향상된 비디오 이상 탐지를 위한 객체 인식 재구성 프레임워크

홍민우[01], 조영완[2], 이효정[3], 홍인표[2], 김은지[2], 김정은[2], 박상현[2†]
상명대학교 소프트웨어학과[1], 연세대학교 컴퓨터과학과[2], 연세대학교 인공지능학과[3]
hkmw7800@gmail.com, {jyy1551, hip9863, kejh66, wjddms2216, hyojoy, sanghyun}@yonsei.ac.kr

Object-Aware Reconstruction Framework for Improved Video Anomaly Detection

Minwoo Hong[01], Youngwan Jo[2], Hyojeong Lee[3], Inpyo Hong[2], Eunji Kim[2], Jeongeun Kim[2], Sanghyun Park[2†]
Dept. of Software, Sangmyung University[1],
Dept. of Computer Science, Yonsei University[2],
Dept. of Artificial Intelligence, Yonsei University[3]

요 약

재구성 기반 비디오 이상 탐지는 작은(또는 원거리) 객체의 미약한 움직임으로 인해 이상 신호가 희석되는 성능적 한계가 존재한다. 본 논문은 이러한 한계를 보완하기 위해, 바운딩 박스에서 추출한 객체 수준 정보를 FiLM(Feature-wise Linear Modulation)으로 주입하여 특징 맵을 객체 맥락에 맞게 조절하는 객체 정보 기반 재구성 성능 개선 기법을 제안한다. 제안 기법은 기존 백본(ML-MemAE-SC)에 모듈로 부착되며, Ped2와 Avenue 벤치마크에서 Baseline 대비 AUC가 +0.6%pt, +0.7%pt 향상되어, 객체 정보에 기반한 조건화가 재구성 품질 향상에 기여함을 확인하였다.

1. 서 론

비디오 이상 탐지(Video Anomaly Detection, VAD)는 영상 내에서 기대되는 행동과 일치하지 않는 사건을 식별하는 문제로[1], 실제 환경에서는 이상 사건이 드물고 형태가 다양해 개방적이면서도 도전적인 특성을 지닌다[2]. 모든 유형의 이상 패턴을 사전에 수집하기 어렵기 때문에, 정상 데이터만으로 모델을 학습하고 테스트 시 이상치로 판별되는 사건을 이상으로 간주하는 비지도 학습 접근이 널리 사용된다. 딥러닝의 발전과 함께 전통적 수작업 특징 기반 기법[3]을 넘어 다양한 신경망 기반 방법들이 제안되어 왔다[2, 4].

이들 중 재구성 기반(reconstruction-based) 방법[4]은 정상 데이터만으로 오토인코더를 학습시키고, 입력 영상의 구조·질감·동적 패턴을 재구성하도록 훈련된다. 테스트 시 이상 데이터는 학습된 정상 패턴과 다르기 때문에 더 큰 재구성 오류를 발생시키며, 이 오류를 이상 점수(anomaly score)로 사용하여 정상과 구별한다. 이러한 특성으로 재구성 기반 방법은 이상 데이터를 필요로 하지 않고 단순하며, 다양한 정상 특징을 동시에 학습할 수 있는 범용성을 가진다.

그러나 실제 환경에서는 객체가 작거나 카메라에서 먼 경우, 재구성 오차가 충분히 커지지 않아 이상 탐지가 되지 않는 한계가 있다. 특히 객체가 카메라로부터 멀어질수록 이상 점수가 작아지는 경향이 있다. 이는 절대적인 optical flow 크기가 감소하기 때문이며, 카메라로부터 먼 객체의 이상 행동이 카메라 근처의 정상 객체보다 더 작게 측정되어 이상 신호가 충분히 부각되지 않는다. 그 결과, 이러한 객체들의 이상 탐지가 실패할 수 있다.

