

약한 감독 하에 보이지 않는 비정상적인 활동에 적응하는 법 배우기

박재유*, 김준하*, 한보형

대한민국 서울대학교, ECE & ASRI
{bellos1203, junha.kim, bhhan}@snu.ac.kr

요약. 우리는 비디오에서 약하게 감독되는 이상 징후 탐지를 위한 메타 학습 프레임워크를 제시하며, 이 프레임워크는 이진 레이블의 비디오 수준 주석만 사용할 수 있을 때 탐지기가 보이지 않는 유형의 이상 활동에 효과적으로 적응하는 방법을 학습합니다. 저희의 연구는 기존 방법들이 눈에 보이지 않는 다양한 이상 징후에 대한 일반화가 잘 이루어지지 않는다는 사실에 착안했습니다. 메타 학습 체계를 갖춘 이상 탐지기는 모델을 더 나은 최적화를 위한 초기화 지점으로 유도함으로써 이러한 한계를 완화할 수 있다고 주장합니다. 우리는 두 가지 까다로운 데이터 세트인 UCF-Crime과 ShanghaiTech에서 프레임워크의 성능을 평가합니다. 실험 결과, 저희 알고리즘은 감시가 취약한 환경에서 보이지 않는 비정상적인 이벤트를 찾아내는 능력을 향상시켰음을 입증했습니다. 기술적인 기여 외에도, 우리는 UCF-Crime 데이터 세트에서 누락된 레이블에 주석을 달아 작업을 효과적으로 평가할 수 있도록 했습니다.

키워드: 이상 징후 탐지; 메타 학습; 약하게 감독된 학습; 학습

1 소개

인간은 이상 이벤트 감지의 정의가 명확하지 않음에도 불구하고 사전 지식을 자연스럽게 일반화하여 비디오에서 비정상적인 이벤트를 쉽게 식별합니다. 반대로 컴퓨터 비전 알고리즘은 비정상 이벤트 탐지를 위한 모델을 얻기 위해 수많은 주석이 달린 훈련 예제를 기반으로 한 광범위한 학습 프로세스에 의존합니다. 비디오의 이상 이벤트 탐지를 위해 다양한 접근 방식이 제안되고 있습니다. 생성 모델에

기반한 방법[1, 2]은 정상 패턴을 재구성할 수 있다고 주장하며, [3, 4]는 이진 분류기에 기반한 판별 기법을 제안합니다. 동영상에 대한 이상 징후 탐지의 상당한 발전에도 불구하고 [1-5], 두 가지 범주의 기존 방법 모두 중요한 단점을 가지고 있습니다. 최근 연구[3]에 따르면 생성적 접근 방식은 보이지 않는 정상 패턴을 비정상적으로 예측하기 쉽기 때문에 장면 변화가 심한 비디오의 인식 문제에 적합하지 않다고 합니다. 또한 생성된 비디오는 훈련 예제를 구성하는 데 사용된 카메라와 동일한 시점을 갖는 등 다양성이 제한되는 경우가 많습니다. 반면, 범죄 분류기는 보이지 않는 정상 또는 비정상 유형에 대해 강력하지 않을 수 있습니다.

* 이 저자들은 동등하게 기여했습니다.

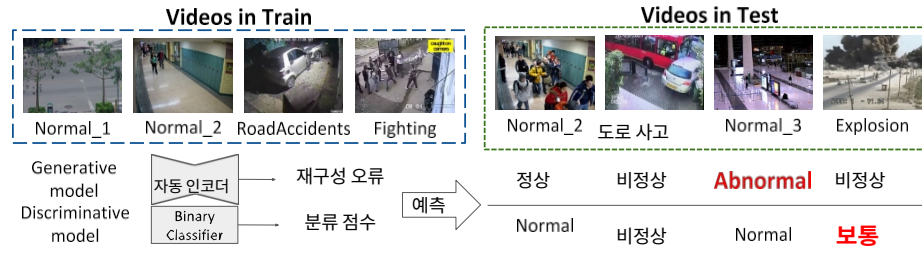


그림 1: 기존 비디오 비정상 이벤트 탐지 접근 방식의 한계. 생성 모델(예: 자동 인코더)은 학습 데이터에서 정상 패턴을 학습하려고 시도하며, 학습 중에 보이는 정상 비디오는 성공적으로 재구성하지만(정상 2) 새로운 시점에서 캡처한 비디오는 재구성하지 못합니다(정상 3). 반면, 판별 접근법(예: 이진 분류기)은 주어진 데이터에서 비정상 패턴을 학습하여 각 프레임을 비정상과 정상이라는 두 가지 클래스로 분류하는 데 중점을 둡니다. 따라서 판별 분류기는 학습 중에 보이는 비정상 이벤트(도로 사고)는 감지하지만 테스트 시에는 보이지 않는 유형의 비정상 이벤트(폭발)는 인식하지 못합니다. 빨간색으로 굵은 글자는 개별 접근 방식에 대한 잘못된 예측에 해당합니다.

활동을 감지합니다. 특히 사전 정의된 유형의 이상 이벤트만 탐지할 수 있고 학습 데이터에 과적합하는 경향이 있습니다. 그림 1은 위에서 언급한 기존 방법의 한계를 보여줍니다.

모든 종류의 정상 및 비정상 활동 패턴이 포함된 비디오를 수집하는 것은 불가능하기 때문에, 탐지기는 제한된 사전 정보로도 비정상적인 것을 발견할 수 있어야 합니다. 여기서 한 가지 중요한 질문이 생깁니다. 비디오에서 보이지 않는 패턴을 탐지하는 모델을 학습할 때 사전 지식을 어떻게 활용할 수 있을까요? 가장 간단한 해결책은 보이는 패턴의 데이터를 사용하여 모델을 사전 학습한 다음 보이지 않는 유형의 예제를 사용하여 모델을 미세 조정하는 것입니다. 훈련 예제의 다양성이 약하다는 현실적인 한계에도 불구하고 이 문제를 해결하기 위해, 우리는 이상 징후 탐지를 보편적인 표현을 학습하는 것이 아니라 보이지 않는 다양한 이상 징후에 적응하는 학습으로 공식화했습니다. 이를 위해 새로운 데이터 분포에 적응하기 위한 학습 과정을 시뮬레이션하여 딥 뉴럴 네트워크의 모델 파라미터가 반드시 빠른 수렴이 아니라 더 나은 최적화를 위해 바람직한 초기 지점에 위치할 수 있다고 주장하는 메타 학습 개념[6, 7]을 활용합니다. 개별 예제 간의 분산이 큰 다양한 비정상 이벤트가 포함된 학습 에피소드를 구성하여 모델이 적절한 초기화 지점에 도달하도록 학습함으로써 새로운 비정상 이벤트에 잘 적응하도록 유도합니다.

또한 메타 학습 프레임워크에서 약한 감독 하에서 보이지 않는 비정상 이벤트 탐지가 효과적으로 수행될 수 있는지 살펴봅니다. 비디오에서 개별 비정상 이벤트의 시간적 위치에 대한 정확한 주석을 얻는 데는 많은 비용이 들기 때문에 비디오 수준의 이진 레이블(정상 대 비정상)이 있는 예제를 사용하여 기본 학습자 모델을 구축하는 것을 선호합니다. 다음 사항에 유의하세요.

