따라서 일부 제안된 방법[116,128,129]은 광학적 흐름을 사용하지 않고 모션 정보를 캡처하려고 시도했습니다.

반면, 가장 최근에 제안된 방법들은 최신 모델을 적용하여 이상 징후 탐지 시스템의

정확도를 향상시키려는 시도를 하고 있습니다[130,131]. 이상 징후 탐지 시스템은 실시간 3448  
애플리케이션으로 실행되어야 하므로 모델의 복잡성을 고려해야 합니다.

## 6.2. 도전 과제

이 섹션에서는 비디오 감시를 위한 AbHAR의 일반적인 과제와 가능한 해결책 및 향후 연구에 대해 설명합니다. 전반적으로 연구자들은 지도 학습 방법에 의존해 접근해 왔습니다. 이를 위해서는 모델 정확도를 훈련하고 개선하기 위해 대량의 레이블이 지정된 데이터가 필요합니다. 하지만 데이터 수집은 지루하고 힘든 과정입니다. 또한 실제 상황의 모든 비정상적인 행동을 포괄할 수 있는 표준 데이터 세트도 존재하지 않습니다. 이러한 문제는 전이 학습 기법이나 클라우드 소싱을 활용하면 해결할 수 있습니다[132]. 또한, 다양한 접근 방식을 공정하게 비교할 수 있는 표준 지표가 필요합니다[133].

추론하는 동안 사람의 움직임과 외모의 미묘한 디테일로 인해 오경보가 발생하는 등 대부분의 접근 방식에는 몇 가지 한계가 있습니다. 어떤 경우에는 조기에

범죄 예방을 위해서도 탐지가 필요합니다. 이러한 문제는 감시 시스템이 사람의 행동을 학습하고 시뮬레이션할 수 있는 AbHAR[134]의 대표 특징을 선택함으로써 접근할 수 있습니다. 클래스 내 변형 및 클래스 간 유사성 문제는 모델이 동일한 활동 또는 서로 다른 클래스 간의 유사하지 않은 대표 패턴을 고려할 때 발생합니다. 이로 인해 정상 이벤트와 비정상 이벤트 사이의 경계가 모호해질 수 있으며, 모델은 분류를 위해 고유하고 차별적인 특징만 학습해야 합니다[135]. 또한, 한 그룹의 다중 피험자 상호 작용을 추적하는 것은 모델이 피험자로부터 시공간 정보를 캡처할 수 있어야 하는 까다로운 문제입니다[136].

시끄러운 환경은 많은 감시 시스템에서 주요 문제로 간주됩니다. 복잡하고 동적인 배경은 정확하게 모델링하기 어렵습니다. 또한 실시간 감시에서는 오클루전, 낮은 조도, 저화질 영상, 다양한 시점도 발생합니다. 멀티모달 데이터는 이러한 문제를 극복할 수 있습니다[137,138](예: 비디오 세그먼트에서 깊이 정보를 추출하기 위해 D-RGB 이미지 사용). 또한, 감시 시스템에는 실시간 감지가 필요하며, 이는 에너지 소비가 많고 많은 컴퓨팅 성능을 필요로 합니다. 따라서 샘플링 주파수와 반복적 분할 원리에 기반한 적응형 방법을 사용해야 할 필요성이 높아집니다[139].

### 6.3. 향후 연구

향후 연구는 비디오 세그먼트의 시공간적 관계를 탐색하기 위해 사전 학습된 모델과 같은 딥러닝 접근법에서 전이 학습 방법을 사용하는 데 초점을 맞출 것입니다[140]. AbHAR의 맥락에서 전이 학습은 대규모 데이터 세트에서 얻은 지식을 소규모 데이터 세트로 옮기는 데 사용될 수 있으며, 후자는 다른 분포와 문제를 가질 수 있습니다. 향후 연구에서는 딥러닝 기반 방법을 사용하여 AbHAR에서 전이 학습의 효과를 조사할 수 있습니다. 현재 연구자들은 활동 클래스를 분류하기 위해 키 프레임 정보[64,77]를 선택하여 비디오 모델을 해석하는 방법에 더 많은 주의를 기울이고 있습니다. 또한, 멀티모달 데이터 세트에서 풍부한 의미적 특징을 추출하는 것이 연구 트렌드입니다[75,141]. 이러한 정보를 통해 상호 작용하는 객체 간의 장기적인 관계를 설명할 수 있기 때문입니다[142]. 마지막으로, 인간 또는 인간-사물 간의 물리적 상호작용[143]은 인간의 행동을 이해하는 데 유용한 특징을 제공하고 모델 학습에 더 많은 해석적 특징을 제공할 수 있기 때문에 더 많은 조사가 필요합니다.

## 7. 결론

공공장소, 사무실, 가정 등 다양한 환경에서 범죄 활동을 감지하고 예방하는 데 있어 영상 감시 시스템의 중요성이 점점 더 커지고 있습니다. 영상 감시 시스템의 필수 구성 요소 중 하나는 비정상적인 인간 활동 인식(AbHAR)입니다. AbHAR은 모니터링 대상 환경의 안전과 보안에 잠재적인 위협이 될 수 있는 비정상적인 행동을 감지하고 이해하는 데 매우 중요합니다. 이 백서에서는 영상 감시에서 딥러닝 기반 방법에 대한 포괄적인 검토

를 제공합니다. 사람의 활동 인식은 비정상적인 행동을 감지하고 이해하는 데 중요한 역할을 합니다. 이 리뷰에서는 관련 연구와 배경 지식의 필수적인 측면을 다루었습니다. 또한 AbHAR에서 사용되는 다양한 벤치마크 데이터베이스와 각 데이터베이스에 대한 자세한 설명도 제시했습니다. 이러한 데이터베이스는 양과 질 모두에서 증가하고 있으며, 연구자들은 실시간 시나리오와 관련된 다양한 측면을 연구할 수 있습니다. 딥러닝 기반 방법에 대한 포괄적인 검토에서는 재구성 기반 접근법, 다중 클래스 분류, 미래 프레임 예측, 점수화 방법 등의 방법을 소개했습니다. 이러한 다양한 접근 방식은 연구자들이 더 나은 해결책을 찾도록 강요하는 AbHAR의 어려움을 보여줍니다. 또한 전이 학습 방법이나 멀티모달 데이터 세트에 대한 향후 연구는 이상 징후 탐지에서 더 나은 분석을 위한 초기 단계를 제공할 것입니다. 사실, 특징 엔지니어링 단계에서 대표 특징을 선택하는 것은 모든 보안 감시 시스템에서 가장 어려운 부분입니다. 동적 배경에서 강력한 피처를 추출하는 것은 필수적입니다.

