Background Suppression Network for Weakly-supervised Temporal Action Localization

AAAI 2220

01. Introduction

(1) Temporal Action Localization?

untrimmed video에서 action을 추출하는 것 중요해짐.

TAL : Temporal Action Localization

- untrimmed video에서 action을 포함하는 frame을 찾는 것을 의미
- **supervised learning**을 이용해 deep network 학습
- 개별 frame은 action/background로 라벨 지정
- **단점**: 비싸다. localization에 subjective, error-prone

01. Introduction

(2) Weakly Supervised Temporal Action Localization?

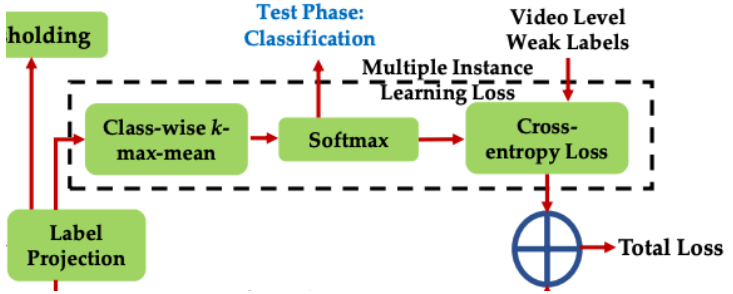
WTAL : Weakly supervised temporal action localization

- frame 단위의 라벨을 predict하지만, weak supervision

  ex) 비디오 레벨의 라벨, 비디오 action의 빈도, action의 시간적 순서 사용

- video 레벨의 라벨은 가장 일반적으로 사용됨

- 비디오는 여러 액션 클래스를 가질 수도 있음



- **MIL** : WTALC을 label을 bag으로 사용 (개별 instance가 아니라)

  - 개별 frame을 action class로 분류

  - frame 레벨의 점수로 집계하여 비디오의 class 예측

  → video level의 분류 loss는 frame level의 predict를 할 수 있음

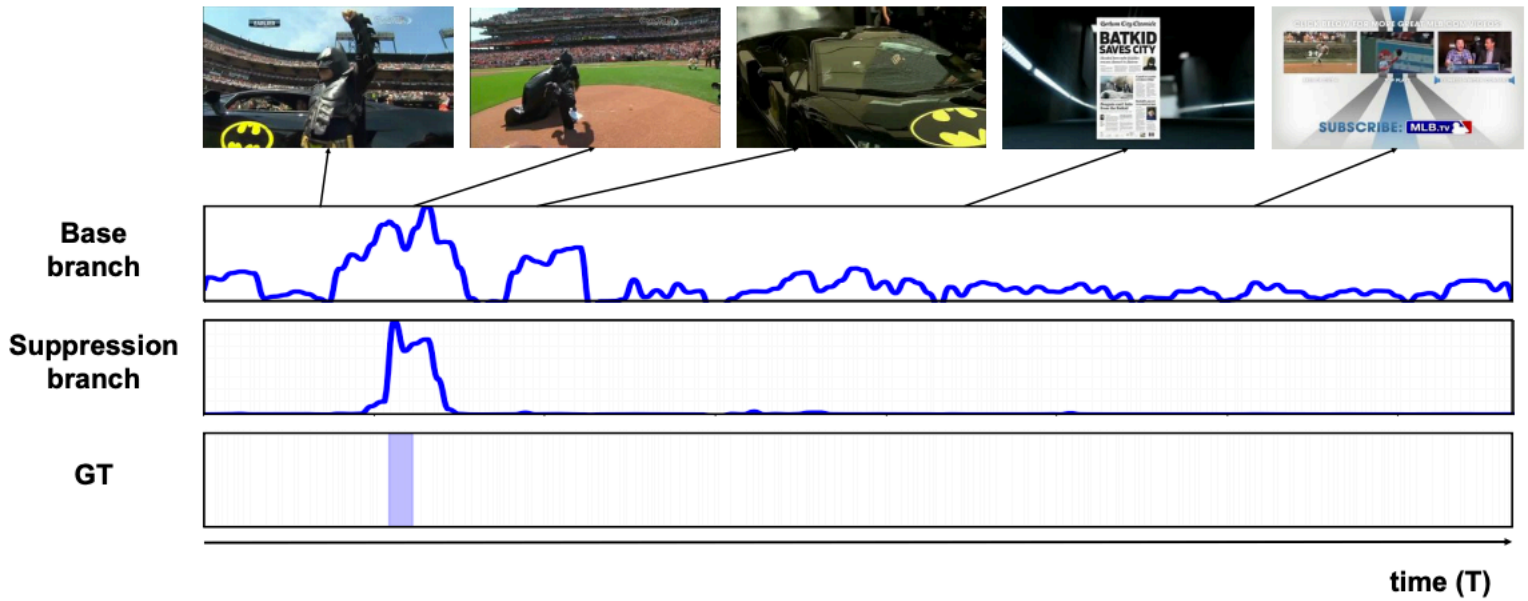
  - 단점: **background가 어느 액션 클래스에도 속하지는 않지만 그걸 모델링하지는 않음**