는 비디오 레벨의 이상 징후에 대한 이진 주석 기반 모델 학습을 통해 시간적 영역에서 비정상적인 활동을 찾아내는 것을 목표로 합니다.

우리는 두 가지 까다로운 데이터 세트인 UCF- 범죄[3]와 상하이테크[8]에서 제안한 훈련 체계를 검증했습니다. UCF-Crime 데이터 세트는 비디오의 일부에만 비정상적인 이벤트에 대한 시간적 주석을 제공하기 때문에 실험을 위해 실측 데이터 없이 예제에 주석을 달았습니다. 실험 결과, 제안된 알고리즘이 새로운 비정상 이벤트를 탐지하는 데 있어 기준선보다 뛰어난 성능을 보였습니다. 소스 코드와 새로운 주석은 프로젝트 페이지 에서 확인할 수 있습니다.¹

제안된 접근 방식의 주요 기여 사항은 아래에 요약되어 있습니다:

- 기존 접근 방식의 한계를 해결하기 위해 보이지 않는 비정상적인 활동에 대한 학습 적응 작업으로 비디오의 이상 징후 탐지를 공식화합니다.
- 기본 학습자는 비디오 수준의 바이너리 레이블만 학습에 활용하고 최종 모델은 보이지 않는 비정상 이벤트의 위치 정보를 추정하는 약한 감독 하의 새로운 메타 학습 접근 방식을 제안합니다.
- UCF-Crime 데이터 세트에서 비정상적인 이벤트의 시간적 위치에 대해 누락된 지상 실측 데이터에 레이블을 지정했습니다.
- 라벨 증강과 상하이테크를 사용한 UCF-Crime 실험은 우리의 방법이 새로운 유형의 비정상적인 이벤트를 학습하는 데 효과적이라는 것을 보여줍니다.

백서의 나머지 부분은 다음과 같이 구성됩니다. 먼저 섹션 2에서 관련 작업에 대해 설명합니다. 섹션 3과 섹션 4에서는 각각 전체 절차와 분석 실험 결과를 설명합니다. 섹션 5에서 이 백서를 마무리합니다.

2 관련 작업

2.1 이상 징후 탐지

많은 연구자들이 비디오에서 이상 징후를 탐지하는 데 관심을 가져왔습니다 [1-5, 9-11]. 비디오가 주어지면 탐지기는 거의 관찰되지 않는 예기치 않은 사건을 찾아냅니다. 이 작업은 정의되지 않은 특성, 태생적인 복잡성, 다양한 사례로 인해 어려운 과제입니다.

심층 신경망에 기반한 생성적 모델링 기법의 발전으로 이상 징후 탐지기를

생성적 방식으로 구축할 수 있게 되었습니다 [2, 12-14]. 이들은 자동 인코더[1, 2, 15] 및 생성적 애드버셜 네트워크[16]와 같은 생성 모델을 사용하여 분포 내 데이터 포인트의 일반적인 패턴을 찾으려고 시도합니다. 생성 모델은 비정상적인 이벤트가 드물다는 가정을 바탕으로 정상적이고 일반적인 패턴을 재구성하는 방법을 학습합니다. 이 모델은 재구성 오류가 큰 예시를 분포 외 표본으로 간주합니다. 하지만 모든 동영상의

¹ <https://cv.snu.ac.kr/research/Learning-to-Adapt-to-Unseen-Abnormal-Activities/>

관점. 그 결과, 모델은 학습 데이터에 과적합하고 익숙하지 않은 정상 패턴을 비정상적으로 예측하는 경향이 있습니다.

최근 연구[3]에 따르면 기존의 생성적 접근 방식은 카메라가 포착한 정상적인 패턴을 새로운 관점에서 일반화할 수 없다고 주장합니다. 또한 기존보다 더 복잡하고 다양한 이벤트로 구성된 새로운 이상 징후 탐지용 데이터 세트인 UCF-Crime을 소개합니다. 이 데이터셋을 기반으로 [3, 4]는 변별력 있는 방식으로 이상 징후 점수를 예측할 것을 제안합니다. 이들은 이상 탐지 작업을 약한 감독 하에서 이진 분류 문제로 처리하며, 모델은 비디오 레벨의 레이블만을 기반으로 비디오에 비정상 이벤트가 포함되어 있는지 또는 정상 이벤트가 포함되어 있는지 분류합니다. 구체적으로 [3]에서는 다중 인스턴스 학습에 기반한 이진 분류기를 제안하고, [4]에서는 그래프 컨볼루션 신경망을 이용한 라벨 노이즈 클리너를 사용합니다. 그럼에도 불구하고 이러한 방법들은 여전히 일반화 가능성이 부족하며, 특히 보이지 않는 이상 유형에 직면할 때 더욱 그렇습니다. 이러한 사실은 보이지 않는 이상 징후를 효과적으로 처리하는 합리적인 초기 모델의 필요성을 제기합니다. 따라서 우리는 사전 지식으로부터 기본 정보를 얻기 위한 메타 학습 프레임워크를 제안합니다.

2.2 메타 학습

메타 학습의 목적은 메타 학습자가 기본 학습자의 학습 과정을 감독하는 학습-학습 기능을 실현하는 것입니다 [6, 17, 18]. 이 문제를 해결하기 위한 일반적인 접근 방식으로는 1) 메타 학습자가 과제 내의 유사성 메트릭에 초점을 맞추는 메트릭 학습 기반 방법 [19-21], 2) 메타 학습자가 훈련 예제 또는 클래스 임베딩 기능을 저장하는 메모리 증강 방법 [22-25], 메타 학습자가 베이스 학습자의 정보(예: 기울기 등)에 의해 직접 파라미터화되는 최적화 기반 방법 [6, 26-28] 등이 있습니다. 우리의 작업은 대중적 최적화 기반 프레임워크인 모델에 구매받지 않는 메타 학습(MAML)을 사용합니다[6]. MAML에 대한 자세한 설명은 섹션 3.3에서 제공될 것입니다. 메타 학습 체계를 활용하는 다른 애플리케이션의 선행 연구[29-34]에 이어, 저희는 메타 학습 프레임워크를 사용하여 비디오에서 이상 징후 탐지를 용이하게 합니다. 우리의 작업은 약한 감독 하에서, 즉 비디오 수준의 라벨만 사용할 수 있을 때 기본 학습자를 기반으로 메타 학습 모델을 구축한다는 점에서 기존 모델과 차별화됩니다. 도메인 일반화 과제에서 메타학습의 일반화 능력에 초점을 맞춘 선행 연구가 있습니다 [35]. 최근 [36]에서는 메타 학습을 기반으로 이상 탐지 문제를 다루고 있습니다. 그러나 생성 모델을 사용하여 장면의 정상성을 학습하려는 반면, 우리의 접근 방식

은 판별 모델을 통해 새로운^기 이상 징후에 적응하는 것을 목표로 한다는 점에서 방향성이 다릅니다.

3 방법

3.1 개요

우리의 목표는 약하게 레이블이 지정된 예제를 사용하여 새로운 유형의 이상 이벤트에 효과적으로 적응하는 이상 징후 탐지 모델을 학습하는 것입니다. 실현 불가능하기 때문에

다양성이 큰 영상에서 일반적으로 허용되는 일반적인 이상 패턴을 포착하기 위해 이상 탐지를 보편적인 표현을 학습하는 것이 아니라 다양한 이상 이벤트에 적응하는 학습으로 공식화했습니다.

