복수 이벤트 분석을 위한 딥러닝 기반 영상 분석 기법

박지선, 문명운, 치옥용, 한우철, 장현준, 서가가, 허언민, 조성재, 조경은
동국대학교 멀티미디어공학과
e-mail : cke@dongguk.edu (교신저자)

A Video Analysis Method based on Deep Learning for Multiple Event Analysis

Jisun Park, Mingyun Wen, Yulong Xi, Woochul Han, Hyeonjun Jang, Jiajia Xu, Yanmin He,
Seoungjae Cho, Kyungeun Cho
Department of Multimedia Engineering, Dongguk University-Seoul

요 약

최근 딥러닝을 활용한 이미지 분석 기술 향상에 힘입어 동영상 분석 연구들이 활발히 진행되고 있다. 하지만 기존 연구들의 경우 특정 영상을 입력으로 단일 이벤트로만 분류한다. 본 논문에서는 복수 이벤트를 분석할 수 있는 딥러닝 기반 영상 분석 기법을 설계하고 실험 및 분석하였다.

1. 서론

동영상에 내제한 장면을 이해하기 위해선 동영상에 포함된 다양한 객체들의 행동 분석 및 분류 작업이 필요하며, 최근 몇 년간 인공 지능 기술의 발전으로 인해, 딥러닝 기반 이미지 및 동영상 분석 연구들이 활발히 진행되고 있다.

하지만 기존 딥러닝 영상 분석 기법의 경우 특정 동영상을 하나의 단일 이벤트로만 분류하기 때문에 단일 동영상에서 복수 이벤트를 분석할 수 없다.

본 논문에서는 Faster-RCNN(Faster-Region Based Convolution Neural Network) 신경망[1]과 LRCN(Long-term Recurrent Convolution Networks) 신경망[2]을 결합하여 단일 동영상에서 복수의 이벤트 영상을 추출하여 분류할 수 있는 영상 분석 기법을 제안한다.

2. 관련연구

딥러닝을 적용한 이미지 분석 기술 향상에 힘입어 동영상을 분석하는 연구들도 활발히 진행되고 있다[3-6]. 현재 딥러닝을 기반으로 동영상을 분석하는 연구는 대부분 CNN 을 적용하여 연속된 동영상 프레임의 특징 벡터를 추출하고 추출된 특징 벡터를 시간 축을 기준으로 통합하는 방법을 사용한다.

[3]의 경우 2D-CNN 을 적용하여 각 프레임 별 이미지 특징 벡터를 추출하고 추출된 특징 벡터를 시간 순으로 LSTM 에 입력하여 하나의 이벤트 클래스로 분류한다. [4]의 경우 3D-CNN 을 적용하여 하나의 동영상에서 추출된 프레임별 이미지를 한번에 CNN 에 입력한 결과 값으로 이벤트가 분류된다. [5]의 경우 동영상에서 2D-CNN 를 통해 한 프레임의 RGB 이미지의 특징 벡터를 추출하고 3D-CNN 을 통해 연속된 프레임 이미지의 특징 벡터를 추출한 뒤 두 개의 특징 벡터를 통합하여 계산하여 하나의 이벤트로 분류한다. [6]은 연속된 프레임 이미지의 RGB 데이터와 이를 변환한 광학 흐름 데이터를 각각 3D-CNN 으로 추출하고 두 개의 특징 벡터를 통합하여 계산하여 하나의 이벤트로 분류된다.

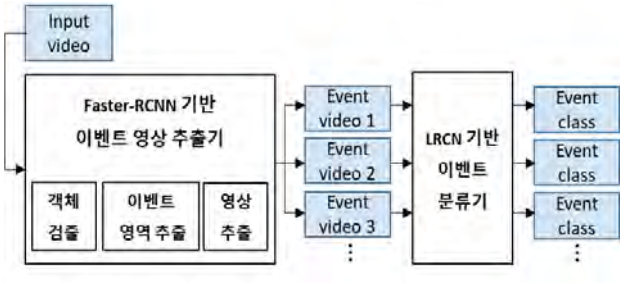
위와 같이 기존 딥러닝 기반의 동영상 분석에 대한 연구는 특정 동영상에 대한 특징 벡터를 추출한 후, 단일 이벤트로 분류하고 있기 때문에, 복수 이벤트를 추출할 수 없는 단점이 있다.