    - background frame은 action이 없어도 action class로 분류됨

01. Introduction

(3) background에 대한 auxiliary class 제안

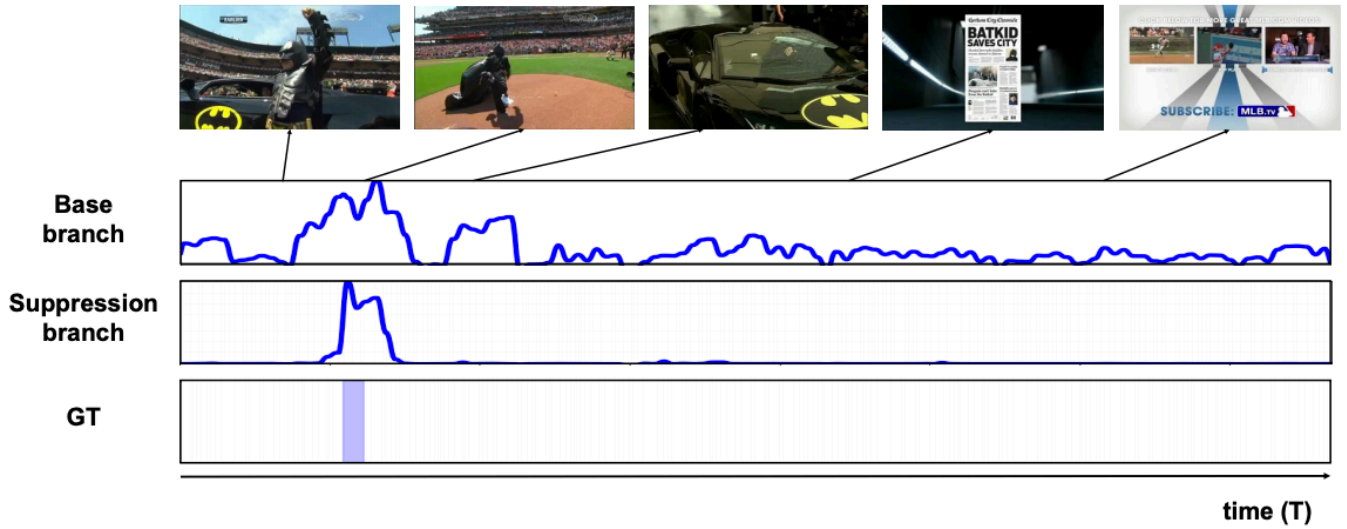
- 모든 untrimmed video에는 background가 포함 → positive sample으로 작용
- video에서 모든 frame에 targeting할 올바른 클래스가 있게 된다
- 그러나 / background 클래스에 대한 negative sample이 없음
- network는 background에서 높은 점수를 얻는 방향으로 학습



01. Introduction

(4) BaSNet : Background Suppression Network

- **Base branch (MIL 구조)**
  - frame 별 feature를 input으로 가진다
  - CAS(frame wise class activation sequence)를 생성 : action/background 클래스에 대한 positive sample로 분류
- **Suppression branch**
  - background frame의 input feature 를 약화
  - base branch를 따르고 가중치를 공유하는 filtering module로 시작
  - 목적: 모든 비디오에서 background class의 점수를 최소화 하는 것. action class를 최적화하는 것



