ISSN : 1226-7244 (Print)
ISSN : 2288-243X (Online)
논문번호 24-02-11
j.inst.Korean.electr.electron.eng.Vol.28,No.2,199~203,June 2024
https://doi.org/10.7471/ikeee.2024.28.2.199
77

단일 프레임 지도 시간적 행동 지역화에서 1D 합성곱 층의 커널 사이즈 변화 연구

A Study on Kernel Size Variations in 1D Convolutional Layer for Single-Frame supervised Temporal Action Localization

조 혜 정, 권 희 원, 조 선 희, 정 찬 호★

Hyejeong Jo, Huiwon Gwon, Sunhee Jo, Chanho Jung★

Abstract

In this paper, we propose variations in the kernel size of 1D convolutional layers for single-frame supervised temporal action localization. Building upon the existing method, which utilizes two 1D convolutional layers with kernel sizes of 3 and 1, we introduce an approach that adjusts the kernel sizes of each 1D convolutional layer. To validate the efficiency of our proposed approach, we conducted comparative experiments using the THUMOS'14 dataset. Additionally, we use overall video classification accuracy, mAP (mean Average Precision), and Average mAP as performance metrics for evaluation. According to the experimental results, our proposed approach demonstrates higher accuracy in terms of mAP and Average mAP compared to the existing method. The method with variations in kernel size of 7 and 1 further demonstrates an 8.0% improvement in overall video classification accuracy.

요 약

본 논문에서는 단일 프레임 지도 시간적 행동 지역화에서 1D 합성곱 층의 커널 사이즈 변화를 제안한다. 본 논문에서는 두 개의 1D 합성곱 층의 커널 사이즈를 각각 3과 1을 사용하는 기존 방법을 기반으로, 각각의 1D 합성곱 층의 커널 사이즈를 변화시키는 방법을 제안하였다. 제안하는 방법의 효율성을 검증하기 위하여 THUMOS'14 데이터셋을 활용하여 비교실험을 수행하였다. 또한 성능 평가를 위해 전체 비디오에 대한 분류 정확도(Accuracy), mAP(mean Average Precision) 그리고 Average mAP를 성능 지표로 사용하였다. 본 논문의 실험 결과에 따르면 제안하는 방법이 기존 방법보다 더 정확한 mAP와 Average mAP를 제공할 수 있음을 관찰하였다. 또한 커널 사이즈를 7과 1로 변화시킨 방법이 전체 비디오에 대한 분류 정확도에서 8.0% 개선된 것을 확인할 수 있었다.

Key words : 1D convolutional Layer, Kernel size, Temporal Action Localization

Dept. of Electrical Engineering, Hanbat National University
★ Corresponding author
E-mail : peterjung@hanbat.ac.kr
※ Acknowledgment
Manuscript received May. 10, 2024; revised Jun. 7, 2024; accepted Jun. 26, 2024.

Ⅰ. 서론

단일 프레임 지도 시간적 행동 지역화는 각 행동 인스턴스에 대해 단일 지점 레이블만 사용하여 행동 인스턴스의 시작 지점과 끝 지점을 지역화하고 해당 행동 클래스를 인식하는 것을 목표로 한다. 본 논문에서는 단일 프레임 지도 시간적 행동 지역화에서 1D 합성곱 층의 커널 사이즈를 변화시키는 방법을 제안한다. 우리는 제안하는 방법의 효율성을 증명하기 위해 LACP[2]를 기존 방법으로 사용하였다. 이때, 기존 방법[2]은 두 개의 1D 합성곱 층의 커널 사이즈를 각각 3과 1로 설정한 방법이다. 제안하는 방법에서는 두 개의 1D 합성곱 층의 커널 사이즈를 변화시키는 방법을 적용하였다. 제안하는 방법의 성능을 검증하기 위해 THUMOS'14 데이터셋[1]에서 실험을 수행하였다. 또한, 제안하는 방법의 우수성을 보여주기 위해 전체 비디오에 대한 분류 정확도, mAP, Average mAP를 성능 지표로 사용하였다. 실험 결과는 제안하는 방법이 기존 방법[2]에 비해 전체 비디오에 대한 분류 정확도에서 상대적으로 높은 성능을 제공함을 보여주었다. 더불어 두 개의 1D 합성곱 층의 커널 사이즈를 3과 5 및 5와 5로 각각 변화시킨 방법이 기존 방법[2]에 비해 mAP 및 Average mAP에서 우수함을 확인할 수 있었다. 이때, 1D 합성곱 층의 커널 사이즈를 7과 1로 변화시킨 방법이 기존 방법[2]에 비해 연산량은 2배 증가하였지만, 중요한 성능 향상을 이루었으며, 이는 의미 있는 결과를 제공한다. 또한, 제안하는 방법이 단일 프레임 지도 시간적 행동 지역화에 대한 최신 최첨단 방법을 현저하게 능가한다는 것을 보여주었다.

