준지도학습의 이상행동감지에서의 이상행동종류별 균형의 중요성 분석

박태경, 박현정, *홍제형
한양대학교 융합전자공학부
tkp0305@hanyang.ac.kr, hjung4337@hanyang.ac.kr, * jhh37@hanyang.ac.kr

Analyzing the Importance of Balanced Action Classes in Weakly Supervised Video Anomaly Detection

Tae Kyeong Park. Hyeon Jeong Park, *Je Hyeong Hong
Department of Electronic Engineering, Hanyang University

요 약

준지도학습 기반의 동영상 이상행동감지는 구하기 어려운 프레임 단위 레이블이 필요하지 않아 더 많은 동영상을 학습에 활용 가능한 장점이 있어 관련 연구가 활발히 진행되고 있다. 최근 제안된 기법들은 주로 UCF-Crime 이라는 실제 CCTV 동영상 데이터셋을 활용하고 있는데, 본 데이터셋은 학습 영상과 테스트 영상에서 이상행동 클래스 별 분포도가 균등하지 않다. 본 연구에서는 해당 불균형으로 인해 학습 모델이 특정 행동 클래스에 과적합될 수 있음을 보이며, 이러한 불균형을 해결하기 위해 Class-Balanced Multiple Instance Learning Loss 를 제안한다. 이를 통해 기존에 특정 클래스에 편중되었던 모델이 이상행동 종류에 좀 더 균등한 성능을 낼 수 있음을 보여준다. 특히 단순히 클래스별 정확도가 제로섬(zero sum)으로 증감하는 것이 아니라 전체적인 이상행동 판별 정확도 또한 향상됨을 실험 결과를 통해 확인할 수 있다.

1. 서론

최근 몇 년간 범죄 예방과 감시의 목적으로 현실 감시 시스템 (CCTV)의 개수는 지속적으로 증가해오고 있는 추세다. 하지만 범죄나 사고가 일어나는 이상현상 (교통사고, 폭발, 폭행 등)은 정상현상에 비해 극히 일부의 상황으로 나타나기 때문에 이러한 이상현상 검출을 위해 사람이 감시하는 것은 고비용과 지루한 문제 해결 방식이다. 그렇기 때문에 CCTV 내에서 자동으로 이상 현상을 감지하는 지능적 현실 감시 시스템의 기술 발전에 대한 수요는 지속적으로 증가하는 추세이다.

이상행동감지(Video Anomaly Detection 이하 VAD)는 지금까지 큰 3 가지 부류인 지도학습 VAD, 준지도학습 VAD, 비지도학습 VAD 으로 연구가 진행되어왔다. 비지도학습 VAD 는 정상행동만 학습하여 기존과 다른 이상행동이 발생하면 감지하는 방식이지만 이상현상에 관한 정보가 없기 때문에 보통 좋지 않은 성능을 보인다. 지도학습 VAD 는 이상현상 비디오 전체를 프레임 단위로 구간을 레이블링(labeling)하는 것은 많은 시간과 노동을 필요로 하기 때문에 많은 데이터를 얻기 힘들다. 이러한 이유로 준지도학습 VAD 환경에서 많은 연구가 진행되고 있는 추세이다.

준지도학습 VAD 에서 가장 많이 사용된 방법은 Multi-Instance Learning(MIL)[1]이다. 비디오를 정상 비디오(Normal video), 이상 비디오(Abnormal video) 두가지로 나누어 모든 비디오를 32 개의 스니펫으로 나눈다. 모든 스니펫의 이상현상 점수를 계산하여 계산된 정보를 바탕으로 이상 비디오에서 가장 이상현상 점수가 높은 스니펫 하나와 정상 비디오에서 가장 이상현상 점수(Anomaly score)가 높은 스니펫 하나를 비교하여 두 스니펫간의 거리를 멀리 벌리는 것이 MIL 방식의 주된 방법이다.