본 논문에서는 위 한계를 완화하기 위해, 바운딩 박스에서 추출한 객체 수준 정보(면적, 종횡비, 중심 좌표)를 FiLM(Feature-wise Linear Modulation)[6] 모듈에 입력하여, 재구성 네트워크의 특징 맵을 객체 맥락에 맞게 조절하는 방법을 제안한다. 제안 모듈은 재구성 백본인 ML-MemAE-SC[5]의 bottleneck과 decoder-2에 삽입되며, 백본 구조를 수정하지 않고도 적용 가능하다. 학습은 기존 재구성 설정과 동일하게 정상 데이터만으로 수행하고, 테스트 시 재구성 오차를 이상 점수로 사용한다.

실험 결과, UCSD Ped2에 대해서 +0.6%pt로 향상되었고, Avenue에 대해서 AUC가 0.7%pt로 향상되어 객체 정보 기반 FiLM 조건화가 재구성 품질과 이상 점수 분리도 향상에 기여함을 확인하였다.

2. 본 론

본 논문에서는 그림1과 같이 바운딩 박스에서 추출한 객체 수준 정보를 FiLM 모듈로 변환해 ML-MemAE-SC의 bottleneck과 decoder-2 계층에 주입함으로써, 작거나 원거리 객체 정보가 반영되도록 특징 맵을 조절하는 방법을 제안한다. 그림 1은 바운딩 박스 특성을 FiLM generator를 통해 MLP로 임베딩하여 γ, β를 생성하고, 이를 이용해 희석되기 쉬운 구간의 재구성 품질을 보강하는 과정을 보여준다.

2.1 재구성 기반 백본(ML-MemAE-SC)

ML-MemAE-SC는 입력 optical flow x로부터 인코더 E가 잠재 표현 z를 산출하고, 메모리 모듈 M이 정상 패턴을 보강한 후, 디코더 D가 재구성 $\hat{x}$를 생성한다.

2.2 객체 정보 추출

각 프레임에서 사전 학습된 RCNN으로 얻은 객체의 바운딩 박스 $b=(x_1, y_1, x_2, y_2)$로부터 다음과 같은 객체 정보를 추출한다.

$$z_{BB} = \mathrm{MLP}(area, aspect, c_x, c_y) \quad (1)$$

바운딩 박스의 너비 $w = x_2 - x_1$, 높이 $h = y_2 - y_1$이며, 수식(1)의 면적 $area = w \times h$, 종횡비 $aspect = w/(h + 1e-6)$, 중심좌표는

* 이 논문은 2025년도 정부(과기정통부)의 재원으로 한국연구재단의 지원(No. RS-2023-00229822)과 국토교통부의 스마트시티 혁신인재육성사업으로 지원을 받아 수행된 연구임.
† 교신 저자: sanghyun@yonsei.ac.kr

