

다중 프레임 예측 에러를 활용한 영상 이상 탐지

김유준^{1,2}, 김영갑^{1,2,*}

¹세종대학교 정보보호학과

²세종대학교 지능형드론융합전공

awakening95@naver.com, alwaysgabi@sejong.ac.kr

Video anomaly detection using multi-frame prediction error

Yujun Kim*, Young-Gab Kim*[†]

*Dept. of Computer and Information Security, and Convergence Engineering
for Intelligent Drone, Sejong University

[†] Corresponding author: Young-Gab Kim

요 약

공공 안전을 위한 영상 감시 시스템이 증가함에 따라 CCTV 관제사가 관제해야 할 영상의 수가 증가하고 있다. 점점 증가하는 관제 영상 수로 인해 CCTV 관제사는 수많은 영상 사이에서 발생하는 살인, 강도, 폭력 등 위급한 이상 상황을 놓치는 문제가 발생할 수 있다. 이러한 문제를 해결하기 위해 최근에는 영상에서 발생하는 이상 상황을 자동으로 탐지하고 CCTV 관제사에게 알려 관제 효율을 향상시키는 연구가 진행되고 있다. 본 논문은 영상에서 발생하는 이상 상황을 자동으로 탐지하기 위해 예측 기반 이상 탐지 방법에 다중 프레임 예측 에러를 활용해서 영상 이상 탐지 정확도를 향상시키는 방법을 제안한다. 결과적으로 제안한 방법을 사용함으로써 프레임 레벨 AUC가 Ped2 데이터셋에서 92.70%에서 94.56%, Avenue 데이터셋에서 87.37%에서 89.17%로 상승하였다.

1. 서론

영상 감시 시스템 수의 증가는 CCTV 관제사마다 관제해야 할 영상의 수를 증가시켰다. 이로 인해 CCTV 관제사는 관제하고 있는 수많은 영상 중 하나의 영상에서 이상 상황(살인, 강도, 폭력 등)이 발생할 때 해당 이상 상황을 인지하지 못하는 경우가 발생할 수 있고, 이는 심각한 문제를 일으킬 수 있다. 이러한 문제를 해결하기 위해 최근에는 딥러닝 모델을 사용해 영상에서 발생하는 이상 상황을 자동으로 탐지해 CCTV 관제사에게 알려 관제 효율을 향상시키는 연구가 진행되고 있다.

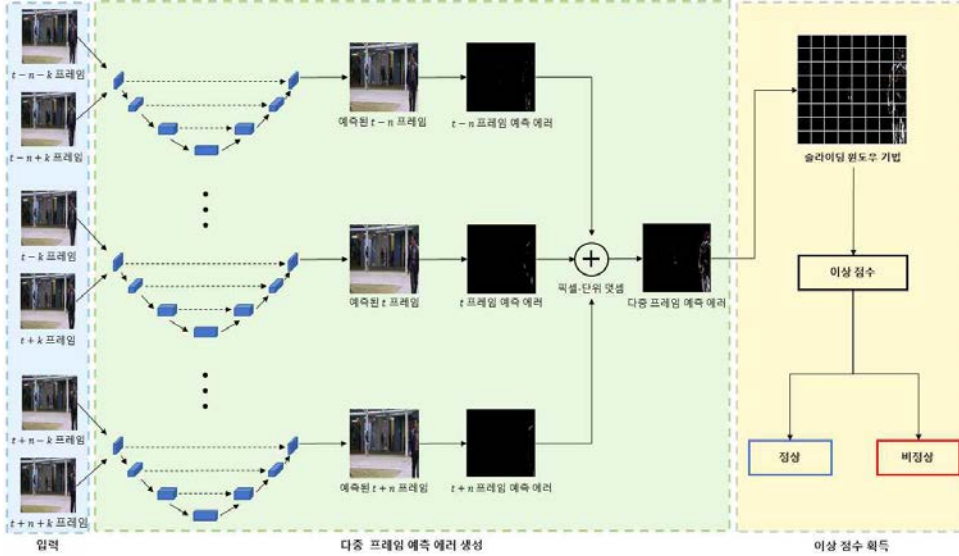
딥러닝 모델을 사용한 이상 탐지 방법의 하나인 예측 기반 이상 탐지 방법은 예측 모델을 사용한다. 예측 모델은 정상 프레임들만을 입력으로 사용해 현재 프레임을 예측하도록 훈련하고, 훈련된 예측 모델은 비정상 프레임들을 입력으로 받을 때보다 정상 프레임들을 입력으로 받을 때 현재 프레임을 더 정확하게 예측한다. 이를 토대로 예측된 현재 프레임과 실제 현재 프레임의 차이인 예측 오류를 활용해 현재 프레임의 이상 상황을 탐지한다. 다만, 딥러닝 모델의 성능, 프레임의 배경 소음(background noise), 정상 프레임인지 비정상 프레임인지 구별하

기 어려운 프레임 등으로 인해 예측 모델이 정상 프레임들을 입력으로 받을 때보다 비정상 프레임들을 입력으로 받을 때 현재 프레임을 더 정확하게 예측하는 경우가 발생할 수 있다. 이로 인해 정상 프레임은 높은 예측 에러를 가지고 비정상 프레임은 낮은 예측 에러를 가지게 되어 이상 상황을 정상 상황으로 이상 상황을 정상 상황으로 추론하게 되는 문제가 발생한다.