이러한 MIL 의 방식은 VAD 분야에서 높은 성능을 보여주었고 이를 토대로 많은 선행연구가 UCF-Crime

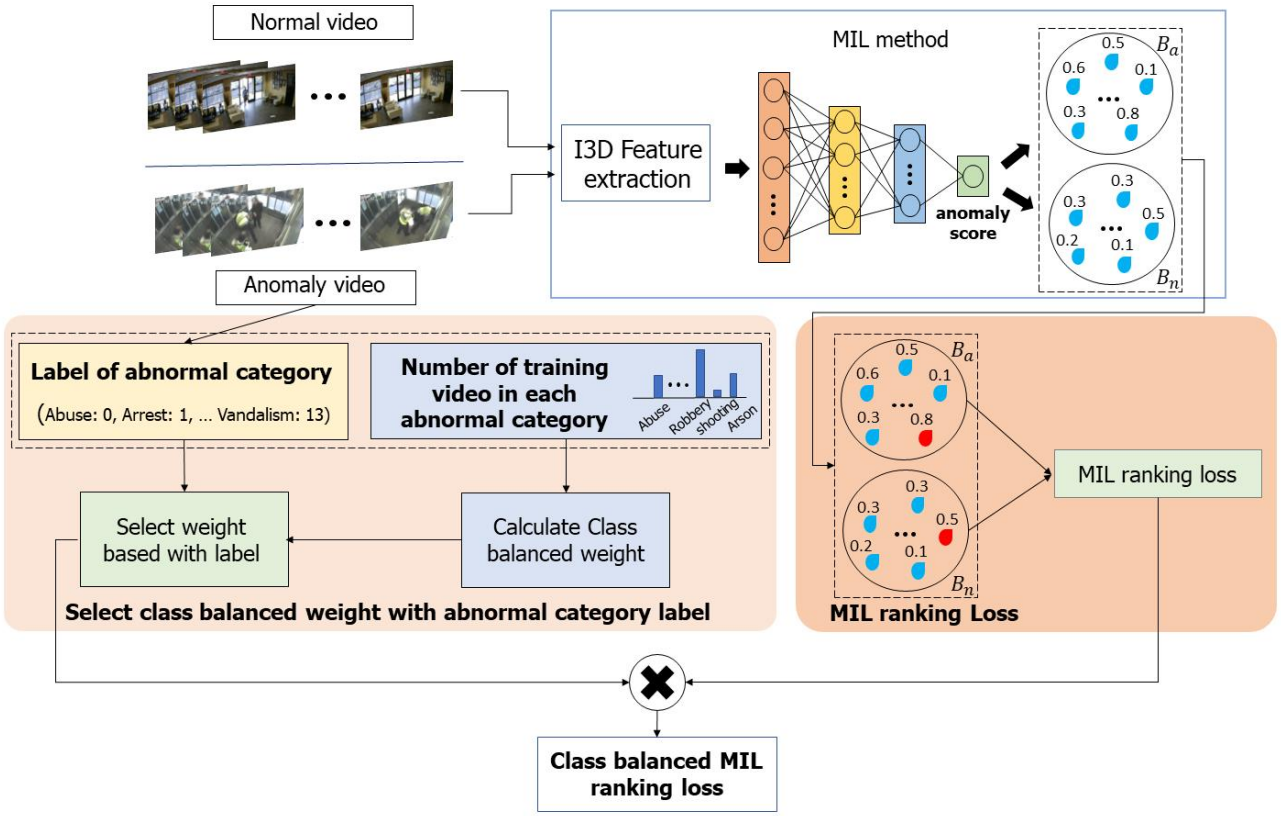

그림 1. CB-MIL loss 의 전체구조. 일반적인 MIL 방식에서는 클래스 정보 없이 무작위의 배치에서의 학습이 진행되었다면 본 논문에서 제시하는 방식은 클래스별 학습 데이터 개수 정보에 따라 Class balanced weight 를 계산하고 영상 클래스 정보 레이블에 따라 가중치를 선택하여 기존 MIL ranking Loss 와 곱하여 균형 있는 VAD 학습을 진행한다.

데이터 셋에 관한 성능 향상도를 보이고 있다. 하지만 UCF-Crime 데이터 셋의 가장 큰 문제점은 이상현상에 관한 비디오의 클래스(class)를 13 개 지니고 있지만 불균형한 분포도를 보인다는 것이다. 이러한 불균형한 분포도를 갖는 데이터셋에서 학습(training)하고 실험(test)을 진행하여 결과를 도출해 냈을 때 특정 행동 클래스에 관한 이상현상만을 잘 감지하는 과적합된 모델이 나타날 수 있다는 점을 본 논문에서는 문제점으로 삼는다. 다양한 선행연구에서 Area Under Cover(AUC)의 향상을 나타내고 있지만 전체적인 종류의 이상현상을 잘 구분하는 모델인지 확인할 필요가 있다는 점이다.