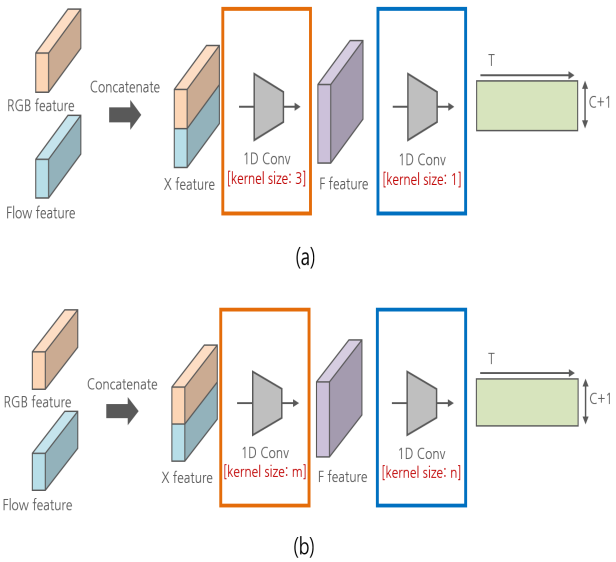

Fig. 1. (a) The structure of the existing method, (b) The structure of the proposed method.
그림 1. (a) 기존 방법 구조, (b) 제안하는 방법 구조

Ⅱ. 제안하는 방법

본 논문에서는 1D 합성곱 층의 커널 사이즈를 다양하게 변화시키는 방법을 제안한다. 그림 1-(a)에서 보이는 바와 같이 기존 방법[2]은 두 개의 1D 합성곱 층의 커널 사이즈를 각각 3과 1로 설정한 구조이다. 그림 1-(b)에서는 우리가 제안하는 방법으로, 두 개의 1D 합성곱 층의 커널 사이즈를 각각 m과 n으로 변화시키는 구조이다. 이때, 커널 사이즈 m 및 n은 각각 1, 3, 5, 7을 나타낸다.

Table 1. Comparison of accuracy results.
표 1. 정확도 결과 비교

Method	Kernel Sizes (m - n)	Accuracy (%)
existing method [1]	3 - 1	74.5
Proposed method	1 - 1	69.5
	1 - 3	73.3
	1 - 5	75.9
	1 - 7	77.1
	3 - 3	75.6
	3 - 5	76.6
	3 - 7	77.2
	5 - 1	81.4
	5 - 3	76.8
	5 - 5	77.8
	5 - 7	78.3
	7 - 1	82.5
	7 - 3	81.0
	7 - 5	79.1
	7 - 7	80.9

Ⅲ. 실험 결과

데이터셋. 제안하는 방법의 학습 및 평가를 위해 THUMOS'14 데이터셋[1]을 사용하였다. 이때, 데이터셋은 200개의 검증 데이터와 213개의 평가 데이터로 이루어진 무편집 비디오로, 20개의 운동 행동 클래스를 포함하고 있다. 기존 방법과 같게 성능을 평가하기 위하여 검

Table 2. Comparison of mAPs and average mAPs results.
표 2. mAPs 및 average mAPs 결과 비교