인간의 행동을 추적하고 이해하기 위한 상대적이고 차별적인 패턴을 포착하는 데 사용됩니다.

**저자 기고:** 모든 저자가 연구 개념화 및 설계에 기여했습니다. 원고의 초안은 H.-T.D.와 V.-T.L.이 작성하고 V.T.H.가 감수했습니다. 모든 저자는 출판된 버전의 원고를 읽고 동의했습니다.

**자금 지원:** 이 연구는 베트남 호치민시 오픈 대학교의 지원을 받았습니다. 이 백서의 연구 중 일부는 중국과학원 자동화 연구소에서 수집한 CASIA 액션 데이터베이스를 사용했습니다.

**데이터 가용성 진술:** 본 연구 중에 생성 및/또는 분석된 데이터 세트는 합리적인 요청이 있을 경우 교신저자로부터 제공받을 수 있습니다.

**이해 상충:** 저자는 이해 상충이 없음을 선언합니다.

## 참조

1. Kumar, V. 이상 탐지: 설문 조사. *ACM Comput. Surv.* **2009**, *41*, 1-58. [CrossRef]
2. 술타니, W.; 첸, C.; 샤, M. 감시 비디오에서 실제 이상 감지. IEEE 컨퍼런스 논문집 컴퓨터 비전 및 패턴 인식, 미국 유타주 솔트레이크시티, 2018년 6월 18-23일; 6479-6488페이지.
3. Sabokrou, M.; Fayyaz, M.; Fathy, M.; Klette, R. Deep-cascade: 혼잡한 장면에서 빠른 이상 감지 및 로컬라이제이션을 위한 캐스케이드 3D 심층 신경망. *IEEE Trans. Image Process.* **2017**, *26*, 1992-2004. [CrossRef] [PubMed]
4. Medel, J.R.; Savakis, A. 예측 컨볼루션 장단기 메모리 네트워크를 사용한 비디오의 이상 감지. **2016**, arXiv:1612.00390.
5. 활동 인식을 위한 새로운 랜덤 포레스트 기반 클래스 점진적 학습 방법(Hu, C.; Chen, Y.; Hu, L.; Peng, X.). *패턴 인식* **2018**, *78*, 277-290. [CrossRef]
6. Xiao, Q.; Song, R. 계층적 동적 베이지안 네트워크에 기반한 동작 인식. *Multimed.* **2018**, *77*, 6955-6968. [CrossRef]
7. Sok, P.; Xiao, T.; Azeze, Y.; Jayaraman, A.; Albert, M.V. 숨겨진 마르코프 모델을 사용한 불완전한 척수 손상 피험자의 활동 인식. *IEEE Sens. J.* **2018**, *18*, 6369-6374. [CrossRef]
8. 아비딘, B.M.; 페르가니, L.; 페르가니, B.; 우살라, M. 일상 활동 인식 개선을 위한 시퀀스 특징 조합 및 수정된 가중 SVM의 공동 사용. *패턴 학문* **2018**, *21*, 119-138. [CrossRef]
9. Hu, X.; Hu, S.; Huang, Y.; Zhang, H.; Wu, H. 심층 증분 느린 특징 분석 네트워크를 사용한 비디오 이상 감지. *IET Comput. Vis.* **2016**, *10*, 258-267. [CrossRef]
10. Ijjina, E.P.; Chalavadi, K.M. 모션 시퀀스 정보와 딥러닝을 사용한 RGB-D 비디오에서의 인간 동작 인식. *패턴 인식* **2017**, *72*, 504-516. [CrossRef]
11. Tan, T.H.; Gochoo, M.; Huang, S.C.; Liu, Y.H.; Liu, S.H.; Huang, Y.F. 스마트 홈에서 다중 거주자 활동 인식 RGB 활동 이미지와 DCNN을 이용한 활동 인식. *IEEE Sens. J.* **2018**, *18*, 9718-9727. [CrossRef]
12. 영, T.; 하자리카, D.; 포리아, S.; 캄브리아, E. 딥러닝 기반 자연어 처리의 최근 동향. *IEEE Comput. Intell. Mag.* **2018**, *13*, 55-75. [CrossRef]
13. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. 심층 컨볼루션 신경망을 이용한 이미지넷 분류. 2012년 12월 3-6일, 미국 네바다주 레이크 타호에서 열린 신경 정보 처리 시스템의 발전, ; 1097-1105쪽.
14. Oyedotun, O.K.; Khashman, A. 비전 기반 정적 손 제스처 인식의 딥 러닝. *Neural Comput. Appl.* **2017**, *28*, 3941-3951. [CrossRef]
15. Weingaertner, T.; Hassfeld, S.; Dillmann, R. 인간 동작 분석: 검토. 1997년 6월 16일, 미국 워싱턴주 산후안, IEEE 비강체 및 관절 모션 워크숍 Proceedings; p. 0090.