본 논문에서는 이러한 문제점들을 해결하기 위해 MIL 기법으로 학습 시 Class-Balanced Loss(CB-loss)[2]에서 사용되는 가중치(weight)를 사용한 Class-Balanced Multiple Instance Learning Loss(CB-MIL Loss)를 제시한다. 제시된 방법을 기반으로 학습을 진행하면 전반적인 학습의 균형과 성능의 향상까지 이루어 낼 수 있다는 사실을 기술한다.

2. 관련 연구

MIL[1] 기법은 준지도학습의 VAD 에서 주로 사용되었던 방법이다. 비디오를 정상 비디오, 이상 비디오 두가지로 분류하여 비디오를 각각 32 개의 스니펫으로 나눈다. 정상 비디오의 경우 모든 스니펫이 이상행동을 지니지 않고 있으며 이상비디오의 경우에는 최소 하나의 스니펫이 이상 현상을 가지고 있다고 가정한다. 영상 각각의 32 개의 스니펫을 C3D[3]로 특징(feature)을 추출하고 모든 스니펫의 이상현상 점수를 추출한다. 이때 이상현상 점수는 높을수록 이상행동일 확률이 높다는 것을 의미한다. 이상 비디오의 스니펫인 경우 양성 가방(B_a)에 정상 비디오의 스니펫인 경우 음성 가방(B_n)에 넣어 각각의 가방에서 가장 높은 이상현상 점수를 가지는 스니펫 하나씩을 선정 후 MIL-ranking Loss 에 적용하여 정상 비디오 스니펫과 이상 비디오 스니펫 간의 점수사이를 멀게 만들어 모델을 최적화한다. 이러한 MIL 기법은 논문이 나온 당시 가장 높은 성능을 이루었고 추후의 연구들 또한 MIL 의 방식을 발전시키는 방향으로 연구가 이루어졌다.

CB Loss[2]는 데이터 불균형을 해결하기 위한 Cost-

Sensitive reweighting 방법중의 하나로 Effective number 를 역수로 사용하여 데이터 비율에 따른 가중치를 부여한다. 많은 개수를 가진 데이터의 경우 낮은 가중치를 부여하고 적은 개수를 가진 데이터의 경우 높은 가중치를 부여한다. 이러한 가중치를 적용하여 제시한 손실함수로는 Class- Balanced Sigmoid Cross-Entropy Loss(CB Sigmoid CE-loss), Class- Balanced Softmax Cross-Entropy Loss(CB Softmax CE-loss), Class-Balanced Focal Loss(CB Focal loss)가 있으며 사진 분류에서의 Long-tailed 데이터 셋의 불균형을 해결하여 성능의 향상을 보였다.

3. 제안 방법

MIL[1] 방식의 VAD 의 기본 구조를 따르며 특징 추출기만 C3D 에서 최신 비디오 분류기인 I3D[4]로 바꾸어 재구성한 코드를 사용하였다. 추출된 특징은 3-layer 전결합(FC) 신경망에 들어가며 Epoch 은 10000 번, 학습률은 0.001 을 사용했다. λ_1, λ_2 는 8×10^{-5} 이 사용되었다. abnormal video 를 하나의 클래스로 보는 것이 아닌 UCF-Crime 데이터셋에 존재하는 전체 13 개의 클래스로 나누어 각각의 클래스별로 손실정보를 CB-loss 에 사용되는 가중치를 곱하여 손실함수의 출력을 조절하여 학습을 진행한다. 기존 CB Loss[2]에서 제시했던 손실함수는 CB Softmax CE-loss(1), CB Sigmoid CE-loss, CB Focal loss 가 있다. C 는 클래스의 개수를 나타내며 클래스별로 모델로부터 예측된 결과값을 $\mathbf{z} = [z_1, z_2, \ldots, z_C]^T$ 라 한다. 클래스 y 는 n_y 개의 학습 동영상 개수를 가지고 있으며 하이퍼파라미터(Hyperparameter) β 값에 따라 클래스별 가중치를 구해 손실함수 결과와 곱해진다. 하지만 제시된 세가지 손실함수는 VAD 에서 좋은 성능을 내지 못하는 손실함수이다.