Method	Kernel Sizes (m - n)	mAP@IoU(%)							AVG (0.1:0.5)	AVG (0.3:0.7)
		0.1	0.2	0.3	0.4	0.5	0.6	0.7		
existing method [2]	3 - 1	75.8	71.1	64.5	56.1	44.8	33.9	20.4	62.4	43.9
Proposed method	1 - 1	70.1	65.9	58.3	50.0	38.5	27.4	15.7	56.6	38.0
	1 - 3	74.7	70.2	63.3	54.5	43.8	32.5	19.7	61.3	42.7
	1 - 5	75.9	71.5	65.0	56.9	46.9	34.3	20.9	63.2	44.8
	1 - 7	76.0	72.0	65.0	57.5	47.6	35.2	20.3	63.6	45.1
	3 - 3	77.1	72.2	65.4	56.8	46.4	35.3	20.9	63.6	45.0
	3 - 5	77.4	73.2	**66.8**	**58.5**	**48.2**	35.7	20.9	64.8	46.0
	3 - 7	76.8	72.5	65.9	57.8	47.0	35.1	20.0	64.0	45.2
	5 - 1	76.9	72.4	65.6	56.7	45.4	33.6	19.4	63.4	44.1
	5 - 3	77.1	72.8	66.4	58.1	47.2	35.7	**21.4**	64.3	45.8
	5 - 5	**78.0**	**73.8**	66.7	58.3	48.1	**36.2**	21.1	**65.0**	**46.1**
	5 - 7	77.1	73.1	65.8	57.9	47.9	35.8	**21.4**	64.4	45.8
	7 - 1	75.7	71.3	63.9	55.3	43.8	31.9	17.2	62.0	42.4
	7 - 3	76.9	72.3	65.0	56.3	45.4	33.4	19.4	63.2	43.9
	7 - 5	77.8	72.8	65.4	57.0	46.6	34.7	20.8	63.9	44.9
	7 - 7	77.8	73.6	66.0	57.4	46.9	35.0	20.6	64.3	45.2

증 데이터를 학습 데이터로 사용하여 학습을 진행하였다.

성능 지표. 본 논문에서는 제안하는 방법의 성능을 증명하기 위해 전체 비디오에 대한 분류 정확도, mAP 그리고 Average mAP를 성능 지표로 사용하였다. 전체 비디오에 대한 분류 정확도는 $\frac{\text{정답인 비디오 개수}}{\text{전체 비디오 개수}}$로 연산을 수행하였다. 또한, mAP는 IoU (Intersection over Union) 임계값(threshold) 0.1:0.7에서 0.1 간격으로 연산을 진행하였다. Average mAP는 IoU 임계값 0.1:0.5 및 0.3:0.7에서 0.1 간격으로 계산하였다.

정량적 결과. 표 1은 기존 방법[2] 및 제안하는 방법 간의 전체 비디오에 대한 분류 정확도의 성능 비교 결과를 보여준다. 표 1에서 보는 바와 같이 제안하는 방법이 기존 방법[2]에 비해 전반적으로 효율적인 성능을 보여줌을 확인할 수 있다. 또한, 두 개의 1D 합성곱 층의 커널 사이즈를 각각 7과 1로 변화시킨 방법이 기존 방법[2]과 비교하였을 때 8.0% 상승함을 확인할 수 있었다. 표 2는 기존 방법[2] 및 제안하는 방법 간의 mAP 및 Average mAP의 성능 비교 결과를 보여준다. 표 2에서 보는 바와 같이 제안하는 방법이 기존 방법[2]에 비해 전반적으로 높은 성능을 보여줌을 알 수 있었다. 더불어 두 개의 1D 합성곱 층의 커널 사이즈를 3과 5 및 5와 5로 변화시킨 방법이 기존 방법[2]에 비해 우수한 성능을 제공함을 확인할 수 있었다. 표 1과 표 2에서 제안하는 방법이 기존 방법[2]에 비해 전체 비디오에 대한 분류 정확도, mAP 그리고 Average mAP에서 전반적으로 높은 성능을 보여줌을 알 수 있었다. 더불어, 기존 방법의 연산량은 36n이며, 두 개의 1D 합성곱 층의 커널 사이즈를 7과 1로 변화시킨 방법의 연산량은 72n으로 연산량은 2배이다. 이때, n은 시퀀스의 길이이다. 본 논문에서는 1D 합성곱 층의 커널 크기를 기존 방법에서 7과 1로 변경함으로써 전체 비디오에 대한 분류 정확도를 8.0% 향상시킨 방법을 제안하였다. 이때, 연산량이 2배로 증가하지만, 중요한 성능 향상을 달성할 수 있었으며, 이는 유의미한 이점을 제공한다.