본 논문에서는 이러한 프레임의 잘못된 예측 에러 차이를 줄이고 영상 이상 탐지 정확도를 높이기 위해 현재 프레임의 이상 상황을 추론할 때 현재 프레임의 예측 에러뿐만 아니라 인접 프레임들의 예측 에러를 함께 사용하는 다중 프레임 예측 에러를 제안한다.

2. 제안 방법

예측 기반 이상 탐지 방법에서 예측 모델의 성능, 프레임의 배경 소음, 정상인지 비정상인지 애매 모호한 프레임 등으로 정상 프레임이 높은 예측 에러를 가지고 비정상 프레임이 낮은 예측 에러를 가지는 경우가 발생한다. 본 논문은 이러한 프레임의 잘못된 예측 에러 차이를 인접 프레임들의 예측 에



(그림 1) 다중 프레임 예측 에러를 활용한 영상 이상 탐지 방법.

리로 완화하기 위해 현재 프레임의 예측 에러와 인접 프레임들의 예측 에러로 다중 프레임 예측 에러를 생성하여 현재 프레임의 이상 상황을 탐지한다. 본 논문에서 제안한 이상 상황 탐지 방법은 그림 1과 같이 다중 프레임 예측 에러 생성 단계와 이상 점수 획득 단계로 구성되어 있다.

2.1. 다중 프레임 예측 에러 생성

현재 프레임의 다중 프레임 예측 에러를 생성하기 위해 현재 프레임의 예측 에러 $PE_{i,j}^t$ 와 인접 프레임들의 예측 에러($PE_{i,j}^{t-n}, \dots, PE_{i,j}^{t-2}, PE_{i,j}^{t-1}, PE_{i,j}^{t+1}, PE_{i,j}^{t+2}, \dots, PE_{i,j}^{t+n}$)를 함께 사용한다. 각 프레임의 $PE_{i,j}^t$ 는 식(1)과 같이 실제 프레임 F_{ij}^t 과 예측된 프레임 \hat{F}_{ij}^t 에 대한 차이의 제곱으로 획득한다.

$$PE_{i,j}^t = \sum_{i=1}^h \sum_{j=1}^w (\hat{F}_{ij}^t - F_{ij}^t)^2 \quad (1)$$

t 는 시간, h 는 프레임의 높이 w 는 프레임의 넓이를 의미하고, $PE_{i,j}^t$ 는 하나의 이미지가 된다. $PE_{i,j}^t$ 의 픽셀은 실제 프레임과 예측된 프레임의 대응하는 픽셀들의 차이이며, 픽셀이 큰 값을 가지면 이상 상황을 포함할 가능성이 크다는 것을 의미한다. 다만, 다양한 원인으로 일부 정상 픽셀들이 큰 값을 가지거나 일부 비정상 픽셀들이 낮은 값을 가질 수 있다. 따라서, 본 논문에서는 현재 프레임의 잘못된 예측 에러 차이를 줄이기 위해 인접 프레임들의 예측 에러를 함께 사용한 다중 프레임 예측 에

러 $MPE_{i,j}^t$ 를 사용한다. $MPE_{i,j}^t$ 는 식 (2)와 같이 현재 프레임의 예측 에러와 인접 프레임들의 예측 에러의 픽셀-단위 덧셈으로 생성된다.

$$MPE_{i,j}^t = \sum_{k=t-n}^{t+n} PE_{i,j}^k \quad (2)$$

2.2 이상 점수 획득

본 논문에서는 현재 프레임의 이상을 추론하기 위해 이상 점수를 사용한다. 이상 점수는 높을수록 현재 프레임이 높은 확률로 이상 상황을 포함한다는 것을 의미한다. 우리는 컨볼루션(convolution) 연산과 같이 $MPE_{i,j}^t$ 를 일정 크기의 필터(filter)가 스트라이드(stride)만큼 이동할 때 필터들과 상응하는 $MPE_{i,j}^t$ 의 패치 중 평균값이 가장 큰 패치를 이상 점수로 하는 슬라이딩 윈도우 기법[1]을 사용한다.

3. 실험 및 결과

3.1. 실험 데이터

본 논문에서 제안한 다중 프레임 예측 에러가 영상 이상 탐지 정확도에 미치는 영향을 확인하기 위해 기존 영상 이상 탐지 연구에 많이 사용되는 UCSD Ped2 데이터셋과 CUHK Avenue 데이터셋을 사용하였다. Ped2 데이터셋은 360×240의 해상도를 가진 16개의 훈련 영상과 12개의 테스트 영상을 제공한다. Avenue 데이터셋은 640×360의 해상도를 가진 16개의 훈련 영상과 21개의 테스트 영상을 제공한다.