16. Aggarwal, J.K.; Cai, Q. 인간 동작 분석: 검토. *Comput. Vis. Image Underst.* **1999**, *73*, 428-440. [[CrossRef](#)]
17. 발레라, M.; 벨라스틴, S.A. 지능형 분산 감시 시스템: 검토. *IEEE Proc. 비전 이미지 신호 프로세스*. **2005**, *152*, 192-204. [[CrossRef](#)]
18. Ko, T. 국토 안보 애플리케이션을 위한 영상 감시에서의 행동 분석에 관한 조사. 2008년 제37회 IEEE 응용 이미지 패턴 인식 워크숍, 미국 워싱턴 DC, 2008년 10월 15-17일; 1-8쪽.
19. Popoola, O.P.; Wang, K. 비디오 기반 비정상적인 인간 행동 인식-리뷰. *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.* **2012**, *42*, 865-878. [[CrossRef](#)]
20. Tsitsoulis, A.; 부르바키스, N. 비전 기반 인간 활동 인식에 대한 첫 번째 단계 비교 조사. *Int. J. Tools* **2013**, *24*, 782783783.
21. 공공장소를 위한 지능형 비디오 감시 시스템 - 설문조사. *J. Theor. Appl. Comput.* **2014**, *8*, 13-27.
22. Li, T.; Chang, H.; Wang, M.; Ni, B.; Hong, R.; Yan, S. 혼잡한 장면 분석: 설문 조사. *IEEE Trans. Circuits Syst. 비디오 테크놀.* **2014**, *25*, 367-386. [[CrossRef](#)]
23. 비디오 스트림에서 인간 활동 인식을 위한 도메인 및 컨텍스트 지식 사용에 관한 설문조사(Onofri, L.; 소다, P.; 페체니즈키, M.; Iannello, G.). *Expert Syst.* **2016**, *63*, 97-111. [[CrossRef](#)]

24. 헤라스, S.; 하란디, M.; 포리클리, F. 행동 인식에 대해 더 깊이 알아보기: 설문 조사. *Image Vis. Comput.* **2017**, *60*, 4-21. [CrossRef]
25. 지능형 비디오 감시 시스템을 위한 비정상 행동 인식: 검토. *Expert Syst.* **2018**, *91*, 480-491. [CrossRef]
26. 왕, P.; 리, W.; 오군보나, P.; 완, J.; 에스칼레라, S. 딥러닝을 통한 RGB-D 기반 인간 동작 인식: 설문 조사. *Comput. Vis. Image Underst.* **2018**, *171*, 118-139. [CrossRef]
27. 압달라, Z.S.; 가버, M.M.; 스리니바산, B.; 크리슈나스와미, S. 진화하는 데이터 스트림을 통한 활동 인식: 검토. *ACM Comput. Surv.* **2018**, *51*, 1-36. [CrossRef]
28. Kiran, B.R.; Thomas, D.M.; Parakkal, R. 동영상에서 비지도 및 반지도 이상 탐지를 위한 딥러닝 기반 방법 개요. *J. 이/미/정* **2018**, *4*, 36. [CrossRef]
29. Zhang, H.B.; Zhang, Y.X.; Zhong, B.; Lei, Q.; Yang, L.; Du, J.X.; Chen, D.S. 비전 기반 인간 행동에 대한 종합적인 조사 인식 방법. *Sensors* **2019**, *19*, 1005. [CrossRef]
30. Raval, R.M.; Prajapati, H.B.; Dabhi, V.K. 감시 비디오에서 인간 활동 인식에 대한 조사 및 분석. *Intell. Decis. Technol.* **2019**, *13*, 271-294. [CrossRef]
31. 디만, C.; 비슈와카르마, D.K. 비정상적인 인간 활동 인식을 위한 최첨단 기술 검토. *Eng. Appl. Artif. Intell.* **2019**, *77*, 21-45. [CrossRef]
32. 라마찬드라, B.; 존스, M.J.; 바사바이, R.R. 단일 장면 비디오 이상 탐지에 대한 설문 조사.
33. 당, L.M.; 민, K.; 왕, H.; 피란, M.J.; 이, C.H.; 문, H. 센서 기반 및 비전 기반 인간 활동 인식: 종합 설문 조사. *Pattern Recognit.* **2020**, *108*, 107561. [CrossRef]
34. Ahmed, M.; Mahmood, A.N.; Hu, J. 네트워크 이상 탐지 기법에 대한 조사. *J. Netw. Comput.* **2016**, *60*, 19-31. [CrossRef]
35. 네트워크 이상 징후 탐지 시스템에 대한 총체적인 검토: 포괄적인 조사. *J. Netw. Comput.* **2019**, *128*, 33-55. [CrossRef]
36. Fernandes, G.; Rodrigues, J.J.; Carvalho, L.F.; Al-Muhtadi, J.F.; Proença, M.L. 네트워크 이상에 대한 포괄적인 조사 탐지. *Telecommun. Syst.* **2019**, *70*, 447-489. [CrossRef]
37. CASIA. CASIA 액션 데이터베이스. 2007. 온라인에서 이용 가능: <http://www.cbsr.ia.ac.cn/english/Action/%20Databases/%20EN.asp> (2020년 11월 20일에 액세스).
38. Adam, A.; Rivlin, E.; Shimshoni, I.; Reinitz, D. 여러 고정 위치 모니터를 사용한 강력한 실시간 비정상 이벤트 탐지. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *30*, 555-560. [CrossRef] [PubMed]
39. Mehran, R.; Oyama, A.; Shah, M. 사회적 힘 모델을 이용한 비정상적인 군중 행동 감지. 2009 IEEE 컴퓨터 비전 및 패턴 인식 컨퍼런스, 미국 플로리다주 마이애미, 2009년 6월 20-25일; 935-942페이지.
40. Zaharescu, A.; Wildes, R. 시공간 지향 에너지, 하위 집합 포함 히스토그램 비교 및 이벤트 중심 처리를 사용한 비정상 행동 감지. 컴퓨터 비전에 관한 유럽 컨퍼런스, 헤라클리온 크레타, 그리스, 2010년 9월 5-11일; Springer: 베를린/하이델베르크, 독일, 2010년; 563-576페이지.
41. Lu, C.; Shi, J.; Jia, J. MATLAB에서 150 FPS의 비정상 이벤트 감지. 2013 IEEE 국제 컴퓨터 비전 컨퍼런스(), 호주 시드니, 2013년 12월 1-8일; 2720-2727쪽.
42. 통계 시각 컴퓨팅 랩. UCSD 이상 데이터 세트. 2014. 온라인에서 사용 가능: <http://www.svcl.ucsd.edu/projects/anomaly/> (2020년 11월 20일에 액세스).
43. 이상 탐지를 위한 미래 프레임 예측-새로운 기준선. 2018 Proceedings of the 2018 IEEE 컴퓨터 비전 및 패턴 인식 컨퍼런스, 솔트레이크 시티, 미국 유타주, 2018년 6월 18-23일; 6536-6545쪽.
44. 라마찬드라, B.; 존스, M. 거리 풍경: 비디오 이상 감지를 위한 새로운 데이터 세트 및 평가 프로토콜. 2020년 3월 1~5일, 미국 콜로라도주 스노우매스, 2020 IEEE 동계 컴퓨터 비전 응용 컨퍼런스, ; 2569-2578쪽.
45. Babae, M.; Dinh, D.T.; Rigoll, G. 비디오 시퀀스 배경 제거를 위한 심층 컨볼루션 신경망. *Pattern Recognit.* **2018**, *76*, 635-649. [CrossRef]
46. Subudhi, B.N.; Ghosh, S.; Shiu, S.C.; Ghosh, A. 로컬 변화 감지를 위한 통계적 특징 백 기반 배경 차감. *Inf.* **2016**, *366*, 31-47. [CrossRef]
47. Shen, Y.; Hu, W.; Yang, M.; Liu, J.; Wei, B.; Lucey, S.; Chou, C.T. 임베디드 카메라 네트워크를 위한 실시간 및 강력한 압축 배경