01. Introduction

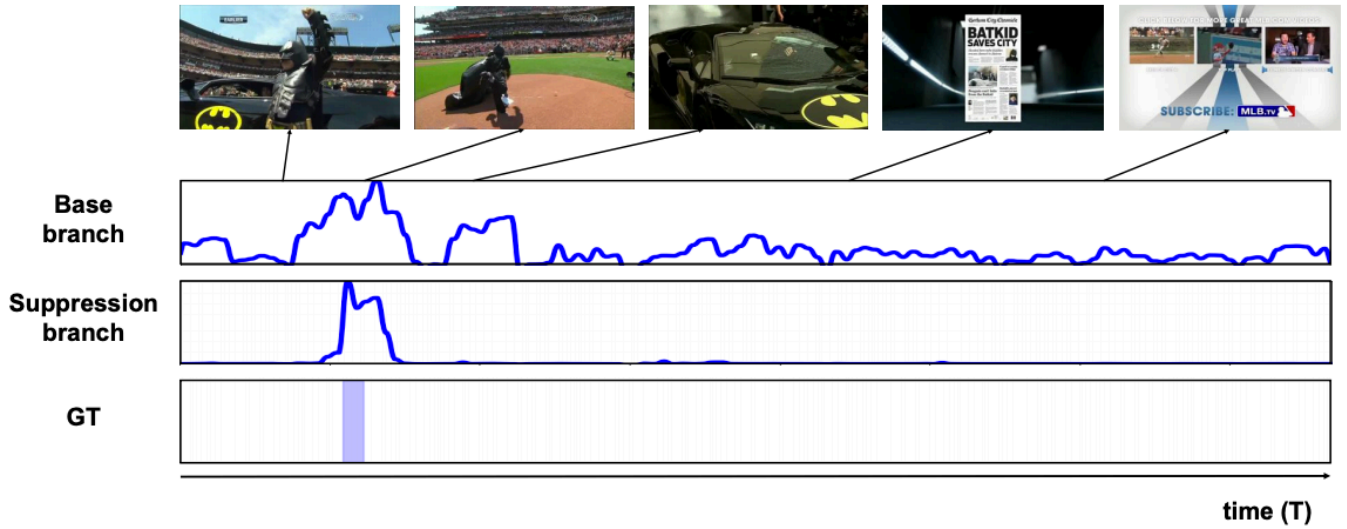
(4) BaSNet : Background Suppression Network

- base branch, suppression branch는 **가중치를 공유**

  → 같은 입력 주어지면 대조되는 것에 대해 서로 최적화를 하지는 못함

  - 이를 해결하기 위해 filtering module은 background에서 활성화를 막음

-> suppression branch는 **background로부터 더 자유로워지고 localization을 더 잘함**



01. Introduction

(5) Contribution

- 기존에 누락된 background를 모델링하기 위해 WTAL에서 **보조 클래스** 제안
- 2 branch **weight 공유 아키텍쳐 제안**
- **filtering module**

  → backgroudn frame에서 활성화를 억제

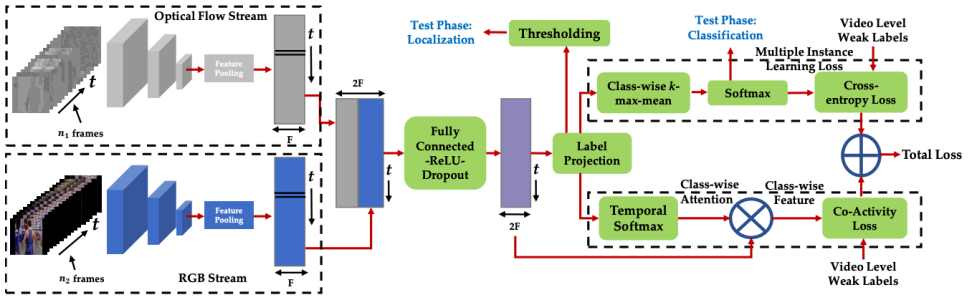
02. Related Work

TAL, WTAL

(1) TAL (Fully-supervised Temporal Action Localization)

: action의 시간적 간격을 같이 본다. 이전의 것들은 full supervision에 의존

- **SCNN** : 슬라이딩 → C개의 액션 + 1 백그라운드 클래스로 분류
- **TAL-Net** : 객체 감지 알고리즘을 TAL로 생성
- **BSN** : 정교한 proposal generation
- **GTAN** : 가우시안 시간 모델링



(2) WTAL (Weakly-supervised Temporal Action Localization)

- **H&S, MAAN, CMCS** : CAS가 소수의 프레임에 중점을 두는 문제 해결
- **STPN** : CAS와 함께 class에 구애받지 않는 가중치 사용
- **Autoloc** : threholding대신 regression을 이용해 proposal 생성
- **UNT, W-TALC, STAR** : MIL사용. 그러나 background class를 고려하지 않아 background는 모두 action class로 분류

03. Method

(1) Background Class

- background class없을 때 background에 대한 activation은 노이즈로 작용

→ background 나타내는 **auxiliary class** 제안

그러나

- training에서 background에 대한 positive sample 작용, negative sample 존재 X
- CAS가 항상 높은 경우, **data imbalance 문제 발생**