D_{base} 동영상에 비정상 이벤트가 포함되어 있는지 여부에 대한 이진 레이블과 긍정적인 동영상에 대한 비정상 이벤트 범주(하위 클래스)라는 두 가지 유형의 주석이 있는 동영상 데이터 세트 D_{base} 가 존재한다고 가정합니다. D_{novel} 에 속하는 동영상에서 새로운 비정상 이벤트에 적응하는 데 최적화된 초기 모델 파라미터를 C_{novel} 서브클래스로 식별하는 방법을 학습합니다. 제안한 프레임워크의 학습 절차에서 서브클래스 정보를 사용하지는 않지만 D_{base} 와 D_{novel} 의 서브클래스는 서로 분리되어 있다는 점에 유의하세요. An-

또 다른 가정은 비디오 $v = \{v_i\}_{i=1}^N$ 가 주어졌을 때 $v \in D (= D_{base} \cup D_{novel})$ 세그먼트가 N 개인 경우, 비디오 레벨 레이블 $y \in \{0, 1\}$ 만 갖게 되는데, 여기서 1은 비디오에 비정상 세그먼트가 하나 이상 있음을 나타내고, 그렇지 않으면 0입니다. 다음 사항에 유의하세요.

약하게 감독된 검출기는 세그먼트별 레이블을 예측해야 합니다, $\hat{y} = \{\hat{y}_i\}_{i=1}^N$ 세그먼트 수준 또는 로컬라이제이션 지상 실측 데이터 없이도 가능합니다.

우리는 약한 감독을 기반으로 보이지 않는 비정상 이벤트의 로컬라이제이션 정확도를 높이기 위해 메타 학습을 활용할 것을 제안합니다. 이상 이벤트의 변이가 크고, 보이는 이상 이벤트에서 얻은 사전 지식을 보이지 않는 이상 이벤트로 일반화하기 어렵기 때문에 사전 학습된 모델로부터의 단순한 지식 전달은 우리의 시나리오에서 잘 작동하지 않을 수 있습니다. 따라서 변화가 큰 에피소드를 기반으로 메타 학습을 활용함으로써 전이 학습의 한계를 극복하고 메타 테스트를 통해 보이지 않는 이상 징후 탐지를 위한 모델을 쉽게 학습할 수 있도록 합니다. 구체적으로, D_{base} 에서 비디오의 작은 하위 집합을 샘플링하여 에피소드를 구성하고, 이 에피소드를 사용하여 메타 트레이닝을 반복합니다. 메타 테스트 단계에서는 D_{novel} 에서 샘플링한 동영상을 사용하여 모델을 미세 조정하여 최종 모델을 얻습니다. 전체 훈련 절차는 비디오 수준의 이진 클래스 레이블에만 의존한다는 점에 유의하세요.

이 섹션의 나머지 부분에서는 1) 약한 레이블에 의존하는 기본 이상 징후 탐색기와 2) 더 나은 일반화 가능한 모델을 얻기 위한 메타 학습 알고리즘을 포함하여 프레임워크의 개별 구성 요소에 대한 세부 사항을 설명합니다.

3.2 약하게 감독되는 이상 징후 탐지기

저희는 [3]에서 제안한 이상 감지 방법을 기본 감지기로 채택했습니다. 이 탐지기는 약한 감시 아래에서 각 비디오 세그먼트가 얼마나 비정상적인지 점수를 매기는 방법을 학습합니다. 각 세그먼트의 점수는 비정상 이벤트와 정상 이벤트를 구분하는 이진 분류기에 의해 주어집니다. 비디오 수준 주석만을 기반으로 세그먼트별 이상 감지기를 학습시키기 위해 순위 손실이 있는 다중 인스턴스 학습(MIL)을 사용합니다.

MIL 정상 가방과 비정상 가방 사이의 순위를 학습하기 위해 문제에서 MIL의 개념을 사용합니다. 비디오를 N 개의 세그먼트로 나누고, 각 세그먼트는 다음과 같습니다.

v_i ($i = 1, \dots, N$)로 표시됩니다. 동영상 $v = \{v_i\}_{i=1}^N$ 는 N 세그먼트로 간주됩니다.

세그먼트 중 하나 이상이 비정상인 경우, 즉 $y_i = 1$ 인 경우 포지티브 백 B_a 로 구성합니다. 그렇지 않으면 비디오는 정상이며 해당 세그먼트는 네거티브 백 B_n 를 구성합니다.

B_a 및 B_n 의 세그먼트는 ReLU와 시그모이드 함수로 완전히 연결된 세 개의 레이어로 구성된 점수 함수 $f(-)$ 를 통과하여 이상 점수를 예측합니다.

순위 손실 우리는 [3]에서와 같이 MIL에 순위 손실을 적용하여 비정상 세그먼트에 대해 정상 세그먼트보다 더 높은 점수를 생성합니다. 이 설정에서는 세그먼트 수준 레이블을 사용할 수 없으므로 양수 가방과 음수 가방 쌍에 대한 손실은 두 가방 모두에서 최대 점수를 가진 세그먼트로 정의됩니다.

$$L(\theta; \{B_a, B_n\}) = \max(0, m - \max_{v_i \in B_a} f(v_i; \theta) + \max_{v_i \in B_n} f(v_i; \theta)), \quad (1)$$

여기서 θ 는 모델 파라미터를 나타내고, v_i 는 백의 i 번째 세그먼트를 의미하며, m 은 두 백 사이의 점수 마진을 나타냅니다. 또한 손실 함수에는 시간적 평활성 손실과 희소성 손실이라는 두 가지 정규화 조건이 있습니다. 전자는 시간적으로 인접한 세그먼트의 점수가 비슷하도록 장려하는 반면, 후자는 비디오에서 비정상적인 활동이 거의 발생하지 않는다는 가정 하에 비디오의 일부 세그먼트만 높은 점수를 받도록 강제합니다. 모든 손실 조건을 결합하여 최종 손실 함수는 다음과 같이 주어집니다.

$$L(\theta; \{B_a, B_n\}) = \max(0, m - \max_{v_i \in B_a} f(v_i; \theta) + \max_{v_i \in B_n} f(v_i; \theta)) + \lambda_1 \sum_{v_i \in B_a} (f(v_i; \theta) - f(v_{i+1}; \theta))^2 + \lambda_2 \sum_{v_i \in B_a} f(v_i; \theta), \quad (2)$$

여기서 λ_1 및 λ_2 은 개별 용어의 영향을 제어하기 위한 하이퍼파라미터입니다. [3]에 따라 훈련 절차 전반에 걸쳐 $m = 1$, $\lambda_1 = \lambda_2 = 8 \times 10^{-5}$ 으로 설정합니다. $f(v_i; \theta)$ 는 시그모이드 함수의 출력이며 항상 양수이므로 방정식 (2)의 마지막 항은 세그먼트 현명한 점수 벡터의 ℓ_1 노름과 동일하다는 점에 유의하십시오.

기본 검출기 훈련 D_{base} 를 사용하여 식 (2)의 객체 함수를 기반으로 이상 검출기를 훈련합니다. 검출기를 훈련하기 위해 비디오를 여러 세그먼트로 분할하고 각 세그먼트는 16개의 연속 프레임으로 구성되며, 키네틱스 데이터세트에서 사전 훈련된 I3D 네트워크[37]에서 3D 컨볼루션 특징을 추출합니다. [3]에 설명된 대로 32개의 겹치지 않는 변환된 특징을 사용하여 각 가변 길이 비디오를 표현하고 이를 기본 검출기 모델에 공급하여 훈련합니다.