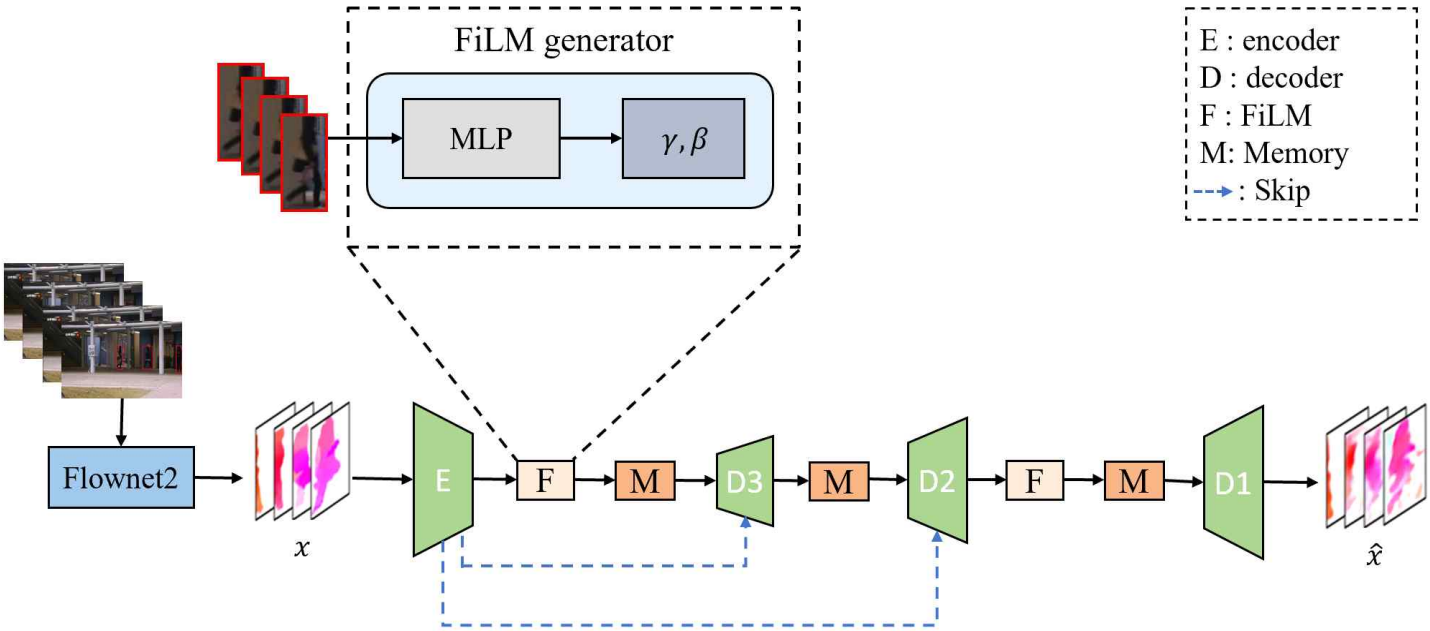

그림 1 바운딩 박스 특성으로 생성한 FiLM(γ, β)를 bottleneck, decoder-2에 주입해 특징을 조절하고, $E-M-D$로 입력x를 $\hat{x}$로 재구성하는 구조이다.

$c_x = \frac{x_1 + x_2}{2}$, $c_y = \frac{y_1 + y_2}{2}$이다.

2.3 FiLM 기반 조건화

바운딩 박스에서 추출된 객체 정보 z_{BB}를 FiLM generator g에 입력으로 주어, bottleneck과 decoder-2의 특징 맵 채널 c에 대응하는 FiLM 파라미터 γ, β를 생성한다. 구체적으로, FiLM generator g는 z_{BB}를 MLP를 통해 $2C$-차원 벡터를 생성하며 벡터의 앞 C개 요소를 스케일링 계수 γ, 뒤 C개 요소를 시프트 계수 β로 사용한다.

$$[\gamma_c, \beta_c] = g(z_{BB}), \quad c = 1, \ldots, C \tag{2}$$

수식(2)의 C는 특징 맵의 채널 수를 나타낸다. 생성된 γ, β는 bottleneck과 decoder-2의 특징 맵에 적용되어, 객체 특성에 따라 각 채널을 스케일링 및 시프트한다.

$$F'_{c,h,w} = \gamma_c \bullet F_{c,h,w} + \beta_c \tag{3}$$

수식(3)의 $F_{c,h,w}$는 bottleneck 또는 decoder-2의 특징 맵을 나타내며, $F'_{c,h,w}$는 FiLM이 적용된 후 특징 맵, h, w는 특징 맵의 높이와 너비를 나타낸다.

2.4 손실 함수 및 이상 점수

본 연구에서는 손실 함수 및 이상 점수 산출 방식은 기존 baseline과 동일한 구성으로 채택하였다. 이에 기반하여, 입력 데이터를 효과적으로 재구성하고 메모리 사용을 최소화하기 위해 재구성 손실과 메모리 희소성 손실을 결합한 총 손실 함수를 다음과 같이 정의하여 모델을 학습하였다.

$$L = \lambda_{recon} L_{recon} + \lambda_{ent} L_{ent} \tag{4}$$

수식(4)의 λ_{recon}과 λ_{ent}는 각각 재구성 손실과 메모리 희소성 손실에 대한 가중치이다. 재구성 손실 함수(L_{recon})는 입력 x와 재구성 $\hat{x}$의 차이를 L2 norm 연산을 통해 최소화하도록 다음과 같이 정의된다.

