

심층학습 기반 멀티 모달 데이터를 활용한 차량 사고 검출

조영완[†], 조익현[†], 김대희[†], 김인정[†], 김수현[†], 강승은[†], 엄주석[†], 이재구^{†*}

국민대학교 ^{†*}{소프트웨어학부, 컴퓨터공학과} [†]쏘카

*jaekoo@kookmin.ac.kr

Deep Learning Based Car Accident Detection Using Multi-Modal Data

Youngwan Jo, Ikhyeon Jo, Daehee Kim, Inkyung Kim, Suhyun Kim, Seungeun Kang, Jusuk Um,

Jaekoo Lee*

^{†*}College of Computer Science, Kookmin University, [†]Socar

요 약

기존의 블랙박스 데이터를 활용한 차량 사고 자동 검출 시스템은 블랙박스의 메타데이터만을 활용하였다. 이 방법은 동영상 데이터를 활용하지 못하는 한계가 있으므로, 본 논문에서는 블랙박스 동영상 데이터와 메타데이터를 융합하여 사용하는 ConvLSTM 기반 모델을 설계하여 제안한다. 실험적으로 동영상 데이터와 가속도 데이터를 함께 학습한 모델은 하나의 데이터만 학습한 모델보다 검출 정확도가 7.5% 향상되어, 97.5%를 달성하였다. 제안하는 멀티 모달 데이터 기반 차량 사고 검출 모델은 실제 환경에서 블랙박스의 모든 데이터를 활용하여 더 정확한 차량 사고 검출을 목적으로 사용될 수 있을 것으로 예상된다.

I. 서 론

쏘카(Socar)는 이용자들이끼리 자율적으로 차량을 공유하여 사용하는 차량 공유 플랫폼으로, 회사 측에서는 차량에 발생한 미세한 사고가 언제 발생하였는지 검출할 수 있는 “자동 차량 사고 검출 시스템”을 필요로 한다. 우리는 이러한 시스템으로 블랙박스의 가속도 등 센서 데이터를 포함하여 메타데이터를 활용한 기계학습 기반 사고 탐지[1]를 제안한 바 있다. 하지만 이 방법은 블랙박스의 동영상을 적절히 활용하지 못한다는 한계가 있다. 이 문제를 극복하기 위해, 본 논문에서는 차량 사고 검출에 블랙박스 동영상과 메타데이터를 함께 활용하는 멀티 모달(Multi-modal) 기반 차량 사고 검출을 제안한다.

우리는 블랙박스 동영상의 순차적인 특징을 적절히 추출하기 위해 ConvLSTM[2] 모델을 사용하였다. 동영상 데이터만 사용한 모델과 동영상과 함께 메타데이터를 융합하여 입력하는 모델 두 가지에 대해 실험하였다. 그 결과로 멀티 모달 기반 모델은 7.5%만큼 정확도가 향상되는 것을 정량적으로 검증하였다. 우리는 이 실험을 바탕으로 멀티 모달 기반 차량 사고 검출 모델이 더 정확한 실제 사고 검출을 위해 활용될 수 있을 것으로 예상된다.

II. 본론

미세한 차량 사고는 시간에 따른 순차적인 특징이 있으므로 이미지 한 장으로는 사고를 파악하기 어렵다. 따라서 동영상에서 임의의 프레임들을 추출하여 순차적으로 사고 이미지 여러 장을 데이터로 사용해야 한다. ConvLSTM 모델은 이와 같은 형태의 데이터에서 시간적인 특징을 추출할 수 있는 합성곱 기반 LSTM[3] 모델이다. 따라서 본 논문에서는 순차적인 이미지 데이터의 특징을 추출하기 위해 [그림 1]과 같은 구조의 ConvLSTM 기반 검출 모델을 설계하여 사용하였다.