- 감산. *IEEE Trans. Mob. Comput.* **2015**, *15*, 406-418. [[CrossRef](#)]
48. Jiang, S.; Lu, X. WeSamBE: 배경 감산을 위한 가중치 샘플 기반 방법. *IEEE Trans. Circuits Syst. 비디오 테크놀.* **2017**, *28*, 2105-2115. [[CrossRef](#)]
49. Sakkos, D.; Liu, H.; Han, J.; Shao, L. 3D 컨볼루션 신경망을 사용한 엔드투엔드 비디오 배경 차감. *Multimed.* **2018**, *77*, 23023-23041. [[CrossRef](#)]
50. Minematsu, T.; Shimada, A.; Uchiyama, H.; Taniguchi, R.I. 심층 신경망 기반 배경 빼기 분석. *J. 이/디/징* **2018**, *4*, 78. [[CrossRef](#)]
51. Ullah, A.; Muhammad, K.; Del Ser, J.; Baik, S.W.; de Albuquerque, V.H.C. 시간적 광 흐름을 이용한 활동 인식 컨볼루션 특징 및 다층 LSTM. *IEEE Trans. Ind. Electron.* **2018**, *66*, 9692-9702. [[CrossRef](#)]
52. Ladjailia, A.; Bouchrika, I.; Merouani, H.F.; Harrati, N.; Mahfouf, Z. 광학적 흐름을 통한 인간 활동 인식: 활동을 기본 동작으로 분해. *Neural Comput.* **2020**, *32*, 16387-16400. [[CrossRef](#)]
53. Mliki, H.; Bouhlel, F.; Hammami, M. UAV로 캡처한 비디오 시퀀스에서 인간 활동 인식. *패턴 인식* **2020**, *100*, 107140. [[CrossRef](#)]

