움직임 기반 주의 정보 신경망을 이용한 행동 인식 방법

장희창, 송민수, 김원준
건국대학교
jhc6696@konkuk.ac.kr, tjqansthd@konkuk.ac.kr, wonjkim@konkuk.ac.kr

Motion-based Attention Network for Action Recognition

Heechang Jang, Minsoo Song, Wonjun Kim
Konkuk University

요 약

본 논문에서는 움직임 정보와 시공간 주의 정보를 심층신경망을 이용하여 함께 활용한 행동 인식 방법을 제안한다. RGB 영상을 입력으로 사용하는 기존 방법과 달리 제안하는 방법은 움직임 정보를 입력으로 사용하여 시간적 특징 및 시공간 주의 정보를 추출하고, RGB 영상에서 추출한 공간적 특징에 시공간 주의 정보를 고려하게 하여 행동 인식 정확도를 향상시킨다. 실험 결과를 통해 행동 분류 정확도 및 연산 효율성이 기존 신경망보다 우수함을 보인다.

1. 서론

합성곱 심층신경망(Convolutional Neural Network, CNN)을 이용한 행동 인식은 여러 가지 방법을 통해 연구되고 있으며, 크게 단일 채널(Single-channel)과 이중 채널(Dual-channel) 기반의 방법들로 나누어 볼 수 있다. 이중 채널을 사용하는 방법들은 다시 LSTM(Long short term memory)을 사용하는 신경망과 이중 흐름(Two-stream) 신경망으로 나눌 수 있고, 대표적인 이중 흐름 신경망으로 SlowFast 신경망이 있다[1].

이중 흐름 신경망은 입력 영상의 공간 정보와 시간 정보를 따로 고려하기 위해 설계된 신경망으로, 대부분 광학 흐름(Optical flow)을 입력으로 함께 사용한다. SlowFast 신경망은 광학 흐름을 사용하지 않는 이중 흐름 신경망으로, 사람의 시각 시스템을 모방하여 신경망 내 시간 간격(Time stride)과 채널 깊이(Channel depth)에 차이를 주어 시간 및 공간 정보를 포착하게 하고, 영상의 스케일이 줄어들 때 마다 시간 정보를 공간 정보와 합성한다. 사람의 시각 및 인식 체계에서 물체의 경계가 미치는 영향이 크다는 점에 착안하여, 본 연구는 움직임의 경계 부분을 활용한 효율적인 시간 정보 추출 및 시간 정보를 시공간 주의 정보로 사용하는 방법을 제안한다. 본 논문에서는 KTH[3] 데이터셋을 이용하여 제안하는 신경망을 학습하고, 기존 신경망과 행동 인식 정확도 비교를 통해 제안하는 방법이 행동 인식에 효과적임을 보인다.

2. 제안하는 방법

이 장에서는 움직임 정보 추출 및 시공간 주의 정보를 통해 이중 흐름 신경망을 학습시키는 방법에 대해 자세히 설명한다.

2-1. 움직임 경계 영상의 입력

본 논문에서 제안하는 신경망은 그림 1 과 같이 SlowFast 신경망 구조를 기반으로 한다. 제안하는 방법에서는 Fast Pathway 의 입력으로 기존 RGB 영상 대신 Persistence appearance 모듈[2]을 통해 추출한 움직임 경계 영상을 사용하여 움직임에 대한 정보를 효율적으로 추출할 수 있다.