$$CB_{softmax}(\mathbf{z}, y) = -\frac{1-\beta}{1-\beta^{n_y}} \sum_{i=1}^{C} \log\left(\frac{\exp(z_y)}{\sum_{j=1}^{C}\exp(z_j)}\right). \quad (1)$$

본 논문에서는 CB Loss 를 VAD 에 적용하기위해 CB-MIL Loss(2)를 제시한다. β값과 클래스별 학습영상 개수에 따라 가중치가 계산되며 ① MIL-ranking Loss 과 곱해진다.

$$CB_{MIL}(B_a, B_n, y) = -\frac{1-\beta}{1-\beta^{n_y}} \overbrace{\times \max\left(0, 1 - \max_{i \in B_a} f(V_a^i) + \max_{i \in B_n} f(V_n^i)\right)}^{①}. \quad (2)$$

기존 방식과 달리 B_a, B_n 과 해당하는 abnormal video 의 클래스가 무엇인지를 전달하는 레이블 y 도 함께 들어간다. V_a^i, V_n^i 는 이상행동과 정상행동의 스니펫을 나타내며 i 는 스니펫의 위치 정보를 나타낸다. $f(V_a^i)$, $f(V_n^i)$ 는 이상행동과 정상행동 스니펫의 이상현상 점수를 나타낸다.

최종 손실함수(3)의 형태는 기존 MIL[1] 방식에서 제시한 ② smoothness term 과 ③sparsity term 을 더하여 나타낸다. 이상행동은 전체영상에서 희소하게 나타나며 이상행동과 인접한 스니펫의 이상현상 점수의 경우는 부드럽게 상승하고 감소해야 한다는 점을 강제하기 방법이다. 이러한 손실함수의 결과 값으로 하여금 클래스별로 균형 있게 모델을 학습하도록 최적화한다.

$$l(B_a, B_n, y) = CB_{MIL}(B_a, B_n, y) + \lambda_1 \overbrace{\sum_{i}^{(n-1)} (f(V_a^i) - f(V_a^{i+1}))^2}^{②} + \lambda_2 \overbrace{\sum_{i}^{n} f(V_a^i)^2}^{③} \quad (3)$$

4. 실험 결과

UCF-crime 데이터셋을 기반으로 한 MIL 방식의 재구성된 코드를 사용하였다. 그림 2 는 이상행동 종류와 학습 테스트 데이터셋의 구성을 나타내며 데이터 불균형을 확인할 수 있다.