3.2. 실험 방법

본 실험에서는 영상 이상 탐지 정확도를 측정하기 위해 프레임 레벨 AUC(area under the curve)를 측정한다. AUC는 ROC(receiver operating characteristic) curve의 아래 면적이며, 제안한 방법에서 ROC는 비정상 상황에 대한 이상 점수의 임계값을 변경함으로써 획득할 수 있다.

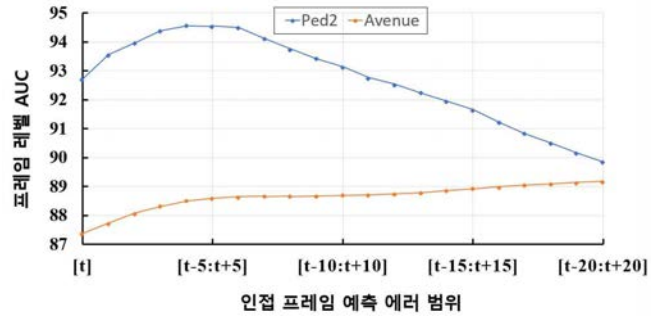
본 실험은 데이터셋마다 다중 프레임 예측 에러를 사용하지 않았을 때와 다중 프레임 예측 에러를 사용했을 때 인접 프레임 예측 에러 범위 $[t-1:t+1]$ 부터 $[t-20:t+20]$ 까지의 각 프레임 레벨 AUC를 측정했다. 예측 모델은 $t-1$ 프레임과 $t+1$ 프레임을 입력으로 받아 t 프레임을 예측하는 U-Net 모델[2]을 사용했다. 슬라이딩 윈도우 기법을 사용할 때 필터의 크기는 32×32 , 스트라이드는 32로 설정하였다.

3.3. 실험 결과

그림 2는 Ped2와 Avenue 데이터셋에서 다중 프레임 예측 에러에 활용된 인접 프레임 예측 에러 범위에 따른 프레임 레벨 AUC를 보여준다. 다중 프레임 예측 에러를 활용하지 않았을 때, 즉 인접 프레임 예측 에러 범위가 $[t]$ 일 때 Ped2에서는 92.70% Avenue에서는 87.37%의 프레임 레벨 AUC를 획득했다. Ped2는 인접 프레임 예측 에러 범위가 $[t-4:t+4]$ 일 때까지 프레임 레벨 AUC가 증가하였고 최대 프레임 레벨 AUC는 94.56%이었다. Avenue는 인접 프레임 예측 에러 범위가 $[t-20:t+20]$ 일 때까지 프레임 레벨 AUC가 계속 증가하였고 최대 프레임 레벨 AUC는 89.17%이었다. 각 데이터셋에서 다중 프레임 예측 에러를 활용하기 전에 비해 다중 프레임 예측 에러를 활용한 후 프레임 레벨 AUC는 표 1과 같이 Ped2는 최대 1.86%, Avenue는 최대 1.80% 증가하였다. 이는 다중 프레임 예측 에러가 영상 이상 탐지 정확도를 향상시켜준다는 것을 보여준다.

4. 결론

본 논문에서는 영상 이상 탐지 정확도를 향상시키는 방법으로 다중 프레임 예측 에러 활용을 제안하였다. 우리는 실험을 통해 다중 프레임 예측 에러를 활용하지 않았을 때보다 다중 프레임 예측 에러를 활용했을 때 프레임 레벨 AUC가 Ped2와 Avenue 데이터셋에서 증가하는 것을 확인할 수 있



(그림 2) Ped2와 Avenue 데이터셋에서 다중 프레임 예측 에러에 활용된 인접 프레임 예측 에러 범위에 따른 프레임 레벨 AUC

<표 1> Ped2와 Avenue 데이터셋에서 다중 프레임 예측 에러를 사용하기 전과 후의 프레임 레벨 AUC 및 증가량

데이터셋	다중 예측 에러 사용 전	다중 예측 에러 사용 후	증가량
Ped2	92.70%	94.56%	+1.86%
Avenue	87.37%	89.17%	+1.80%

었다. 이를 통해 우리가 제안한 다중 프레임 예측 에러가 영상 이상 탐지 정확도를 향상시켜준다는 것을 보여주었다. 향후 연구에서는 다중 프레임 예측 에러에 적합한 예측 모델과 영상 이상 탐지 방법을 연구해 영상 이상 탐지 정확도를 향상시킬 것이다.

사사

이 논문은 2019년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2019-0-00231, 공공 인프라 안전을 위한 인공지능 기반 영상보안 기술 및 시스템 개발)

참고문헌

[1] T.-N. Nguyen and J. Meunier, "Anomaly detection in video sequence with appearance-motion correspondence," Proceedings of the IEEE/CVF international conference on computer vision, 2019, pp. 1273-1283.

[2] O. Ronneberger, P. Fischer and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," Proceedings of International Conference on Medical image computing and computer-assisted intervention, Springer, 2015, pp. 234-241.