[그림 1]에서 L 은 순차 이미지 1 개의 총 길이를 의미한다. 예를 들어, 순차 이미지의 시간 간격을 3 초로 하고 FPS를 2로 추출한다면 L 은 6이 된다. 모델의 추론 과정으로는 3개의 ConvLSTM Block을 지나 128 채널의 특징 맵을 출력하고, 전역 평균 풀링(Global Average Pooling)을 통해 각 특징 맵의 평균값을 특징으로 사용한다. 이후 3개의 완전연결층(Fully Connected Layer)으로 구성된 분류기(Classifier)를 통해 다 대일(Many to One)의 형태로 사고에 대한 예측값을 출력한다.

제안 모델은 블랙박스 가속도 데이터를 이미지 데이터와 함께 활용하기 위해 가속도 데이터를

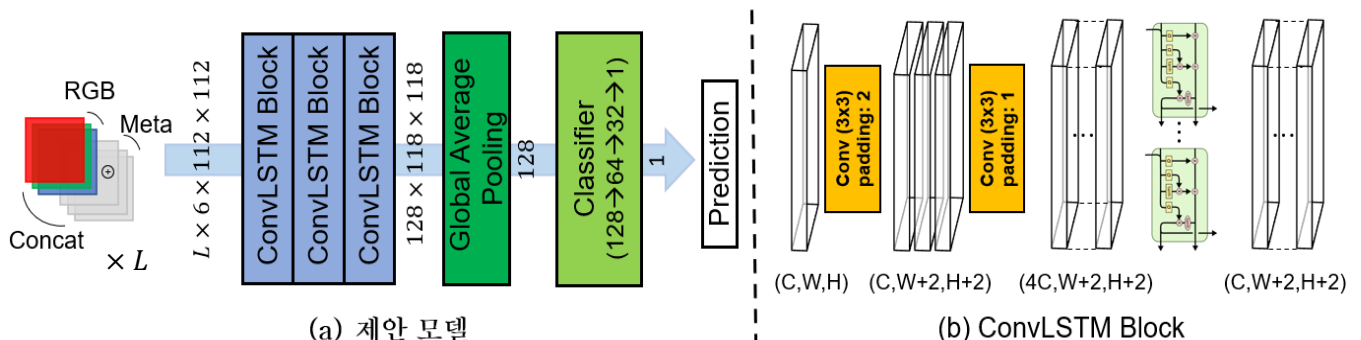


그림 1. 제안 모델 묘사

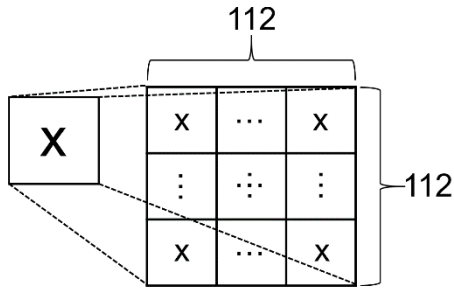


그림 2. 가속도 데이터를 이미지 크기만큼 확장



그림 3. 사고 동영상에서 추출한 이미지 묘사

[그림 2]와 같은 방법으로 확장하여 사용하였다. 구체적으로, 가속도 데이터 x, y, z 값을 이미지의 크기인 112×112 만큼 확장하고 픽셀마다 각 가속도 값을 복사하였다. 이를 [그림 1]과 같이 이미지의 RGB 3 채널과 접합(Concatenation)하여 입력으로 함께 사용하였다. 즉, 블랙박스 사고 동영상을 임의의 프레임 단위로 추출한 이미지인 RGB 채널과 메타 데이터의 가속도 x, y, z 값을 융합하여 모델에 입력하였다.

III. 실험

쏘카에서 제공한 동영상 데이터와 메타 데이터는 사고 43 개와 비사고 48 개로 구성되어 총 91 개이며, 우리는 이 중 88 개의 데이터로 모델을 학습하였다. 사고인 3 개 데이터는 테스트를 위해 사용하였다. 영상 데이터는 30FPS, 메타 데이터는 2Hz 로 구성되어 있으므로 각 데이터의 시점을 동기화하기 위해 동영상 데이터의 초당 30 프레임 중 1 번과 16 번 프레임을 추출하여 데이터 집합을 구성하였다.