54. 상, R.; 딜론, J.K.; 쿠시와하, A.K.S.; 스리바스타바, R. 활동 인식을 위한 3D 심층 컨볼루션 네트워크의 깊이 기반 확대된 시간적 차원. *Multimed.* **2019**, *78*, 30599-30614. [[CrossRef](#)]
55. 불확실성 속에서 살아가는 일상의 복잡한 활동을 인식하기 위한 복잡한 시간적 규칙을 마이닝합니다. *패턴 인식*. **2016**, *60*, 1015-1028. [[CrossRef](#)]
56. Sun, B.; Li, Y.; Guosheng, C.; Zhang, J.; Chang, B.; Min, C. 적외선 비디오에서 마르코프 랜덤 필드 기반 평가 지표를 사용한 이동 대상 세분화. *Opt.* **2018**, *57*, 013106. [[CrossRef](#)]
57. 라이, Y.; 한, Y.; 왕, Y. 프로토타입 유도 차별적 잠복 임베딩을 이용한 이상 탐지. In Proceedings of the 2021 IEEE 국제 데이터 마이닝 컨퍼런스(ICDM), 뉴질랜드 오클랜드, 2021년 12월 7-10일; 300-309페이지.
58. Doshi, K.; Yilmaz, Y. 감시 비디오에서 애니샷 순차적 이상 감지. 2020 IEEE/CVF 컴퓨터 비전 및 패턴 인식 워크숍 컨퍼런스, 미국 워싱턴주 시애틀, 2020년 6월 14-19일; 934-935쪽.
59. Cai, R.; Zhang, H.; Liu, W.; Gao, S.; Hao, Z. 비디오 이상 감지를 위한 외관-동작 메모리 일관성 네트워크. In 인공 지능에 관한 AAAI 컨퍼런스, 온라인, 2021년 2월 2-9일; 35권, 938-946쪽.
60. 쿠마르, K.S.; 바바니, R. HOG, GiST 및 컬러 기능을 사용한 자기 중심 비디오에서의 인간 활동 인식. *Multimed. 도구 Appl.* **2020**, *79*, 3543-3559. [[CrossRef](#)]
61. 왕, Z.; 우, D.; 첸, J.; 고님, A.; 호세인, M.A. EEMD- 기반 특징 및 게임 이론 기반 특징 선택을 통한 3축 가속도계 기반 인간 활동 인식. *IEEE Sens. J.* **2016**, *16*, 3198-3207. [[CrossRef](#)]
62. Roy, P.K.; Om, H. 감시 비디오에서 HOG 특징과 SVM 분류기를 사용한 사람의 의심스럽고 폭력적인 활동 감지. *이미지 처리에 서 소프트웨어 컴퓨팅 및 머신 러닝의 발전*, Springer: Berlin/Heidelberg, Germany, 2018; pp. 277-294.
63. Patel, C.I.; Garg, S.; Zaveri, T.; Banerjee, A.; Patel, R. 비제약적 비디오 시퀀스에 대한 특징 융합을 사용한 인간 동작 인식. *Comput. Electr.* **2018**, *70*, 284-301. [[CrossRef](#)]
64. Jagadeesh, B.; Patil, C.M. SVM을 사용한 비디오 기반 인간 활동 감지, 인식 및 행동 분류. *Trans. Mach. Learn. Artif. Intell.* **2018**, *6*, 22-22.
65. Kale, G.V. 실시간 및 오프라인 데이터 세트에서 인간 활동 인식. *Int. J. Intell. Syst.* **2019**, *7*, 60-65. [[CrossRef](#)]
66. 벅스, A.; 안젤로프, P.; 하비브, Z. 비전 기반 인간 활동 인식: 검토. *컴퓨터 지능의 발전 시스템*, 스프링거: 베를린/하이델베르크, 독일, 2017; 341-371쪽.
67. Ahad, M.A.R.; Tan, J.K.; 김, H.; 이시카와, S. SURF 기반 템플릿을 통한 활동 표현. *Comput. 방법 생체 공학. Biomed. Eng. Imaging Vis.* **2018**, *6*, 573-583. [[CrossRef](#)]
68. 파트로나, F.; 차치토피스, A.; 자르팔라스, D.; 다라스, P. 모션 분석: 동작 감지, 인식 및 모션 기반 평가 캡처 데이터. *패턴 인식*. **2018**, *76*, 612-622. [[CrossRef](#)]
69. Tang, J.; Cheng, H.; Zhao, Y.; Guo, H. 연속 손 궤적 제스처 인식을 위한 구조화된 동적 시간 워핑. *패턴 인식*. **2018**, *80*, 21-31. [[CrossRef](#)]
70. 김, K.; 잘랄, A.; 마흐무드, M. 깊이 실루엣을 이용한 비전 기반 인간 활동 인식 시스템: 거주자 모니터링을 위한 스마트 홈 시스템. *J. Electr. Eng. Technol.* **2019**, *14*, 2567-2573. [[CrossRef](#)]
71. 바우만, F.; 엘러스, A.; 로젠한, B.; 리아오, J. 새로운 시공간 볼륨 바이너리 패턴을 사용하여 인간의 행동을 인식합니다. *뉴로컴퓨팅* **2016**, *173*, 54-63. [[CrossRef](#)]
72. 행동 분류를 위한 국소 다항식 시공간 기술자, 피카드, D.; 고셀린, P.H. 행동 분류를 위한 국소 다항식 시공간 기술자. *Mach. Vis. Appl.* **2016**, *27*, 351-361. [[CrossRef](#)]
73. 행동 인식을 위한 스파스 코딩 기반 시공간 비디오 표현. *Multimed. Tools Appl.* **2017**, *76*, 12645-12658. [[CrossRef](#)]
74. Seyfiog̃ lu, M.S.; Özbayog̃ lu, A.M.; Gürbüz, S.Z. 유사한 지원 및 비지원 인간 활동의 레이더 기반 분류를 위한 딥 컨볼루션 자동 인코더. *IEEE Trans. Aerosp. Electron. Syst.* **2018**, *54*, 1709-1723. [[CrossRef](#)]

75. Shahroudy, A.; Ng, T.T.; Yang, Q.; Wang, G. 심층 비디오에서 동작 인식을 위한 멀티모달 멀티파트 학습. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *38*, 2123-2129. [[CrossRef](#)]
76. Li, M.; Zhou, Z.; Liu, X. 바운딩 박스 제약 조건과 LSTM을 사용한 다인칭 포즈 추정. *IEEE Trans. Multimed.* **2019**, *21*, 2653-2663. [[CrossRef](#)]
77. 복잡한 사람 포즈 인식을 위한 신체 부위 레이블이 있는 사람 깊이 이미지 생성. *Pattern Recognit.* **2017**, *71*, 402-413. [[CrossRef](#)]
78. Xu, D.; Yan, Y.; Ricci, E.; Sebe, N. 외모 및 모션의 심층 표현을 학습하여 동영상에서 비정상적인 이벤트 감지. *Comput. Vis. Image Underst.* **2017**, *156*, 117-127. [[CrossRef](#)]
79. 장, Y.; 루, H.; 장, L.; 루안, X. 이상 징후 탐지를 위한 동작 및 외관 단서의 결합. *패턴 인식*. **2016**, *51*, 443-452. [[CrossRef](#)]
80. Ji, X.; Cheng, J.; Feng, W.; Tao, D. 깊이 시퀀스를 사용한 인간 동작 인식을 위한 스켈레톤 임베디드 모션 바디 파티션. *신호 프로세스*. **2018**, *143*, 56-68. [[CrossRef](#)]
81. 하산, M.; 최, J.; 노이만, J.; 로이-쇼두리, A.K.; 데이비스, L.S. 비디오 시퀀스에서 시간적 규칙성 학습. 2016 IEEE 컴퓨터 비전 및 패턴 인식 컨퍼런스, 미국 네바다주 라스베이거스, 2016년 6월 27~30일; pp. 733-742.
82. Sabokrou, M.; Fathy, M.; Hoseini, M. 자동 인코더의 희소성 및 재구성 오류에 기반한 비디오 이상 감지 및 로컬라이제이션. *Electron.* **2016**, *52*, 1122-1124. [[CrossRef](#)]