3.3 메타 트레이닝

메타 학습 단계에서 우리의 목표는 클래스 내 변동이 큰 분포에서 샘플링된 데이터를 사용하여 학습 절차를 반복적으로 시뮬레이션함으로써 모델이 새로운

유형의 비정상 사례에 적응하도록 학습하게 하는 것입니다. 이 목표를 달성하¹¹
기 위해 우리는 MAML [6]에 기반한 메타 학습 접근 방식을 채택합니다. 메타
학습 방식은 보이지 않는 유형의 데이터에 적응하는 데 적합한 최적의 초기 모
델 파라미터 집합을 찾는 것을 목표로 합니다. MAML에는 외부 메타 학습자가
없기 때문에 모델 파라미터는 경사 하강 방식으로만 업데이트됩니다.



그림 2: 일반적인 N-방향 K-샷 분류와 이상 징후 탐지 간의 에피소드 구성 방법 비교. (a) 작업은 에피소드마다 다르며 클래스 내 변화는 상대적으로 작습니다. 이미지는 *미니이미지넷*[38] 데이터 세트에서 샘플링했습니다. (b) 모든 에피소드는 비정상 클래스와 정상 클래스로 이분법적으로 분류됩니다. 비정상 클래스는 D_{base} 의 하위 클래스로 구성되므로 클래스 내 변동이 큼니다.

이상 탐지에서의 에피소드 이상 탐지를 위한 에피소드를 구성하는 방법에 대해 먼저 설명합니다. 대부분의 샷 분류 연구는 에피소드를 분류 작업으로 공식화하여 각 에피소드에 대한 모델을 고유한 클래스 집합에 최적화하고 메타 테스트를 위한 목표 작업과 메타 학습을 위한 목표 작업을 분리합니다. 이 섹션의 나머지 부분에서는 이러한 종류의 전략을 기존 메타 학습이라고 합니다.

이와 대조적으로, 이상 탐지의 모든 작업은 정상과 *비정상* 사이의 이진 분류라는 점에서 동일합니다. D_{base} 에서 정상 비디오와 비정상 비디오를 모두 c_{base} 서브클래스로 샘플링하여 태스크 T 를 구성합니다. 샘플링된 비정상 동영상은 어느 서브클래스에 속해도 상관없지만, 메타 트레이닝과 메타 테스트를 위한 서브클래스는 분리되어 있어야 합니다. 기존 메타 학습과 우리의 메타 학습의 가장 큰 차이점은 다양성의 원천에 있습니다. 이상 징후 탐지의 클래스 내 변이는 기존 메타 학습보다 훨씬 큼니다. 기존의 소수점 학습 연구는 과제 분포에 대한 모델을 일반화하려고 시도하는 반면, 우리의 접근 방식은 메타 학습 중에 개별 클래스 내의 데이터 분포, 비정상 및 정상에 대한 모델을 일반화하는 데 중점을 둡니다. 기존 메타 학습과 우리의 접근 방식은 그림 2에 나와 있습니다.

훈련 방법 비정상 및 정상 표본을 샘플링하여 각각 T^{trn} 및 T^{tst} 로 표시되는 훈련과 테스트로 구분되는 각 작업 T_i 를 구성합니다.

의 일반 동영상 c_{base} . 훈련 및 테스트 스플릿을 사용하여 메타 훈련은 기본 감지기의 이중 레벨 최적화에 의해 수행됩니다.

보이지 않는 비정상적인 활동에 적응하는 법 배우 13
먼저 일반적인 트레이닝 시나리오에서와 같이 기본 감지기를 T^{nn}
를 방정식 (2)의 목적 함수에 따라 계산합니다. 이 적응 단계를

를 내부 루프로 설정합니다. i 번째 작업 T_i 에 대해 θ 로 표시되는 기본 검출기의 모델 파라미터는 손실 함수 $L_T(\theta; T_i^{trn})$ 를 사용하여 경사 하강 방법에 의해 θ^{T_i} 로 업데이트되며, 이는 다음과 같이 표현됩니다.

$$\theta^{T_i} = \theta - \alpha \nabla L_{\theta T_i}(\theta; T_i^{trn}), \quad (3)$$

여기서 α 는 기본 감지기의 학습 속도입니다.

다음으로, 적응된 기본 감지기는 T^{tst} 에 의해 평가되고, 메타 학습기는 결과 오류를 사용하여 최적화됩니다. 메타 최적화에는 적응 단계가 포함되어 있으므로 이를 외부 루프라고도 합니다. 메타 학습기는 $T(\theta^{T_i}; T^{tst})$ 에 의해 최적화됩니다. MAML 기반 접근 방식에서 메타 업데이트는 다음과 같습니다.

메타 목표 함수를 사용하여 적응 단계 전에 기본 검출기의 모델 매개 변수(θ 로 표시)를 업데이트하여 수행되며, 이는 다음과 같이 주어집니다.

$$\sum_{\theta} L_T(\theta^{T_i}; T_i^{tst}) = \sum_{T_i \sim p(T)} L_T(\theta - \alpha \nabla L_{\theta T_i}(\theta; T_i^{trn}); T_i^{tst}). \quad (4)$$

따라서 모델 매개 변수는 다음과 같이 메타 업데이트됩니다.

$$\theta = \theta - \beta \nabla_{\theta} \sum_{T_i \sim p(T)} L_T(\theta^{T_i}; T_i^{tst}), \quad (5)$$

여기서 β 는 메타 학습률입니다.

3.4 메타 테스트

메타 테스트 단계에서는 모델이 새로운 유형의 비정상 이벤트에 잘 적응하는지 평가합니다. 이를 위해 D_{novel} 에서 샘플링한 예제를 사용하여 메타 테스트용 에피소드를 구성하여 c_{novel} 정상 서브클래스와 분리된 c_{base} 서브클래스의 비정상 이벤트에 대한 모델을 미세 조정합니다.

유효성 검사 집합 D_{val} 가 없으므로 D_{trn} 데이터 세트의 크기가 작기 때문입니다, 이를 활용하여 10배 교차 검증을 수행합니다. D_{trn} 숫자를 결정하기 위해 미세 조정을 위한 반복 횟수입니다.

4 실험

4.1 데이터 세트

두 가지 벤치마크 데이터 세트, 즉 라벨 증강이 적용된 UCF-Crime [3]과 ShanghaiTech [8]를 대상으로 실험을 진행했습니다.

UCF-Crime 이 대규모 데이터 세트는 다양한 상황에서 촬영된 실제 감시 영상으로 구성되어 있습니다. 학대, 체포, 방화, 폭행, 강도, 폭발, 싸움, 도로 사고, 강도, 총격, 절도, 절도, 기물 파손 등 13개의 비정상적인 이벤트 하위 클래스가 포함되어 있습니다. 이 데이터 세트에는 비정상 동영상 950개와 정상 동영상 950개를 포함하여 1,900개의 트리밍되지 않은 동영상이 포함되어 있습니다.