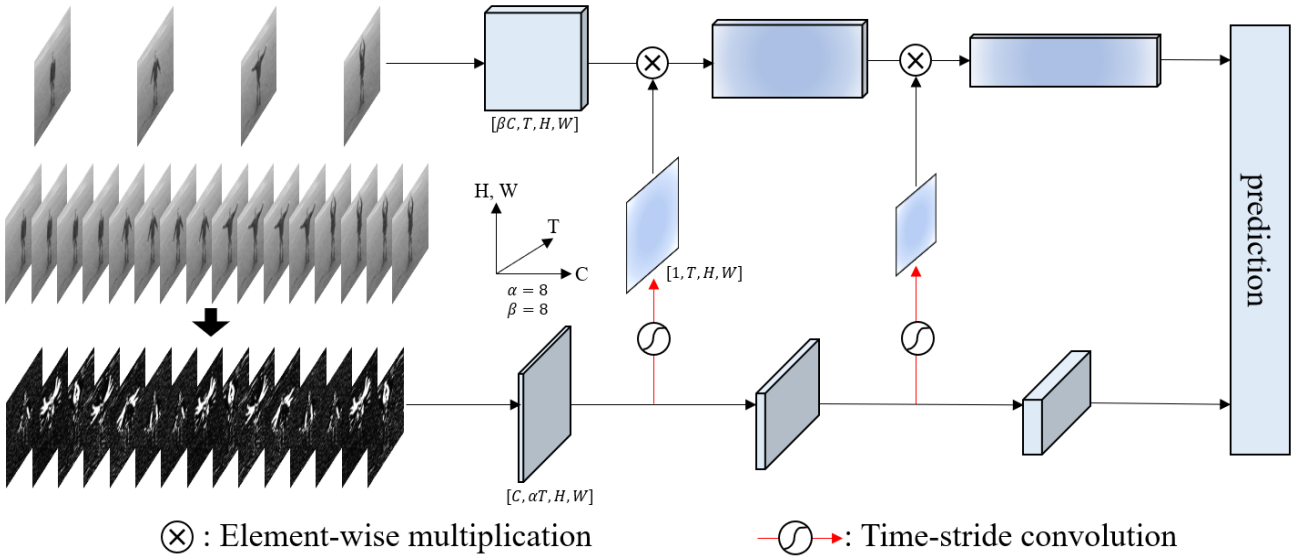

그림 1. 신경망 구조도

2-2. 시공간 주의 정보

기존 SlowFast 신경망에서 Fast Pathway 의 시간 정보를 그대로 Slow Pathway 의 공간 정보에 채널 단위로 연결하여 전달해주는 방법과 달리, 제안하는 방법에서는 특징 정보에 시간 간격이 넓은 합성곱 연산을 적용하여 단일 채널 시공간 주의 지도를 추출하고 Slow Pathway 의 특징 정보에 곱해주어 특정 시간 및 공간의 중요도가 고려될 수 있게 한다.

3. 실험 결과

제안하는 방법의 효과를 확인하기 위해 다양한 경우에 대한 성능 평가를 수행하였고, 그 결과를 표 1 에 제시하였다. 표 1 에서 PA 는 PA 모듈을 통해 추출된 움직임 경계 영상을 단독으로 Fast Pathway 입력으로 사용함을 나타내고, RGB_PA 는 RGB 영상과 움직임 경계 영상을 함께 사용한 경우를 나타낸다. attention 은 시공간 주의 지도를 Slow Pathway 의 특징 정보에 픽셀 단위 곱셈을 통해 적용함을 의미한다. 움직임 경계 영상만을 사용하고 attention 을 적용한 경우가 가장 우수한 성능을 보임을 확인할 수 있다.

4. 결론

본 논문에서는 움직임 경계 영상을 기반으로 한 시공간 주의 정보를 통해 행동 인식 정확도를 향상시키는 방법을 제안하였다. 입력 영상으로 움직임 경계 영상을 사용하고, 움직임 경계 영상 기반의 특징 정보에서 추출된 시공간 주의 정보를 RGB 영상에 곱해주는 방법을 통해 행동 인식의 정확도를 향상시킬 수 있으며, 다양한 실험을 통해 제안하는 방법이 기존 방법보다 우수한 성능을 보임을 확인할 수 있었다.

Model	Params	Accuracy (100 epochs)	Accuracy (200 epochs)
Baseline (SlowFast, R50)	34.48 M	83.84	89.90
Baseline, PA w/o attention	34.47 M	87.37	92.93
Baseline, RGB_PA w/o attention	34.48 M	90.91	94.44
Proposed Method, PA w/ attention	**32.37 M**	**92.42**	**97.47**
Proposed Method, RGB_PA w/ attention	33.07 M	87.37	95.45

표 1. 제안된 방법의 성능 분석

감사의 글

이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. 2020R1F1A1068080).

참고문헌

[1] C. Feichtenhofer, H. Fan, J. Malik and K. He. “SlowFast Networks for Video Recognition”, in *Proc. IEEE International Conference on Computer Vision (ICCV)*, pp. 6201-6210, Oct 2019.

[2] C. Zhang, Y. Zou, G. Chen, L. Gan. “PAN: Towards Fast Action Recognition via Learning Persistence of Appearance”, *arXiv:2008.03462*, Aug 2020.

[3] C. Schuldt, I. Laptev and B. Caputo, “Recognizing human actions: a local SVM approach”, in *Proc. IEEE International Conference on Pattern Recognition*, pp. 32-36, Aug 2004.