→ **background class 추가하는 것만으로는 큰 성능향상이 일어나지는 않음**

03. Method

(2) Two-branch Architecture (video level score predict 관점에서의 차이)

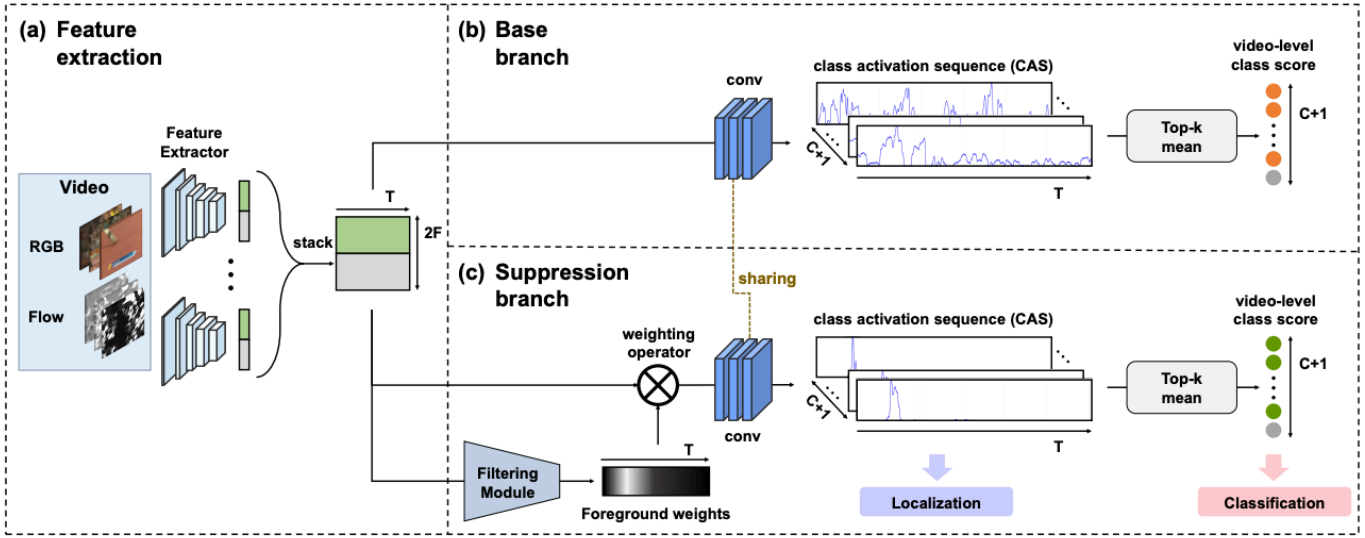
A. suppression branch은 filtering module을 가짐