3. 딥러닝 기반 영상 분석 기법 설계

본 장에서는 단일 동영상을 입력으로 복수 이벤트를 추출하여 분류할 수 있는 딥러닝 기반의 영상 분석 기법을 설계한다. 제안된 기법은 1) 복수 이벤트 영상 추출 단계와 2) 영상 분석 기반 시나리오 생성 단계로 구성된다. 제안된 딥러닝 영상 분석 기법을 통해 복수 이벤트를 분석하는 프로세스는 [그림 1]과 같으며, 실행 순서는 다음과 같다.

첫 번째 단계로, 복수 이벤트 영상을 추출하기 위해 특정 동영상을 입력 받아 객체 검출에 높은 성능을 보인 Faster-RCNN 신경망[1]을 적용하여 동영상 내에 존재하는 객체들을 검출한다. 검출된 객체들 중에서 일정거리 이하로 인접한 객체들을 하나의 이벤트 영역으로 통합하여 복수의 이벤트 영상을 추출한다.

두 번째 단계에서는, 앞 단계에서 추출된 복수 이벤트 영상을 딥러닝 기반 동영상 분류 모델의 일종인 LRCN 신경망[2]에 입력하여 각각의 영상을 하나의 이벤트 클래스로 분류한다. 이를 통해 각 영상이 어떤 이벤트에 속하는지 알 수 있다. 최종적으로 하나의 동영상을 입력 받아 복수 이벤트 분석 결과를 추출하게 된다.



[그림 1] 제안하는 영상 분석 기법 구조도

4. 실험 및 결과

본 실험은 윈도우 기반의 파이참-community 개발 환경을 Intel i7, Nvidia GTX 980 GPU 및 DDR4 H/W 상에 구축하였으며, 딥러닝 기반 동영상 분석 모델은 대표적인 딥러닝 라이브러리인 Keras (Backend-Tensorflow) 상에서 구현하였다. <표 1>과 같이 입력 영상으로부터 Faster-RCNN 기법을 통해 검출된 객체들을 통해 이벤트 영역이 추출되고, 이를 기반으로 추출된 복수의 이벤트 영상들은 LRCN 기법을 통해 각각 분류하였다.

<표 1> 복수 이벤트 분석 결과 예시

NO.	입력 영상	복수 이벤트 분석 결과
1		E1-Car crash[Car:2] E2-Driving a car [Car:1]
2		E1-Pushing wheelchair [Human:1]
3		E1-Car crash[Car:2] E2-Driving a car [Bus:1] E3-Driving a car [Car:1]

5. 결론

본 논문에서는 단일 이벤트가 아닌 복수 이벤트를 분석할 수 있는 딥러닝 기반 영상 분석 기법을 설계하고 실험 및 분석하였다. 영상으로부터 Faster-RCNN 기법을 통해 객체를 검출하여 이를 기반으로 복수 이벤트 영상을 추출하고 이들을 LRCN 기법을 통해 각각 분류한 결과, 단일 영상을 입력으로 하여 복수 이벤트를 분석할 수 있음을 확인하였다. 향후 복수 이벤트 영역 추출 시 객체간 인접거리뿐만이 아니라 객체들의 이동 방향, 속도와 같은 다양한 요소를 고려한 이벤트 영상 추출이 가능하도록 발전시키려 한다

감사의 글

이 논문은 2018 년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(2018R1A2B2007934).

참고문헌

[1] S. Ren, K. He, et. al., “Faster R-CNN: Towards Real-time Object Detection with Region Proposal Networks,” Advances in Neural Information Processing Systems, pp. 91-99, 2015.

[2] J. Donahue, L. Anne Hendricks, et. al., “Long-term Recurrent Convolutional Networks for Visual Recognition and Description,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2625-2634, 2016.

[3] A. Karpathy, G. Toderici, et. al., “Large-scale Video Classification with Convolutional Neural Networks,” Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, pp. 1725–1732, 2014.

[4] S. Ji, W. Xu, et. al., “3d Convolutional Neural Networks for Human Action Recognition,” Proceedings of IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 35, pp. 221– 231, 2013.

[5] C. Feichtenhofer, A. Pinz, and A. Zisserman, “Convolution Two-stream Network Fusion for Video Action Recognition,” Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, pp. 1933-1941, 2016.

[6] J. Carreira and A. Zisserman, “Quo vadis, Action Recognition? A new model and the Kinetics dataset,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4724–4733, 2017.