하나의 순차 데이터는 3 초간의 데이터로, 6 장의 이미지와 각 이미지 시점에 해당하는 가속도 데이터 x, y, z 로 구성하였으며 보폭은 1 로 설정하였다. 이와 같은 과정을 거쳐 훈련 집합은 사고 데이터 177 개, 비사고 데이터 4,431 개로 총 4,608 개의 순차 데이터로 구성하였다. 테스트 집합은 동영상 3 개에서 추출한 사고 데이터 19 개와 비사고 데이터 296 개 중 무작위로 19 개를 추출하여 사용하여 클래스 불균형(Class Imbalance) 문제를 해소하였다. 학습 시에도 클래스 불균형 문제를 해소하기 위해 언더 샘플링(Under Sampling)을 수행하여 학습하였다. 목적 함수로는 교차 엔트로피(Cross Entropy)를 사용하였으며, 배치 크기는 4, Adam 최적화 기법[4], 학습률은 0.001 로 50 세대(Epcohs)를 학습하였다.

우리는 멀티 모달 방법의 정확도를 검증하기 위해 영상 데이터만 학습한 모델과 멀티 모달 데이터를

표 1. 데이터 형태에 따른 모델별 정확도

Threshold	2 FPS Images	2 FPS Images & Metadata
0.4	0.775	0.975
0.5	0.900	0.975
0.6	0.925	0.950
0.7	0.700	0.950

학습한 모델의 성능을 각각 실험하였다. 모델에서는 출력으로 입력이 사고인지 예측하는 1 개의 신뢰도를 예측하는데, Sigmoid 함수를 통해 신뢰도 값을 0~1 사이의 값으로 변환하여 임계값 0.4, 0.5, 0.6, 0.7 마다 정확도를 계산하였다.

각 모델에서 예측한 신뢰도 점수의 임계값별 정확도는 [표 1]과 같다. 일반적으로 사용하는 임계값 0.5 를 기준으로, 멀티 모달을 사용하여 영상과 메타 데이터를 함께 학습한 모델은 영상만을 학습한 모델보다 임계값을 정확도 7.5%가 향상되었다. 이 결과는 블랙박스의 모든 데이터를 활용하는 멀티 모달 방법이 영상 데이터만 사용하여 학습한 모델보다 더 차량 사고 검출에 우월하다는 것을 의미한다.

IV. 결론

본 논문에서는 차량 사고 검출을 위해 블랙박스의 동영상 데이터와 메타데이터를 멀티 모달로 활용하였다. ConvLSTM 기반 검출 모델을 설계하여 제안하였으며, 멀티 모달 기반 검출 모델은 97.5% 정확도를 달성하였다. 제안 모델은 다른 형태의 데이터를 함께 활용하였고, 차량 사고 검출에 실제로 더 정확하게 사용될 수 있을 것으로 예상된다.

ACKNOWLEDGMENT

본 연구는 과학 기술 정보 통신부 및 정보 통신 기획 평가원의 글로벌 핵심 인재 양성 지원 사업(2019-0-01607)의 연구 결과로 수행되었으며, 2021 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원 (No.2021-0-00994, ICT R&D 혁신바우처 (지속가능하고 견고한 자율주행 인공지능 교육/개발 통합 플랫폼))을 받아 수행된 연구임.

참 고 문 헌

- [1] Youngwan Jo, Inkyung Kim, Suhyun Kim, Seungeun Kang, Jusuk Em, Jakoo Lee, 사고 검출을 위한 기계학습 기반 블랙박스 메타데이터 이상 탐지. In ISET 2021 Proceeding p.130-134.
- [2] Shi, Xingjian, Chen, Zhourong, Wang, Hao, Yeung, Dit-Yan, Wong, Wai-kin, and Woo, Wang-chun, ConvolutionalLSTM network: A machine learning approach for precipitation nowcasting. In NIPS, 2015.
- [3] Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735-1780.
- [4] KINGMA, Diederik P.; BA, Jimmy. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014.