83. 나라심한, M.G.; 카마스, S. 스파스 노이즈 제거 자동 인코더를 사용한 동적 비디오 이상 감지 및 로컬라이제이션. *Multimed.* **2018**, *77*, 13173-13195. [CrossRef]
84. Zhao, Y.; Deng, B.; Shen, C.; Liu, Y.; Lu, H.; Hua, X.S. 비디오 이상 감지를 위한 시공간적 자동 인코더. 2017년 10월 23-27일, 미국 캘리포니아주 마운틴뷰, 제25회 ACM 멀티미디어 국제 컨퍼런스, 1933-1941페이지.
85. Sabokrou, M.; Fayyaz, M.; Fathy, M.; Moayed, Z.; Klette, R. 심층 이상: 혼잡한 장면에서 빠른 이상 징후 탐지를 위한 완전 컨볼루션 신경망. *Comput. Vis. Image Underst.* **2018**, *172*, 88-97. [CrossRef]
86. 리베이로, M.; 라자레티, A.E.; 로페스, H.S. 비디오의 이상 감지를 위한 딥 컨볼루션 자동 인코더에 대한 연구. *Pattern Recognit.* **2018**, *105*, 13-22. [CrossRef]
87. 사잘리안, B.; 마르비, H.; 아흐마디파드, A. 비디오의 이상 감지 및 로컬라이제이션을 위한 딥 및 스파스 기능. 2019 제4회 패턴 인식 및 이미지 분석 국제 컨퍼런스(IPRIA), 이란 테헤란, 3월 6-7일, 2019; 173-178페이지.
88. Landi, F.; Snoek, C.G.; Cucchiara, R. 비디오 감시의 이상 위치.
89. Zhou, J.T.; Du, J.; Zhu, H.; Peng, X.; Liu, Y.; Goh, R.S.M. Anomalynet: 비디오 감시를 위한 이상 감지 네트워크. *IEEE Trans. Inf. Forensics Secur.* **2019**, *14*, 2537-2550. [CrossRef]
90. 향상된 비디오 이상 감지를 위한 모션 인식 기능. *arXiv* **2019**, arXiv:1907.10211.
91. Lin, S.; Yang, H.; Tang, X.; Shi, T.; Chen, L. Social MIL: 군중 이상 탐지를 위한 상호 작용 인식. 2019 제16회 첨단 비디오 및 신호 기반 감시(AVSS) 국제 컨퍼런스(대만 타이베이, 18-21일) Proceedings of the 2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS) 2019년 9월; 1-8페이지.
92. 공, D.; 리우, L.; 레, V.; 사하, B.; 만수르, M.R.; 벤카테쉬, S.; 반 덴 헨겔, A. 비정상 감지를 위한 정상성 기억: 비지도 비정상 감지를 위한 메모리 증강 심층 자동 인코더. 2019 IEEE/CVF 국제 컴퓨터 비전 학술대회(), 서울, 대한민국, 2019년 10월 27일 ~11월 2일, 1705~1714페이지.
93. dos Santos, F.P.; Ribeiro, L.S.; Ponti, M.A. 다른 비디오 이상 탐지 도메인에서 전송된 기능 임베딩의 일반화. *J. Vis. Commun. Image Represent.* **2019**, *60*, 407-416. [CrossRef]
94. Ionescu, R.T.; Khan, F.S.; Georgescu, M.I.; Shao, L. 비디오에서 비정상 이벤트 감지를 위한 객체 중심 자동 인코더 및 더미 이상 징후. 2019 IEEE/CVF 컴퓨터 비전 및 패턴 인식 컨퍼런스(CVPR), , 미국 캘리포니아주 롱비치, 2019년 6월 15~20일, 7842-7851페이지.
95. 루오, W.; 리우, W.; 리안, D.; 탕, J.; 두안, L.; 펑, X.; 가오, S. 스파스 코딩으로 딥 뉴럴 네트워크에서 영감을 얻은 비디오 이상 감지. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *43*, 1070-1084. [CrossRef] [PubMed]
96. 적응형 프레임 내 분류 네트워크에 기반한 비디오 이상 탐지 및 로컬라이제이션. *IEEE Trans. Multimed.* **2019**, *22*, 394-406. [CrossRef]
97. Fan, Y.; Wen, G.; Li, D.; Qiu, S.; Levine, MD; Xiao, F. 가우시안 혼합을 통한 비디오 이상 감지 및 로컬라이제이션 완전 컨볼루션 변형 자동 인코더를 통한 비디오 이상 감지. *Comput. Vis. Image Underst.* **2020**, *195*, 102920. [CrossRef]
98. 위키피디아. 자동 인코더-위키 백과, 무료 백과 사전. 2023. 온라인에서 이용 가능: <http://en.wikipedia.org/w/index.php?title=Autoencoder&oldid=1141727025> (2023년 4월 4일에 액세스).
99. Hochreiter, S.; 슈미트 후버, J. 장단기 기억. *Neural Comput.* **1997**, *9*, 1735-1780. [CrossRef]
100. Aich, A.; Peng, K.C.; Roy-Chowdhury, A.K. 타겟 도메인 적응 없이 크로스 도메인 비디오 이상 감지. 2023년 1월 2-7일, 미국 하와이주 와이코올로아, 2023년 IEEE/CVF 컴퓨터 비전 응용 동계 컨퍼런스 논문집(Proceedings of the 2023 IEEE/CVF Winter Conference on Applications of Computer Vision); pp. 2579-2591.
101. 양, Z.; 우, P.; 리우, J.; 리우, X. 이상 탐지를 위한 적응형 클러스터러를 사용한 동적 로컬 집계 네트워크. 컴퓨터 비전-ECCV 2022: 제17회 유럽 컨퍼런스, 텔아비브, 이스라엘, 2022년 10월 23-27일; Proceedings, Part IV; Springer: Berlin/Heidelberg, Germany, 2022; 404-421페이지.
102. 비디오 이상 탐지를 위한 외관-동작 정상성 학습. In 2022 IEEE 국제 멀티미디어 및 엑스포 컨퍼런스(ICME), 대만 타이베이, 2022년

7월 18-22일; 1-6쪽.