정성적 결과. 그림 2, 그림 3 그리고 그림 4는 각기 다른 비디오에 대한 두 개의 1D 합성곱 층의 커널 사이즈를 각각 1과 1, 5와 5 그리고 7과 7로 변화시킨 방법의 검출 결과와 GT를 보여준다. 순서대로 제1행은 두 개의 1D 합성곱 층의 커널 사이즈를 각각 1과 1로 변화시킨 방법의 검출 결과, 제2행은 두 개의 1D 합성곱 층의 커

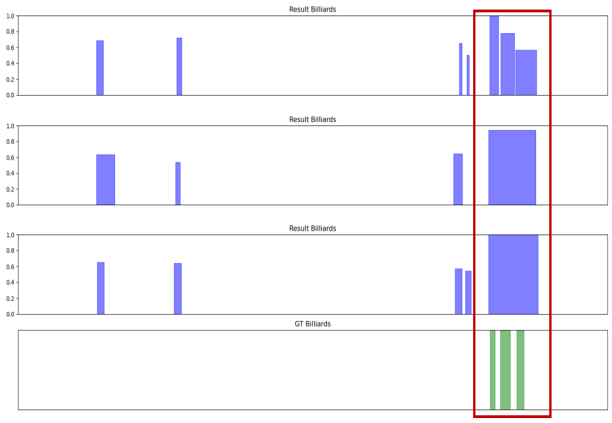

Fig. 2. Detection results for the proposed method and GT for a video_test_0000412.
그림 2. video_test_0000412에 대한 제안하는 방법의 검출 결과 및 GT

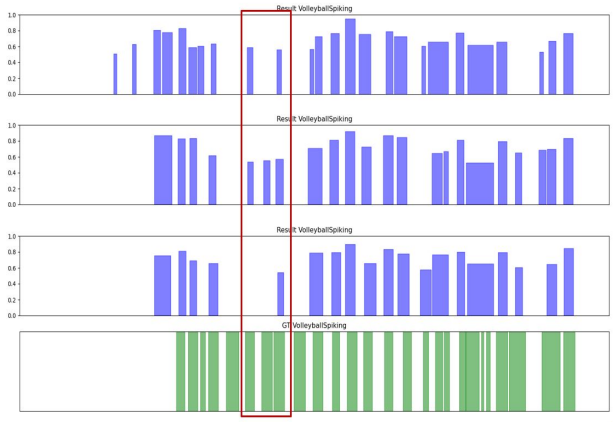

Fig. 3. Detection results for the proposed method and GT for a video_test_0000429.
그림 3. video_test_0000429에 대한 제안하는 방법의 검출 결과 및 GT

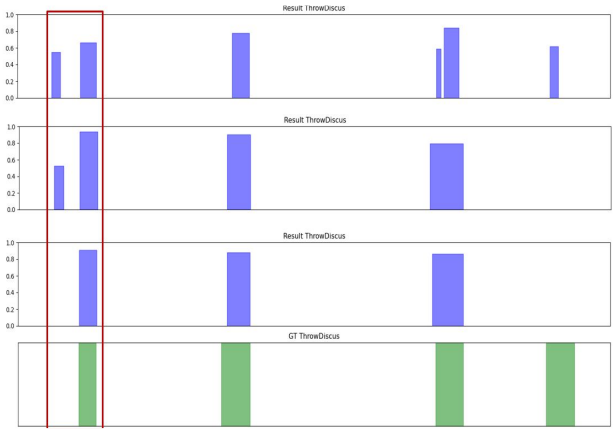

Fig. 4. Detection results for the proposed method and GT for a video_test_0000319.
그림 4. video_test_0000319에 대한 제안하는 방법의 검출 결과 및 GT