표 1: 표 1: 기존 평가 방법의 성능 과대 추정 문제. [3]의 이상 분할을 사용하여 전체 테스트 비디오를 평가할 때 기본 탐지기의 AUC 점수는 약 84%입니다. 이 수치는 [3]에서 보고된 수치보다 약간 높는데, 이는 비디오 표현을 위해 I3D 네트워크[37]의 2스트림 특징을 사용하기 때문입니다. 그러나 테스트 세트에서 정상 동영상을 제외하면 성능이 약 68%까지 떨어집니다.

클래스	AUC(%)	# 테스트 샘플 수
비정상	68.35	140
비정상+정상	84.39	290

이 데이터 세트에는 우리 작업과의 직접적인 호환성을 방해하는 몇 가지 중요한 한계가 있습니다. 첫째, [3]에서 이상 징후 탐지를 위해 제공한 원래 훈련 및 테스트 분할의 하위 클래스 분포가 심각하게 불균형하다는 점입니다. 따라서 [3]에서 제공하는 행동 인식 분할을 사용하여 실험을 수행합니다. 동작 인식 분할에서 모든 하위 클래스에는 38개의 훈련용 비디오와 12개의 테스트용 비디오가 있습니다. 둘째, [3]은 이상 감지 분할에서 테스트 비디오에 대한 비정상 이벤트의 시간적 지속 시간을 제공하는 반면, 동작 인식 분할의 테스트 세트에 있는 일부 비디오에는 이러한 주석이 없습니다. 성능 평가를 위한 완벽한 데이터 세트를 만들기 위해 데이터 세트의 일부 동영상에 대해 이상 이벤트의 실측값 간격을 주석으로 추가합니다.

상하이테크 이것은 13개의 다른 장면에서 437개의 비디오로 구성된 중간 규모의 데이터 세트입니다. 모든 훈련 영상이 정상 영상이므로 [4]에서 제안한 새로운 분할을 사용합니다. 또한 메타 학습을 위한 비정상 이벤트가 포함된 동영상의 수가 충분하지 않기 때문에 이 데이터셋은 메타 테스트에만 사용합니다. 이 데이터 세트의 실험은 제안한 방법의 교차 데이터 세트 일반화 성능을 보여줄 수 있다고 생각합니다.

4.2 평가 지표 및 프로토콜

기존 연구 [2-4]에 따라 프레임별 수신기 동작 특성(ROC) 곡선을 도출하고 곡선 아래 면적(AUC) 점수를 계산합니다. 하지만 다음 두 가지 부분에서 기존 평가 방법과 다른 점이 있습니다.

먼저 비정상 동영상에 대해서만 AUC 성능을 평가합니다. 특히 비정상 동영상과 정상 동영상을 모두 포함할 경우 정상 프레임 수가 비정상 프레임 수보다 훨씬

씬 많기 때문에 두 등급의 동영상을 모두 사용하여 성능을 평가하면 표 1과 같이 정확도 과대 추정으로 편향된 결과가 나옵니다. 따라서 실험에서 AUC 점수를 계산할 때 정상 동영상은 제외했습니다. 통계를 얻기 위해 수정된 분할과 주석 대신 원본 분할과 주석을 사용한다는 점에 유의하세요.

둘째, 기존 방식은 모든 동영상에서 수집한 모든 프레임을 사용하여 점수를 추정하는 반면, 저희는 각 동영상에 대한 평균 프레임별 AUC 점수를 평가합니다.

테스트 데이터 세트입니다. 이는 전체 성능이 평균 길이인 약 4×10^3 프레임보다 훨씬 긴 10개의⁵ 프레임에 달하는 매우 긴 동영상의 작은 하위 집합에 의해 좌우되는 경우가 많기 때문입니다.

4.3 실험 설정 및 구현 세부 정보

novel 메타 학습이 적절한 초기화 지점을 제공한다는 주장을 검증하기 위해 1) 무작위로 초기화된 모델, 2) D_{base} 에서 사전 학습된 모델, 3) D_{base} 에서 메타 학습된 모델에서 탐지기를 미세 조정하여 주어진 다음 세 가지 시나리오를 비교합니다.

실험을 위해 [3]에서 제안한 검출기를 다시 구현하여 기본 학습자로 사용했습니다. 구현 방식은 다음 세 가지 부분을 제외하고는 [3]과 동일합니다. 첫째, [3]에서 사용된 C3D 특징[39] 대신 Kinetics 데이터세트에서 훈련된 사전 훈련된 2스트림 I3D 특징[37]을 활용했으며, 광 흐름은 TVL1 알고리즘[40]으로 계산하고 두 가지 모달리티(RGB와 광 흐름)의 융합은 특징의 연결에 의해 주어집니다. 둘째, MAML 모델[6]의 훈련이 불안정했기 때문에 드롭아웃 레이어 [41]를 제거했습니다. 마지막으로 Adagrad 대신 아담 옵티마이저를 사용했습니다. 사전 학습을 위해 비정상 클래스와 정상 클래스 모두에서 30개의 비디오를 샘플링하여 미니 배치를 구성했습니다. D_{base} 를 동작 인식 분할에 따라 훈련용과 검증용 비디오로 분할한 후 학습률로 모델을 훈련했습니다.

10^{-3} 유효성 검사 AUC 점수가 최고점에 도달할 때까지.

메타 학습을 위해 비정상 및 정상 클래스의 두 범주에서 훈련용 샘플 10개와 테스트용 샘플 30개를 사용하여 각 에피소드를 구성합니다. 내부 루프의 학습률은 10^{-3} 으로 설정하고, 메타 학습률인 외부 루프의 학습률은 10^{-5} 으로 설정합니다. 메타 배치 크기 15로 3,000번의 외부 반복으로 모델을 훈련하고 내부 루프 최적화를 위해 SGD 옵티마이저를 사용했습니다.

학습률 10^{-3} 의 초기화 방법과 관계없이 D_{novel} 에서 미세 조정을 수행했습니다. 최대값으로 300회 반복하여 모델을 미세 조정하고 10배 교차 검증을 수행하여 최상의 모델을 선택했습니다.

4.4 정량적 결과

UCF-Crime 데이터 세트에서 두 개의 동작 인식 데이터 분할에 대한 실험을 진행했습니다. UCF-Crime은 동영상에서 이상 징후를 탐지하기 위한 가장 큰 데

이터 세트이지만, 메타 학습 실험을 수행하기에는 여전히 너무 작습니다. 실험을 위해 13개의 서브태스크를 생성했으며, 각 서브태스크에는 서로 다른 새로운 서브클래스가 있고 나머지 12개의 서브클래스는 *Dbase*를 구축하는 데 사용되었습니다.

표 2에는 13개 하위 작업 모두에 대한 AUC 점수가 표시되어 있으며, 4.3절에 설명된 세 가지 시나리오의 결과를 비교합니다. 각 서브클래스의 통계가 서로 다르기 때문에 공정한 비교를 위한 최적의 모델을 식별하는 것은 간단하지 않습니다. 따라서 각 하위 작업의 평가를 위해 다음 두 가지 모델을 선택합니다. 먼저 10개의 모델을 샘플링합니다.