103. 장, X.; 팡, J.; 양, B.; 첸, S.; 리, B. 혼잡한 장면에서 이상 감지를 위한 하이브리드 주의 및 모션 제약. *IEEE Trans. Circuits Syst. Video Technol.* **2023**, *33*, 2259-2274. [[CrossRef](#)]
104. 왕, T.; 쉬, X.; 셴, F.; 양, Y. 시각적 이상 탐지를 위한 인지 기억 증강 네트워크. *IEEE/CAA J. Autom. Sin.* **2021**, *8*, 1296-1307. [[CrossRef](#)]
105. Bahrami, M.; Pourahmadi, M.; Vafaei, A.; Shayesteh, M. R. 녹화된 비디오 스트림에서 단일 프레임과 다중 프레임 이상 탐지 및 로컬라이제이션 간의 비교 연구. *J. Vis. Commun. Image Represent.* **2021**, *79*, 103232. [[CrossRef](#)]
106. 메모리 증강 흐름 재구성 및 흐름 안내 프레임 예측을 통한 하이브리드 비디오 이상 감지 프레임워크. 2021 IEEE/CVF 국제 컴퓨터 비전 컨퍼런스(Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision), 몬트리올, QC, 캐나다, 2021년 10월 10-17일; 13588-13597페이지.
107. Tang, W.; Feng, Y.; Li, J. 비디오 이상 감지를 위한 메모리 모듈이 있는 자동 인코더. 2021년 5월 28~30일, 중국 난창에서 열린 2021년 제36회 중국 자동화 협회(YAC) 청년 학술 연례회의 논문집; pp. 473-478.
108. 동영상 이상 징후 탐지를 위한 설명 가능하고 효율적인 딥 러닝 프레임워크. *Clust. Comput.* **2022**, *25*, 2715-2737. [[CrossRef](#)]
109. A3N: 비디오 시퀀스에서 이상 징후를 감지하기 위한 주의 기반 적대적 자동 인코더 네트워크. *J. Vis. Commun. Image Represent.* **2022**, *87*, 103598. [[CrossRef](#)]

110. Borja-Borja, L.F.; Azorín-López, J.; Saval-Calvo, M.; Fuster-Guilló, A.; Sebban, M. 로컬 동작 및 컨텍스트를 사용하여 그룹 활동의 자동 인식을 위한 아키텍처. *IEEE Access* **2022**, *10*, 79874-79889. [[CrossRef](#)]
111. Huang, X.; Hu, Y.; Luo, X.; Han, J.; Zhang, B.; Cao, X. Boosting Variational Inference with Margin Learning for Few-Shot 장면 적응형 이상 탐지. *IEEE Trans. Circuits Syst. Video Technol.* **2022**, 조기 액세스. [[CrossRef](#)]
112. Li, D.; Nie, X.; Li, X.; Zhang, Y.; Yin, Y. 생성적 적대적 네트워크를 통한 컨텍스트 관련 비디오 이상 탐지. *Pattern Recognit.* **2022**, *156*, 183-189. [[CrossRef](#)]
113. Yu, G.; Wang, S.; Cai, Z.; Liu, X.; Xu, C.; Wu, C. 정상성 우위와 자기 속도 개선을 통한 라벨링되지 않은 비디오에서 심층적 인 이상 징후 발견. 컴퓨터 비전 및 패턴 인식에 관한 2022 IEEE/CVF 컨퍼런스 논문집, 뉴올리언스, LA, 미국, 2022년 6월 18-24일; 13987-13998페이지.
114. Zhang, Z.; Zhong, S.h.; Fares, A.; Liu, Y. 전경과 배경을 분리하여 이상 감지: 비디오 비정상 이벤트 감지를 위한 상호 생성 적 대적 네트워크. *Comput. Vis. Image Underst.* **2022**, *219*, 103416. [[CrossRef](#)]
115. 왕, Z.; 첸, Y. 듀얼 스트림 메모리 네트워크를 이용한 이상 탐지. *J. Vis. Commun. Image Represent.* **2023**, *90*, 103739. [[CrossRef](#)]
116. 비디오 이상 감지를 위한 주의 기반 잔여 자동 인코더. *Appl. Intell.* **2023**, *53*, 3240-3254. [[CrossRef](#)]
117. Deng, H.; Zhang, Z.; Zou, S.; Li, X. 비지도 비디오 이상 감지를 위한 양방향 프레임 보간. 2023년 1월 2-7일, 미국 하와이주 와이콜로아, 2023년 IEEE/CVF 컴퓨터 비전 응용 프로그램 동계 컨퍼런스 논문집; pp. 2634-2643.
118. Pan, S.J.; Tsang, I.W.; Kwok, J.T.; Yang, Q. 전송 구성 요소 분석을 통한 도메인 적응. *IEEE Trans. Neural Netw.* **2010**, *22*, 199-210. [[CrossRef](#)]
119. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. Ssd: 싱글 샷 멀티박스 검출기. 2016년 10월 11-14일, 네덜란드 암스테르담, 유럽 컴퓨터 비전 학회 논문집, Springer: Cham, Switzerland, 2016; 21-37쪽.
120. 맥킨, J. 다변량 관측의 분류 및 분석을 위한 몇 가지 방법. 제 5회 버클리 수학 통계 및 확률 심포지엄, 미국 캘리포니아주 오클랜드, 1965년 6월 21일~7월 18일, 제 1권, 281~297페이지.
121. 굿펠로우, I.; 푸제-아바디, J.; 미르자, M.; 쉬, B.; 워데-팔리, D.; 오자이르, S.; 쿠르빌, A.; 벤지오, Y. 생성적 적대 신경망. 2014년 12월 8-13일, 미국 매사추세츠주 케임브리지, 신경 정보 처리 시스템 학회 논문집(Proceedings of the Advances in Neural Information Processing Systems, Cambridge, MA); pp. 2672-2680.
122. 론네버거, O.; 피셔, P.; 브록스, T. U-net: 생의학 이미지 분할을 위한 컨볼루션 네트워크. 의료 영상 컴퓨팅 및 컴퓨터 보조 종재에 관한 국제 컨퍼런스, 독일 뮌헨, 10월 5-9일, 2015; Springer: Cham, Switzerland, 2015; 234-241쪽.
123. 위스콧, L.; 세노스키, T.J. 느린 특징 분석: 불변성에 대한 비지도 학습. *Neural Comput.* **2002**, *14*, 715-770. [[CrossRef](#)]
124. Tran, D.; Bourdev, L.; Fergus, R.; Torresani, L.; Paluri, M. 3D 컨볼루션 네트워크로 시공간적 특징 학습. In IEEE 국제 컴퓨터 비전 컨퍼런스 논문집, 칠레 산티아고, 2015년 12월 7-13일; 4489-4497쪽.
125. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. 주의만 기울이면 충분합니다. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 5998-6008.
126. 동영상에서 이상 징후를 탐지하기 위한 고속 스파스 코딩 네트워크. *패턴 인식* **2020**, *107*, 107515. [[CrossRef](#)]
127. Li, N.; Chang, F.; Liu, C. 혼합한 장면에서 비디오 이상 감지를 위한 공간-시간 캐스케이드 자동 인코더. *IEEE Trans. Multimed.* **2020**, *23*, 203-215. [[CrossRef](#)]
128. Chang, Y.; Tu, Z.; Xie, W.; Yuan, J. 비디오 이상 감지를 위한 클러스터링 기반 딥 오토 인코더. 컴퓨터 비전-ECCV 2020: 제16회 유럽 컨퍼런스, 글래스고, 영국, 2020년 8월 23~28일; Proceedings, Part XV 16; Springer: 베를린/하이델베르크, 독일, 2020년; 329~345페이지.
129. Chang, Y.; Tu, Z.; Xie, W.; Luo, B.; Zhang, S.; Sui, H.; Yuan, J. 시공간 해리를 통한 비디오 이상 탐지. *Pattern Recognit.* **2022**,