$$L_{recon} = \| x - \hat{x} \|_2 \tag{5}$$

메모리 희소성 손실 함수(L_{ent})는 각 메모리 모듈이 불필요하게 분산되지 않고 중요한 특징만 선택하도록 다음과 같이 정의된다.

$$L_{ent} = \sum_{i=1}^{M} \sum_{k=1}^{N} -\hat{\omega}_{i,k} \log(\hat{\omega}_{i,k}) \tag{6}$$

수식(6)의 M은 메모리 모듈의 개수, N은 메모리 모듈 안에 slot의 개수를 나타낸다. $\hat{\omega}_{i,k}$은 i번째 메모리 모듈에서 특징 맵이 k slot과 매칭된 확률 값이다. 이상 점수는 입력 x와 재구성 $\hat{x}$의 오차를 L2 norm 연산을 통해 구한다.

$$S = \| x - \hat{x} \|_2 \tag{7}$$

3. 실험 및 결과

3.1 실험 환경

본 연구에서는 제안하는 ML-MemAE-SC + FiLM 모델의 성능 평가를 위해 UCSD Ped2와 CUHK Avenue 데이터셋을 사용하였다. 바운딩 박스는 사전 학습된 RCNN, optical flow는 FlowNet2.0으로 추출하였으며, 연속된 5개의 프레임을 32×32 픽셀 Spatial-Temporal Cube(STC) 형태로 crop 및 resize하여 모델 입력으로 사용하였다. 학습은 batch size 128, epoch 80, 메모리 모듈 3개, 손실 가중치$\lambda_{recon}, \lambda_{ent}$는 각각 $1.0, 2e^{-4}$로 설정하였다. 각 프레임의 이상 점수는 해당 프레임 내 모든 객체 점수의 최대값으로 정의하고, 이를 기반으로 프레임 단위 ROC curve의 AUC를 평가 지표로 사용하였다.

3.2 실험 결과

표 1에서, 기존 ML-MemAE-SC (Baseline) 모델은 UCSD

Ped2에서 98.4%, CUHK Avenue에서 79.3%의 AUC를 기록하였다. 여기에 FiLM 모듈을 결합한 모델은 두 데이터셋 모두에서 성능이 향상되어, UCSD Ped2에서는 99.0%, Avenue에서는 80.0%의 AUC를 달성하였다. 이러한 결과는 FiLM 기반 조건화가 객체 특성을 효과적으로 반영하여 재구성 성능과 이상 탐지 정확도를 향상시킴을 보여준다.

표 1 Ped2, Avenue 데이터셋에 대한 AUC 점수 비교

Model	Ped2	Avenue
ML-MemAE-SC	98.4%	79.3%
ML-MemAE-SC + FiLM	99.0%	80.0%

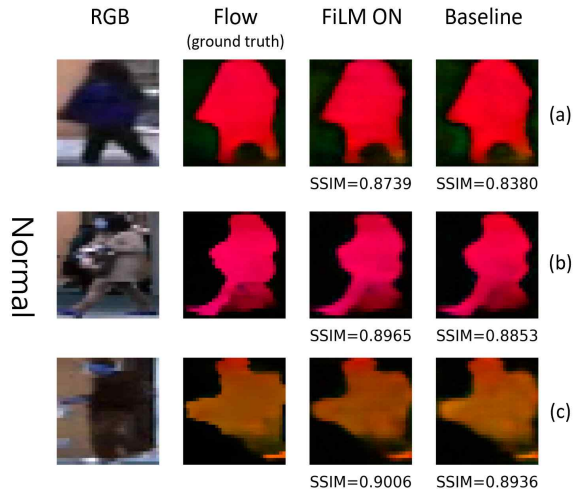

그림 2 정상 데이터 재구성 결과: (a) 작은 객체, (b),(c) 큰 객체. FiLM ON은 제안한 FiLM 기반 조건화 모델, Baseline은 기존 ML-MemAE-SC이며, flow(ground truth)와 비교함.