표 2: 각 타겟 하위 클래스에 대한 세 가지 시나리오 간의 AUC 점수(%) 비교. 비교한 모든 방법의 미세 조정 프로세스는 동일하지만 미세 조정의 초기 시점은 다릅니다. S로 표시된 알고리즘에서는 모델이 무작위 스크래치 모델에서 미세 조정됩니다. 시나리오 P에서 모델은 미세 조정 전에 D_{base} 로 사전 학습됩니다. 초기 모델을 얻기 위해 D_{base} 로 메타 학습을 수행하는 M_S 과 M_G 으로 표시되는 두 가지 버전의 메타 학습 접근 방식은 "샘플링"과 "전역"이라는 두 가지 모델 선택 전략에 해당합니다. "샘플링"과 "글로벌"에 대한 자세한 내용은 4.4절에 설명되어 있습니다. 굵은 글씨로 표시된 숫자는 각 서브클래스에 대한 최고 정확도에 해당합니다.

분할	Algo.	Abus Arre Arso Assa Burg Expl Figh Road Robb Shoo Shop Stea Vand	평균
1	S	62.99 67.91 56.93 80.05 72.02 63.62 70.94 73.19 77.86 75.81 57.70 66.77 72.86	69.13
	P	69.71 67.57 60.22 81.18 77.51 71.85 70.65 77.11 80.48 82.69 52.37 65.41 74.61	71.64
	MS	70.93 72.05 61.26 82.67 81.08 73.32 71.35 76.72 82.56 82.85 59.19 70.75 77.30	74.00
	MG	69.89 71.30 59.97 82.19 78.75 73.32 69.96 74.85 82.56 82.85 56.20 66.37 75.33	72.58
2	S	79.64 61.08 77.57 77.86 74.10 77.31 79.24 74.96 80.02 79.60 67.29 65.63 75.26	74.58
	P	73.60 73.91 83.81 79.62 75.22 73.62 83.21 74.27 71.32 80.13 64.96 72.56 78.56	75.75
	MS	79.01 76.36 82.34 79.75 77.20 73.32 84.39 74.14 76.16 81.58 67.27 77.41 79.26	77.55
	MG	76.01 72.22 83.82 79.08 76.83 68.34 84.62 75.07 73.92 80.92 65.97 77.02 79.25	76.39

을 균일하게 샘플링한 메타 반복에서 각 하위 작업에 대해 추출하고 각 하위 클래스에 가장 적합한 모델을 선택합니다. 우리는 이 모델 선택 전략을 "샘플링"이라고 부릅니다. 둘째, 평가를 보다 엄격하게 하기 위해 동일한 메타 반복에서 모든 하위 클래스를 처리할 수 있는 전역 모델을 선택하는데, 이를 "전역"이라고 합니다. 두 분할 모두에서 메타 훈련된 모델이 다른 모델보다 평균적으로 더 나은 성능을 보였으며, 제안된 모델은 대부분의 하위 클래스에서 정확도를 향상시킨 반면, 두 번째 분할에서 폭발(*Explosion*)과 절도(*Shoplifting*)를 비롯한 몇 가지 사례에서 사전 훈련과 메타 훈련으로 정확도가 떨어지는 것을 관찰했습니다. 이는 아마도 이 두 하위 클래스의 데이터 분포가 다른 하위 클래스와 크게 다르고 사전 지식이 도움이 되지 않기 때문일 것입니다. 그러나 절도의 경우, 우리의 접근 방식이 사전 학습 방법보다 약 5% 정도 더 우수한 성능을 보였는데, 이는 사전 학습이 도움이 되지 않는 시나리오에서도 제안한 기법이 효과적이라는 것을 나타냅니다.

상하이테크 모델의 일반화 능력을 검증하기 위해 상하이테크 데이터 세트에 대한 추가 실험을 수행했습니다[8]. 4.1절에서 언급했듯이, 이 데이터셋은 크기가 작기 때문에 메타트레이닝을 수행하는 것이 불가능합니다. 상하이테크의

사례는 캠퍼스 생활의 비정상적인 사건에 속하며, 이는 UCF-Crime과 비교하여 고유한 것이므로, 상하이테크 데이터셋을 D_{novel} 로 간주하고 UCF-Crime을 D_{base} 로 사용하여 사전 지식을 학습할 수 있습니다. 상하이테크 실험의 모든 하이퍼파라미터는 UCF-Crime 실험의 하이퍼파라미터와 동일합니다.

표 3: 상하이테크 데이터 세트의 정량적 결과. P와 M의 경우_S, UCF- 범죄 데이터 세트가 모델 학습에 사용됩니다. 그런 다음 상하이테크의 훈련 분할을 통해 각 모델을 미세 조정하고 상하이테크의 테스트 분할을 통해 최종 평가를 수행합니다. 그 결과 메타 초기화 모델이 다른 모델인 S와 P보다 새로운 변칙에 더 잘 적응하는 것으로 나타났습니다.

알고리즘	AUC(%)
S	79.53
P	79.34
MS	84.70

표 3은 세 가지 훈련 시나리오의 결과를 보여줍니다. 제안된 전략은 다른 훈련 방법보다 우수한 성능을 보였으며, 이는 메타 학습 접근 방식으로 학습한 데이터 세트의 지식이 다른 데이터 세트로 전이될 수 있음을 의미하며, 제안된 프레임워크가 다양한 상황에서 이상 징후를 찾아내는 데 유망한 초기 모델을 제공한다는 것을 의미합니다.

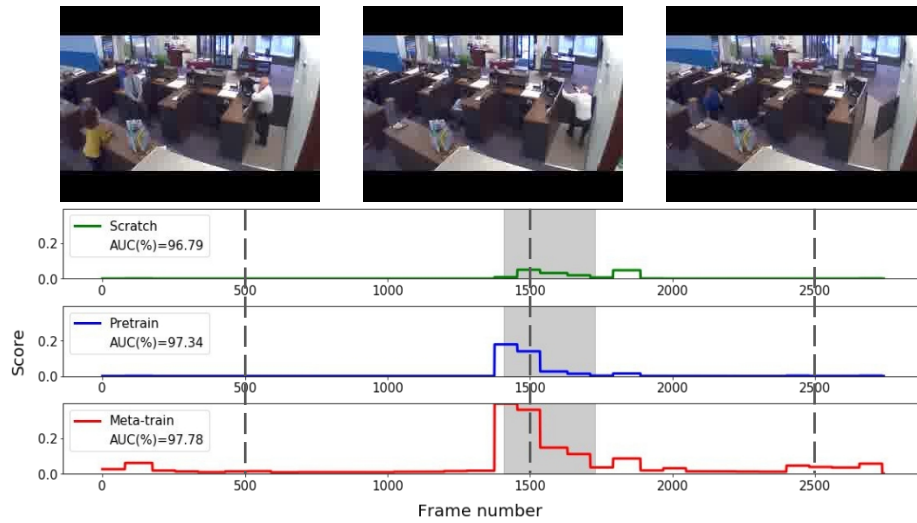
4.5 정성적 결과

그림 3은 (a) UCF-Crime 및 (b) ShanghaiTech 데이터 세트의 두 테스트 비디오에 대한 세 가지 훈련 시나리오의 정성적 결과를 보여줍니다. 그래프에서 음영 처리된 영역은 비정상 이벤트의 실측값 간격에 해당합니다. 각 비디오와 모델에 대한 ROC 곡선(AUROC 또는 AUC) 아래 영역도 그래프에 표시되어 있습니다. AUC 지표는 점수 순위에 따라 계산되므로 세 가지 방법의 성능은 모두 비슷해 보입니다. 그러나 메타트레이닝된 모델의 점수는 다른 두 가지 방법보다 변별력이 더 높습니다. 즉, 처음부터 학습된 모델이나 사전 학습된 모델은 오탐 및/또는 거짓 경보가 발생하기 쉽습니다. 이러한 관찰은 제안된 접근 방식이 더 까다로운 예제에서 다른 접근 방식보다 더 강력할 수 있음을 의미합니다. 더 많은 이상 징후 탐지 샘플 결과를 점수와 함께 보충 문서에 첨부할 예정입니다.