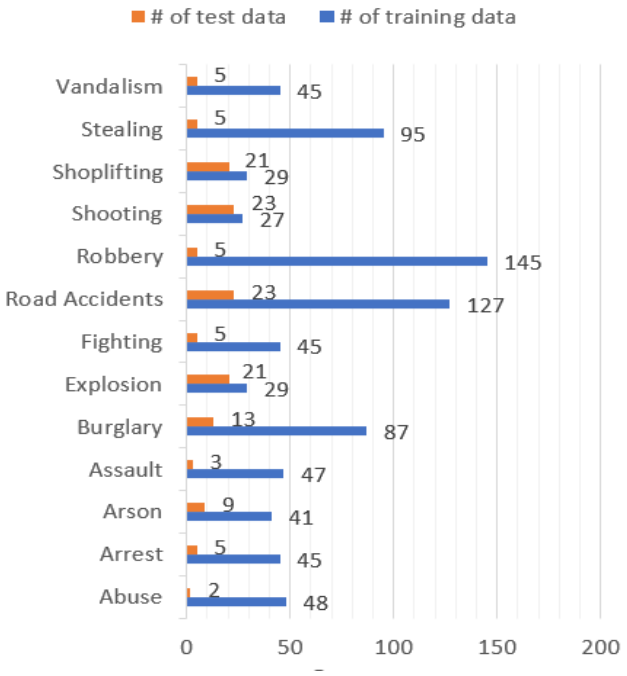

그림 2. UCF-Crime 데이터셋 분포도

기존 MIL 방식은 이상행동과 정상행동을 함께 전체적인 AUC(Total AUC)를 계산하는 것에서 그쳤지만 본 논문에서는

Anomaly Class	AUC change (β=0.999)	AUC change (β=0.9999)
Vandalism	+ 0.2%	+ 0.3%
Stealing	- 1.2%	- 1.6%
Shoplifting	0.0 %	+ 0.2%
Shooting	- 0.7%	- 0.5%
Robbery	+ 0.9%	+ 0.6%
Road Accidents	+ 0.3%	+ 1.0%
Fighting	+ 0.3%	+ 0.3%
Explosion	+ 2.6%	+ 2.7%
Burglary	- 0.3%	+ 0.6%
Assault	+ 1.3%	- 0.4%
Arson	+ 3.5%	+ 3.0%
Arrest	+ 0.5%	+ 1.4%
Abuse	- 0.9%	- 0.4%
Average AUC	+ 0.54%	+ 0.57%

표 1. 이상행동 클래스 별 성능 지표

이상행동의 종류별로 AUC 를 계산하고 각각 계산된 AUC 를 평균 내어 이상행동동영상만의 전체적인 AUC(Abnormal AUC)를 계산하였다. 표 1 은 기존 MIL 방식과 비교하여 CB-MIL 방식으로 모델 학습을 진행하였을 때 이상행동 종류별로 AUC 향상도와 전체 이상행동 평균 AUC 향상도를 나타냈다. CB loss 에서 사용되는 하이퍼파라미터 β는 가장 좋은 성능을 보이는 0.999 와 0.9999 를 사용하였다. 모든 이상행동 종류에서 성능의 향상을 보이지는 않지만 기존 MIL 방식에서 볼 수 없었던 큰 성능 향상폭이 방화(Arson), 폭발(Explosion)에서 나타났다. 성능이 오히려 낮아지는 이상행동 종류도 있었지만 전체 이상행동 평균 AUC 는 상승한 것을 확인할 수 있다.

Method	Abnormal AUC (%)	Total AUC (%)
MIL	76.22	84.95
Ours, β=0.9	76.36	85.06
Ours, β=0.99	76.59	85.25
Ours, β=0.999	76.79	85.40
Ours, β=0.9999	76.80	85.41

표 2. β값에 따른 성능향상표

표 2 에서는 MIL 방식과 비교하여 CB-MIL 방식으로 하이퍼파라미터 β를 바꿈에 따라 Abnormal AUC 와 Total AUC 의 성능을 나타낸다. 기존 MIL 방식보다 CB-MIL Loss 방식을 사용하는 것이 전체적인 AUC 의 성능향상에도 영향을 주었다는 것을 확인할 수 있다. 또한 β값이 커질수록 데이터셋 불균형에 민감하여 가중치를 더 강하게 주게 되는데 실험결과에서 β값이 커질수록 성능이 상승하는 것을 확인할 수 있다.

5. 결론

본 논문은 이상행동감지 분야에서 사용되는 UCF-Crime 데이터셋의 불균형 문제를 지적하고 이를 해결하기 위해 Class balanced loss 에 사용되는 가중치를 MIL 에 적용하여 CB-MIL Loss 를 새롭게 제시하였다. 이상행동 종류 각각에 따른 성능의 증감을 보였으며 특정 분야 영상에서 큰 성능 향상을 확인할 수 있었다. 또한 전체적인 성능 향상도 이루었다. 하지만 가중치가 큰 데이터셋 임에도 성능향상이 미미하거나 하강하는 등 모든 이상행동 종류가 본 실험에서 의도한 가중치를 따르지 않았다는 점은 한계점으로 지목할 수 있다.

감사의 글

이 논문은 2022 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2020-0-01373, 인공지능대학원지원(한양대학교))

참고 문헌

[1] Sultani, Waqas, Chen Chen, and Mubarak Shah. "Real-world anomaly detection in surveillance videos." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.

[2] Cui, Yin, et al. "Class-balanced loss based on effective number of samples." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019.

[3] Tran, Du, et al. "Learning spatiotemporal features with 3d convolutional networks." Proceedings of the IEEE international conference on computer vision. 2015.

[4] Carreira, Joao, and Andrew Zisserman. "Quo vadis, action recognition? a new model and the kinetics dataset." proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017.