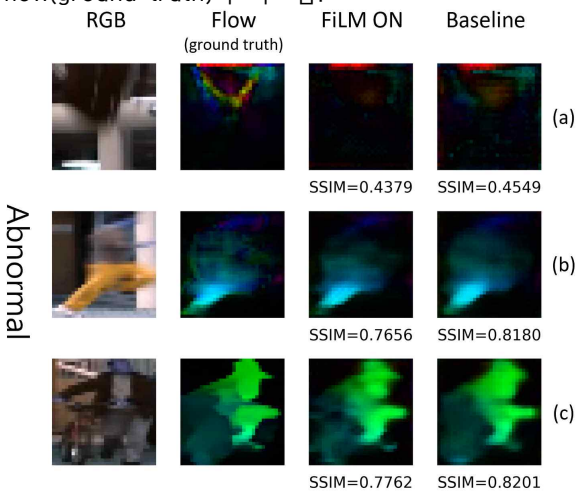

그림 3 이상 데이터 재구성 결과: (a) 작은 객체, (b),(c) 큰 객체. FiLM ON은 제안한 FiLM 기반 조건화 모델, Baseline은 기존 ML-MemAE-SC이며, flow(ground truth)와 비교함.

그림 2와 그림 3은 각각 정상 데이터와 이상 데이터에 대한 재구성 결과를 제시한다. 두 경우 모두 재구성 품질 평가는 구조적 유사도 지수(SSIM)로 수행하였다. 정상 데이터(그림 2)에서 제안한 FiLM 기반 조건화 모델은 Baseline(Ml-MemAE-SC) 대비 높은 SSIM을 보였으며, 특히 작은 객체에서 재구성 향상이 두드러졌다. 반면, 학습에 포함되지 않은 이상 데이터(그림 3)의 경우에는 모델이 입력을 충실히 재현하지 못해 낮은 SSIM을 보이는 것이 바람직한데, 실험 결과 FiLM 적용 모델이 Baseline과 비교하여 더 낮은 재구성 성능(낮은 SSIM)을 나타내어, 이상 패턴에 대한 과도한 복원을 억제한다는 점을 확인하였다.

4. 결 론

본 연구에서는 ML-MemAE-SC 모델에 FiLM 기반 조건화를 결합하여, 각 객체의 바운딩 박스 특성을 활용한 비지도 이상 탐지 방법을 제안하였다. 제안 모델은 UCSD Ped2와 Avenue에서 기존 대비 AUC 성능이 향상되었으며, 작은 객체나 미약한 움직임을 가진 객체에서도 재구성과 이상 탐지 정확도가 개선됨을 확인하였다. 특히 정상 데이터에서 FiLM 적용 모델은 Baseline 대비 높은 재구성 품질을 보였으며, 이상 데이터에서는 재구성 성능이 낮게 나타나 이상 탐지에 유리함을 확인하였다. 향후 연구에서는 다양한 객체 크기와 복잡한 배경 환경에서 성능을 검증하고, 다중 객체 상호작용을 반영한 FiLM 조건화 확장을 통해 이상 탐지의 정확성과 일반화를 향상시킬 계획이다.

참고문헌

[1] Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection: A survey. ACM Computing Surveys, 41(3):1-58, 2009.

[2] Wen Liu, Weixin Luo, Dongze Lian, and Shenghua Gao. Future frame prediction for anomaly detection-a new baseline. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 6536-6545, 2018.

[3] Amit Adam, Ehud Rivlin, Ilan Shimshoni, and Daviv Reinitz. Robust real-time unusual event detection using multiple fixed-location monitors. IEEE Transactions on Pattern Analysis and Machine Intelligence, 30(3):555-560, 2008.

[4] Mahmudul Hasan, Jonghyun Choi, Jan Neumann, Amit K Roy-Chowdhury, and Larry S Davis. Learning temporal regularity in video sequences. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 733-742, 2016.

[5] Liu, Zhian, et al. A hybrid video anomaly detection framework via memory-augmented flow reconstruction and flow-guided frame prediction. Proceedings of the IEEE/CVF international conference on computer vision, 2021.

[6] Perez, Ethan, et al. FiLM: Visual reasoning with a general conditioning layer. Proceedings of the AAAI conference on artificial intelligence, 32(1), 2018.