널 사이즈를 각각 5와 5로 변화시킨 방법의 검출 결과, 제3행은 두 개의 1D 합성곱 층의 커널 사이즈를 각각 7과 7로 변화시킨 방법의 검출 결과이며, 마지막으로 제4행은 해당 비디오에 대한 GT를 나타낸다. 그림 2에서 보는 바와 같이 두 개의 1D 합성곱 층의 커널 사이즈를 각각 1과 1로 변화시킨 방법의 검출 결과가 가장 좋은 것을 확인할 수 있었다. 그림 3에서 보는 바와 같이 두 개의 1D 합성곱 층의 커널 사이즈를 각각 5와 5로 변화시킨 방법의 검출 결과가 가장 우수한 것을 확인할 수 있었다. 그림 4에서 보는 바와 같이 두 개의 1D 합성곱 층의 커널 사이즈를 각각 7과 7로 변화시킨 방법의 검출 결과가 가장 좋은 것을 확인할 수 있었다.

Ⅳ. 결론

본 논문에서는 단일 프레임 지도 시간적 행동 지역화에서 1D 합성곱 층의 커널 사이즈 변화를 제안하였다. 제안하는 방법의 효율성을 검증하기 위하여 THUMOS'14 데이터셋[1]을 활용하였다. 효율성을 검증하기 위하여 전체 비디오에 대한 분류 정확도, mAP 그리고 Average mAP를 평가 지표로 사용하였다. 본 논문의 실험 결과에 따르면 제안하는 방법이 기존 방법[2]에 비해 전반적으로 우수한 성능을 제공함을 알 수 있었다. 더불어 두 개의 1D 합성곱 층의 커널 사이즈를 3과 5 및 5와 5로 변화시킨 제안하는 방법이 기존 방법[2]에 비해 전체 비디오에 대한 분류 정확도, mAP 그리고 Average mAP에서 좋은 성능을 제공함을 알 수 있었다. 실험 결과는 1D 합성곱 층의 커널 사이즈를 변화시키는 제안하는 방법이 단일 프레임 지도 시간적 행동 지역화에 대한 최신 최첨단 방법을 현저하게 능가한다는 것을 보여주었다.

References

[1] Y. G. Jiang, et al., "Thumos challenge: Actionrecognition with a large number of classes," 2014, http://crcv.ucf.edu/THUMOS14/.
[2] LEE, Pilhyeon; BYUN, Hyeran. "Learning action completeness from points for weakly-supervised temporal action localization," *In: Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021. pp.13648-13657.
DOI: 10.1109/ICCV48922.2021.01339

BIOGRAPHY

Hye-Jeong Jo (Member)

2020~current : BS degree course in Electrical Engineering, Hanbat National University.

Hui-Won Gwon (Member)

2019~current : BS degree course in Electrical Engineering, Hanbat National University.

Sun-Hee Jo (Member)

2021~current : BS degree course in Electrical Engineering, Hanbat National University.

Chan-Ho Jung (Member)

2004 : BS degree in Electronic Engineering, Sogang University.
2006 : MS degree in Electronic Engineering, Sogang University.
2013 : PhD degree in Electrical & Electronic Engineering, Korea advanced institute of science and technology.
2006~2008 : Researcher, LG Electronics.
2013 : Research Engineer, Agency for Defense Development.
2013~2016 : Research Engineer, Electronics and Telecommunications Research Institute.
2016~2020 : Assistant Professor, Dept. of Electrical Engineering, Hanbat National University.
2020~current : Associate Professor, Dept. of Electrical Engineering, Hanbat National University.