- 122, 108213. [[CrossRef](#)]
130. 황, X.; 자오, C.; 가오, C.; 첸, L.; 우, Z. 비감독 비디오 이상 탐지를 위한 합성 의사 이상: A 간단하면서도 효율적인 마스크드 자동 인코더 기반 프레임워크. *arXiv* **2023**, arXiv:2303.05112.
  131. Sun, X.; Chen, J.; Shen, X.; Li, H. 비디오 이상 탐지를 위한 시공간적 표현을 갖춘 트랜스포머. 구조적, 구문적, 통계적 패턴 인식 학회 논문집: Joint IAPR 국제 워크숍, S+ SSPR 2022, 몬트리올, QC, 캐나다, 2022년 8월 26-27일; Springer: 베를린/하이델베르크, 독일, 2023; 213-222페이지.
  132. Guo, L.; Wang, L.; Liu, J.; Zhou, W.; Lu, B. HuAc: 클라우드 소싱된 와이파이 신호와 골격 데이터를 이용한 인간 활동 인식. *Wirel. Commun. Mob. Comput.* **2018**, *2018*, 6163475. [[CrossRef](#)]
  133. 카바 하일브론, F.; 에스코르시아, V.; 가넴, B.; 카를로스 니에블스, J. 액티비넷: 인간 활동 이해를 위한 대규모 비디오 벤치마크. 2015 IEEE 컴퓨터 비전 및 패턴 인식 컨퍼런스, 미국 매사추세츠주 보스턴, , 2015년 6월 7-12일; 961-970쪽.
  134. Pei, Z.; Qi, X.; Zhang, Y.; Ma, M.; Yang, Y.H. 사회적 친밀도를 이용한 혼합한 장면에서의 인간 궤적 예측 장단기 메모리. *패턴 인식* **2019**, *93*, 273-282. [[CrossRef](#)]
  135. Akila, K.; Chitrakala, S. 클래스 내 가변성 및 클래스 간 유사성을 처리하기 위한 고도로 정제된 인간 행동 인식 모델. *Multimed.* **2019**, *78*, 20877-20894.

136. Wateosot, C.; Suvonvorn, N. 저수준 특징에 기반한 상호 작용력을 통한 그룹 활동 인식. *IEEJ Trans. Electr. Electron.* **2019**, *14*, 1061-1073. [[CrossRef](#)]
137. Jalal, A.; 김, Y.H.; 김, Y.J.; 카말, S.; 김, D. 시공간적 다중 융합 기능을 사용한 심도 비디오에서 강력한 인간 활동 인식. *Pattern Recognit.* **2017**, *61*, 295-308. [[CrossRef](#)]
138. Carreira, J.; Noland, E.; Hillier, C.; Zisserman, A. kinetics-700 인간 행동 데이터 세트에 대한 짧은 메모. *arXiv* **2019**, arXiv:1907.06987.
139. Xu, C.; Chai, D.; He, J.; Zhang, X.; Duan, S. InnoHAR: 복잡한 인간 활동 인식을 위한 심층 신경망. *IEEE Access* **2019**, *7*, 9893-9902. [[CrossRef](#)]
140. 살라켄, S.M.; 코스라비, A.; 응우옌, T.; 나하반디, S. 딥러닝으로 회귀 문제를 해결하기 위한 전이 학습을 씨앗으로 삼았습니다. *Expert Syst.* **2019**, *115*, 565-577. [[CrossRef](#)]
141. Ordóñez, F.J.; Roggen, D. 멀티모달 웨어러블 활동 인식을 위한 심층 컨볼루션 및 lstm 반복 신경망. *센서* **2016**, *16*, 115. [[CrossRef](#)] [[PubMed](#)]
142. Nguyen, T.N.; Thai, C.H.; Luu, A.T.; Nguyen-Xuan, H.; Lee, J. 기능 등급 탄소 나노튜브 강화 복합 쉘의 NURBS 기반 포스트버클링 분석. *Comput. Methods Appl.* **2019**, *347*, 983-1003. [[CrossRef](#)]
143. Goyal, R.; Kahou, S.E.; Michalski, V.; Materzynska, J.; Westphal, S.; Kim, H.; Haenel, V.; Fruend, I.; Yianilos, P.; Mueller-Freitag, M.; 외. 시각 상식 학습 및 평가를 위한 "Something Something" 비디오 데이터베이스. 2017 IEEE 국제 컴퓨터 비전 컨퍼런스 (ICCV), 이탈리아 베니스, 2017년 10월 22~29일; 1권, 5페이지.

**면책 조항/발행인 참고 사항:** 모든 출판물에 포함된 진술, 의견 및 데이터는 전적으로 개별 저자 및 기고자의 것이며, MDPI 및/또는 편집자의 것이 아닙니다. MDPI 및/또는 편집자는 콘텐츠에 언급된 아이디어, 방법, 지침 또는 제품으로 인해 발생한 인명 또는 재산상의 피해에 대해 책임을 지지 않습니다.