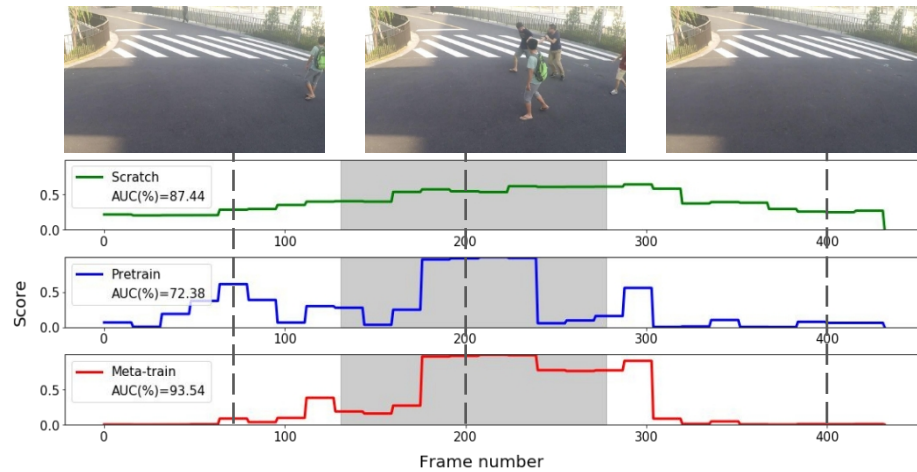
4.6 추가 분석

메타 학습 체계가 새로운 이상 징후에 적응하는 데 실제 영향을 미치는지 분석하기 위해 메타 테스트 단계에서 미세 조정하는 동안의 성능 곡선을 UCF-Crime 및 ShanghaiTech 데이터 세트로 그려보고 그 결과를 그림 4에 제시합니다. 다음은 미세 조정 곡선 관찰을 통해 얻은 교훈입니다.

첫째, 이상 징후 탐지^기는 사전 지식의 이점을 활용합니다. 대부분의 경우, 사전 학습된 모델과 메타 학습된 모델 모두 초기 단계에서 스크래치 사례보다 더 나은 성능을 보여줍니다. 이러한 관찰을 통해 이 모델은 직접적인 사전 지식이 없어도 새로운 유형의 이상 이벤트를 탐지할 수 있는 유용성을 가지고 있다는 결론을 내릴 수 있습니다. 둘째, 미세 조정 과정에서 모델의 성능이 저하되는 경우가 있습니다. 그림 4(b) 및



(a) UCF-범죄 데이터 세트의 Shooting048 시퀀스



(b) 상하이테크 데이터 세트의 03 0059 시퀀스

그림 3: (a) UCF-Crime 및 (b) ShanghaiTech 데이터 세트의 정성적 결과. 세 가지 다른 방법의 점수가 음영 처리된 영역으로 표시된 실측치와 함께 표시됩니다.

(d)는 반복을 거듭할수록 학습 곡선이 아래로 내려가는 것을 보여줍니다. 이러한 경우 노이즈가 심하거나 클래스 내 변동이 큰 데이터 샘플이 존재하며, 약한 감독만으로는 모델이 이상 징후를 탐지하기 어렵다는 결론을 내릴 수 있습니다. 이는 훈련된 탐지기의 내재적 약점에서 비롯됩니다.

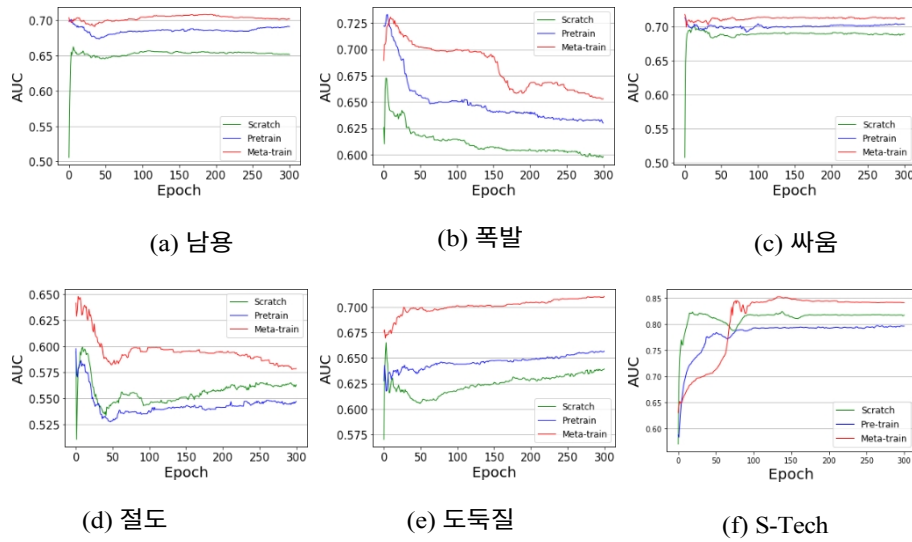


그림 4: 세 가지 훈련 시나리오에 대한 개별 하위 클래스의 미세 조정 곡선

약한 감독만을 기반으로 합니다. 지면 제약으로 인해 나머지 하위 클래스에 대한 학습 곡선은 보충 문서에 포함되어 있습니다.

5 결론

본 논문에서는 기존 방법의 한계를 보완하는 약지도 학습을 통한 동영상 이상 검출 방법을 제시하여, 보이지 않는 다양한 데이터 샘플에 대한 이상 검출의 한계를 극복하고자 합니다. 이를 위해 메타러닝을 고도화하여 보이지 않는 이상 이벤트 유형에 효과적으로 적응할 수 있는 학습 전략을 제안했습니다. 이상 이벤트 탐지와 잘 부합하는 에피소드를 구성하여 모델을 메타 학습시켰습니다. UCF-Crime과 상하이텍의 실험 결과를 통해 제안한 기법이 적용된 모델이 새로운 유형의 이상 영상에 성공적으로 적응하는 것을 확인했으며, 사전 학습된 모델과 비교하여 적응 품질에서 메타러닝의 효과를 검증했습니다. 또한, 주석의 완전성과 데이터 불균형 측면에서 UCF-Crime 데이터 세트의 한계를 지적하고, 이러한 근거 자료가 없는 영상에 대해 비정상적인 활동에 대한 시간적 주석을 유연하게 보완했습니다.

감사의 말 본 연구는 SK텔레콤 T3K 비전 AI 프로젝트 엑셀런스 센터와 정보통신기술진흥센터(IITP)의 정부(과기정통부) 지원 과제[2017-0-01779, 2017-0-01780]의 일부 지원으로 수행되었습니다.

