

# 단일 프레임 지도 시간적 행동 지역화에서 비디오 점수 측정 방법에 관한 연구

조혜정, 권희원, 정찬호

국립한밭대학교

20201566@edu.hanbat.ac.kr, 20191515@edu.hanbat.ac.kr, peterjung@hanbat.ac.kr

## A Study on Video Score Measurement in Single-Frame Supervised Temporal Action Localization

Hyejeong Cho, Huiwon Gwon, Chanhoo Jung

Hanbat National University

### 요약

본 논문에서는 단일 프레임 지도 시간적 행동 지역화에서 비디오 점수 측정 방법에 관한 연구를 진행한다. 본 논문에서는 단일 프레임 지도 행동 지역화에서 비디오 점수 측정 방법에서 앙상블 기법의 효율성을 입증하고, 이를 기존 방법과 비교하여 어떤 결과를 도출할 수 있는지 탐구하였다. 효율성을 입증하기 위해서 THUMOS'14 데이터셋[1]을 사용하였다. 성능을 평가하기 위해 전체 비디오에 대한 분류 정확도(Accuracy) 및 mAPs(Mean Average Precisions)을 성능 지표로 이용하였다. 실험 결과는 전반적으로 앙상블 기법을 도입한 방법 및 기존 방법의 mAPs가 비슷함을 보여주었다. 더불어 앙상블 기법을 도입한 방법이 기존 방법에 비해 전체 비디오에 대한 분류 정확도에서 비교적 우수한 성능을 제공함을 볼 수 있었다.

### I. 서론

단일 프레임 지도 시간적 행동 지역화에서 각 행동 인스턴스의 시작 시점과 끝 시점을 지역화하고 해당 행동 인스턴스의 레이블을 분류하는 것을 목표로 한다. 앙상블 기법은 다양한 모델이나 관점을 결합함으로써 전체적으로 더 좋은 성능을 나타내는 기법이다. 본 논문에서는 단일 프레임 지도 행동 지역화의 비디오 점수 측정 방법에서 앙상블 기법의 성능을 검증하고, 이를 기존 방법과 비교하여 어떤 결과를 도출할 수 있는지 탐구하였다. 본 논문에서는 앙상블 기법의 성능을 평가하기 위해 THUMOS'14 데이터셋[1]을 사용하였다. 앙상블 기법의 효율성을 입증하기 위해 전체 비디오에 대한 분류 정확도 및 mAPs를 성능 지표로 이용하였다. 성능 지표를 활용하여 수행한 실험 결과로 보아 대체로 앙상블 기법을 도입한 방법 및 기존 방법 간의 mAPs가 유사함을 보여주었다. 더불어 앙상블 기법을 도입한 방법은 기존 방법에 비해 전체 비디오에 대한 분류 정확도에서 상당한 향상을 보여줄 수 있었다. 기존 방법과의 성능 비교를 통해 단일 프레임 지도 시간적 행동 지역화의 비디오 점수 측정 방법에서 앙상블 기법의 효율성을 입증하였다. 이를 통해 앙상블 기법을 도입한 방법의 효율성을 뚜렷하게 확인하였다.

### II. 본론

본 논문에서는 앙상블 기법을 활용한 단일 프레임 지도 시간적 행동 지역화의 비디오 측정 방법에 대한 효율성을 증명하였다. 실험 결과를 통해 앙상블 기법을 도입한 방법이 기존 방법보다 효율성이 있음을 보여주었다. 본 논문에서는 앙상블 기법의 효율성을 검증하기 위하여 LACP[2]를 기존 방법으로 이용하여 학습을 진행하였다. 앙상블 기법에 사용된 비디오 점수 측정 방법은 두 가지로 1) 각 레이블에 대한 비디오 점수를 오름차순으로 정렬한다. 그 후 상위 K개의 점수를 추출하여 각 레이블에 대해

추출한 점수들의 평균을 계산하는 방식인 기존 방법, 2) 각 레이블에 대해 비디오 점수의 평균을 연산한다. 그 후 해당 레이블에 대해 평균보다 큰 비디오 점수를 추출하고 각 레이블에 대해서 추출한 점수로 새로운 평균을 도출하는 방법이다. 앙상블 기법을 활용하여 연산을 수행하는 식은 다음과 같이 정의된다.

$$\hat{P}_{ens}^{vid} = \frac{\hat{p}^{vid} + \hat{p}_m^{vid}}{2} \quad (1)$$

식 (1)에서  $\hat{p}^{vid}$  는 1) 방식을  $\hat{p}_m^{vid}$  는 2) 방식을 나타낸다. 이때,  $m$ 은 각 레이블에 대한 비디오 점수의 평균을 나타낸다. 성능 지표는 전체 비디오에 대한 분류 정확도 및 mAPs를 사용하였다. 전체 비디오에 대한 분류 정확도는  $\frac{\text{정답인 비디오 개수}}{\text{전체 비디오 개수}}$  로 연산을 진행하며, mAPs는 0.1:0.1:0.7의 IoU(Intersection over Union)을 적용하여 연산을 진행한다.

### III. 실험 결과

본 논문에서는 앙상블 기법의 효율성을 증명하기 위해서 210개의 무편집 비디오(Untrimmed video)로 구성된 THUMOS'14 데이터셋[1]을 이용하였다. 또한 전체 비디오에 대한 분류 정확도 및 mAPs를 평균 지표로 사용하였다. 표 1은 기존 방법 및 앙상블 기법을 도입한 방법 간에 전체

표 1 기존 방법 및 앙상블 기법을 도입한 방법 간에 전체 비디오에 대한 분류 정확도의 성능 비교 결과

방법	전체 비디오에 대한 분류 정확도
기존 방법[2]	74.5
앙상블 기법을 도입한 방법	83.0

표 2 기존 방법 및 앙상블 기법을 도입한 방법 간에 mAPs의 성능 비교 결과

방법	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		
기존 방법[2]	75.8	71.1	64.5	56.1	44.8	<b>33.9</b>	<b>20.4</b>	62.4	43.9
앙상블 기법을 도입한 방법	<b>76.1</b>	<b>71.4</b>	<b>64.8</b>	<b>56.5</b>	44.8	33.8	20.3	<b>62.7</b>	<b>44.1</b>

비디오에 대한 분류 정확도의 성능 비교 결과를 보여준다. 표 1에서 보는 바와 같이 앙상블 기법을 도입한 방법은 기존 방법에 비해 전체 비디오에 대한 분류 정확도에서 비교적 우수한 성능을 제공함을 볼 수 있다. 표 2는 기존 방법 및 앙상블 기법을 도입한 방법 간에 mAPs의 성능 비교 결과를 나타낸다. 표 2에서 보는 바와 같이 앙상블 기법을 도입한 방법 및 기존 방법 간의 mAPs가 유사함을 보여주었다.

#### IV. 결론

본 논문에서는 단일 프레임 지도 시간적 행동 지역화의 비디오 점수 측정 방법에서 앙상블 기법의 성능을 증명하였다. 실험 결과를 통해 전체 비디오에 대한 분류 정확도에서 앙상블 기법을 도입한 방법이 기존 방법에 비해 비교적 우수함을 알 수 있었다. 즉, 앙상블 기법이 단일 프레임 지도 시간적 행동 지역화의 비디오 점수 측정 방법에서 앙상블 기법의 효율성을 입증하였다.

#### 참 고 문 헌

- [1] Y.-G. Jiang, J. Liu, A. R. Zamir, G. Toderici, I. Laptev, M. Shah, and R. Sukthankar. "Thumos challenge: Action recognition with a large number of classes," 2016.
- [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. p. 13648-13657.