참조

1. 공, D., 리우, L., 레, V., 사하, B., 만수르, M.R., 벤카테쉬, S., 헨겔, A.v.d.: 비정상 감지를 위한 정상성 기억: 비지도 비정상 감지를 위한 메모리 증강 딥 오토인코더. In: ICCV. (2019)
2. 하산, M., 최, J., 노이만, J., 로이-쇼두리, A.K., 데이비스, L.S.: 비디오 시퀀스에서 시간적 규칙성 학습. In: CVPR. (2016)
3. 술타니, W., 첸, C., 샤, M.: 감시 비디오에서 실제 이상 감지. In: CVPR. (2018)
4. Zhong, J.X., Li, N., Kong, W., Liu, S., Li, T.H., Li, G.: 그래프 컨볼루션 라벨 노이즈 클리너: 이상 탐지를 위한 플러그 앤 플레이 액션 분류기 훈련. In: CVPR. (2019)
5. 안티씨, B., 옴머, B.: 이상 감지를 위한 비디오 파싱. In: ICCV. (2011)
6. 핀, C., 아빌, P., 레빈, S.: 딥 네트워크의 빠른 적응을 위한 모델에 구매받지 않는 메타러닝. In: ICML. (2017)
7. 라구, A., 라구, M., 벤지오, S., 빈알스, O.: 빠른 학습 또는 기능 재사용? maml의 효과를 이해하기 위해. arXiv 사전 인쇄물 arXiv:1909.09157 (2019)
8. Liu, W., W. Luo, D.L., Gao, S.: 이상 징후 탐지를 위한 미래 프레임 예측 - 새로운 기준선. In: CVPR. (2018)
9. 크라츠, L., 니시노, K.: 시공간 모션 패턴 모델을 사용한 극도로 혼잡한 장면에서의 이상 감지. In: CVPR. (2009)
10. 우, S., 무어, B.E., 샤, M.: 혼잡한 장면에서 이상 감지를 위한 라그랑지안 입자 궤적의 카오스 불변성. In: CVPR. (2010)
11. Zhao, B., Fei-Fei, L., Xing, E.P.: 동적 스파스 코딩을 통한 동영상에서 비정상적인 이벤트의 온라인 감지. In: CVPR. (2011)
12. Lu, C., Shi, J., Jia, J.: Matlab에서 150fps로 비정상 이벤트 감지. In: ICCV. (2013)
13. Xu, D., Ricci, E., Yan, Y., Song, J., Sebe, N.: 비정상적인 이벤트 감지를 위한 모양과 동작의 심층 표현 학습. In: BMVC. (2015)
14. Zhao, Y., Deng, B., Shen, C., Liu, Y., Lu, H., Hua, X.S.: 비디오 이상 감지를 위한 시공간 자동 인코더. In: 제25회 ACM 멀티미디어 국제 컨퍼런스 논문집. (2017)
15. Nguyen, T.N., Meunier, J.: 출현-모션 대응을 통한 비디오 시퀀스에서의 이상 감지. In: ICCV. (2019)
16. Vu, H., Nguyen, T.D., Le, T., Luo, W., Phung, D.: 다단계 표현을 사용한 비디오의 강력한 이상 탐지. In: AAAI. (2019)
17. Lake, B.M., Ullman, T.D., Tenenbaum, J.B., Gershman, S.J.: 사람처럼 학습하고 사고하는 기계 구축. 행동 및 뇌 과학 40 (2017)
18. Thrun, S., Pratt, L.: 학습하기: 소개 및 개요. In: 배우는 법 배우기. Springer (1998)
19. 스넬, J., 스위스키, K., 제멜, R.: 샷 학습을 위한 프로토타입 네트워크. In: NIPS.

- (2017)
20. Sung, F., Yang, Y., Zhang, L., Xiang, T., Torr, P.H., Hospedales, T.M.: 비교하는 법 배우기: 소수 샷 학습을 위한 관계 네트워크. In: CVPR. (2018)
 21. Vinyals, O., Blundell, C., Lillicrap, T., Wierstra, D. 외: 원샷 학습을 위한 매칭 네트워크. In: NIPS. (2016)
 22. 미슈라, N., 로하니네자드, M., 첸, X., 아빌, P.: 간단한 신경 주의적 메타 학습자. In: ICLR. (2018)
 23. Munkhdalai, T., Yu, H.: 메타 네트워크. In: ICML. (2017)

24. Oreshkin, B., Lo'pez, P.R., Lacoste, A.: Tadam: 개선된 소수 샷 학습을 위한 작업 의존적 적응 메트릭. In: NeurIPS. (2018)
25. 산토로, A., 바르투노프, S., 보트비닉, M., 비어스트라, D., 릴리크랩, T.: 메모리 증강 신경망을 사용한 메타학습. In: ICML. (2016)
26. Andrychowicz, M., Denil, M., Gomez, S., Hoffman, M.W., Pfau, D., Schaul, T., Shillingford, B., De Freitas, N.: 경사 하강으로 경사 하강으로 학습하는 법. In: NIPS. (2016)
27. Antoniou, A., Edwards, H., Storkey, A.: 맘엘을 훈련시키는 방법. In: ICLR. (2019)
28. Rusu, A.A., Rao, D., Sygnowski, J., Vinyals, O., Pascanu, R., Osindero, S., Hadsell, R.: 잠재 임베딩 최적화를 통한 메타 학습. In: ICLR. (2019)
29. 최재원, 권재원, 이강민: 실시간 목표 인식 시각 추적을 위한 딥 메타 학습. In: ICCV. (2019)
30. Gui, L.Y., Wang, Y.X., Ramanan, D., Moura, J.M.: 메타 학습을 통한 소수 샷 인간 동작 예측. In: ECCV. (2018)
31. Park, E., Berg, A.C.: 메타트래커: 시각적 객체 추적기를 위한 빠르고 강력한 온라인 적응. In: ECCV. (2018)
32. Shaban, A., Bansal, S., Liu, Z., Essa, I., Boots, B.: 시맨틱 세분화를 위한 원샷 학습. In: BMVC. (2017)
33. Wang, K., Liew, J.H., Zou, Y., Zhou, D., Feng, J.: Panet: 프로토타입 정렬을 사용한 몇 샷 이미지 시맨틱 분할. In: ICCV. (2019)
34. Yan, X., Chen, Z., Xu, A., Wang, X., Liang, X., Lin, L.: Meta r-cnn : 인스턴스 수준의 로우 샷 학습을 위한 일반 솔버를 향하여. In: ICCV. (2019)
35. Li, D., Yang, Y., Song, Y.Z., Hospedales, T.M.: 일반화 학습: 도메인 일반화를 위한 메타 학습. In: AAAI. (2018)
36. Lu, Y., Yu, F., Reddy, M.K.K., Wang, Y.: 몇 샷 장면 적응형 이상 검출. In: ECCV. (2020)
37. Carreira, J., Zisserman, A.: 쿼바디스, 행동 인식? 새로운 모델과 동역학 데이터 세트. In: CVPR. (2017)
38. 라비, S., 라로셀, H.: 소수의 샷 학습을 위한 모델로서의 최적화. In: ICLR. (2017)
39. Tran, D., Bourdev, L., Fergus, R., Torresani, L., Paluri, M.: 3D 컨볼 루션 네트워크를 사용한 시공간-포칼 특징 학습. In: ICCV. (2015)
40. Zach, C., Pock, T., Bischof, H.: 실시간 tv-l1 광 흐름을 위한 이중성 기반 접근 방식. In: 공동 패턴 인식 심포지엄, Springer (2007) 214-223쪽
41. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: 드롭아웃: 신경망의 과적합을 방지하는 간단한 방법. JMLR (2014)