- CAS에서 background frame를 필터링하는 것을 학습

B. 학습 목표가 다름

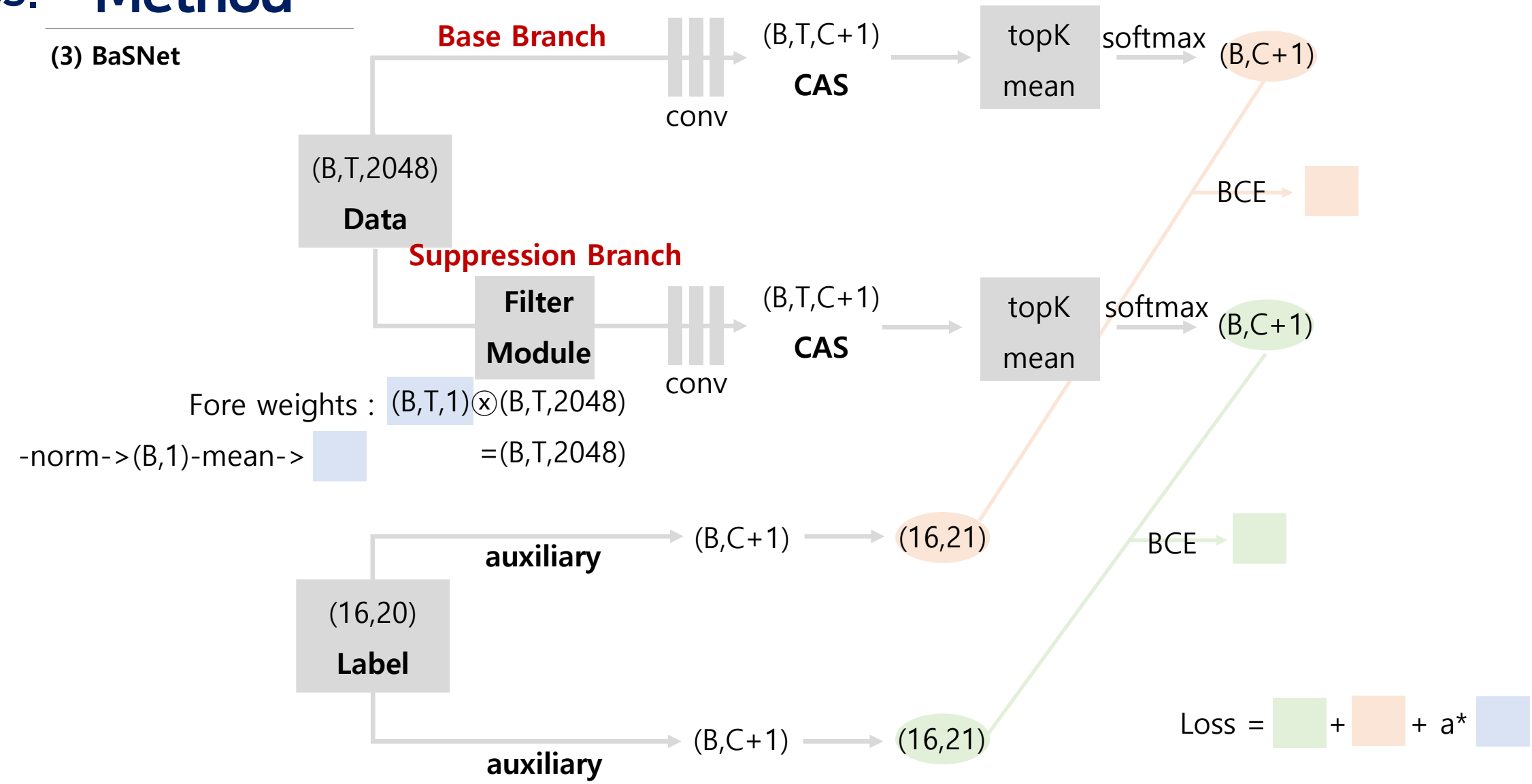
- **Base branch:** input video를 action class+background class의 positive sample로써 분류하는 것
- **Suppression branch:** 기존 action class에서 background class score를 최소화하는 것
- 가중치 공유는 2 브랜치에서 같은 인풋이 주어졌을 때 각자의 목표를 이루는 것을 방해

→ **filtering module으로 해결** : background frame으로부터 suppress activation을 학습, localization 성능 향상



03. Method

(3) BaSNet



Background Suppression Network for Weakly-supervised Temporal Action Localization, *AAAI 2020*

04. Experiments

Thumos14

Supervision	Method	mAP@IoU 0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Full	Richard et al. (2016)	39.7	35.7	30.0	23.2	15.2	-	-	-	-
	S-CNN (2016)	47.7	43.5	36.3	28.7	19.0	10.3	5.3	-	-
	Yeung et al. (2016)	48.9	44.0	36.0	36.0	36.0	26.4	17.1	-	-
	PSDF + T-SVM (2016)	51.4	42.6	33.6	26.1	18.8	-	-	-	-
	CDC (2017)	-	-	40.1	29.4	23.3	13.1	7.9	-	-
	Yuan et al. (2017)	51.0	45.2	36.5	27.8	17.8	-	-	-	-
	CBR (2017)	60.1	56.7	50.1	41.3	31.0	19.1	9.9	-	-
	R-C3D (2017)	54.5	51.5	44.8	35.6	28.9	-	-	-	-
	SSN (2017)	66.0	59.4	51.9	41.0	29.8	-	-	-	-
	SSAD (2017)	50.1	47.8	43.0	35.0	24.6	-	-	-	-
	TPC (2018)	-	-	44.1	37.1	28.2	20.6	12.7	-	-
	TAL-Net (2018)	59.8	57.1	53.2	**48.5**	**42.8**	**33.8**	**20.8**	-	-
	Action Search (2018)	51.8	42.4	30.8	20.2	11.1	-	-		
	BSN (2018)	-	-	53.5	45.0	36.9	28.4	20.0	-	-
	GTAN (2019)	**69.1**	**63.7**	**57.8**	47.2	38.8	-	-	-	-
Weak†	STAR (2019)	**68.8**	**60.0**	**48.7**	**34.7**	**23.0**	-	-	-	-
Weak	UntrimmedNet (2017)	44.4	37.7	28.2	21.1	13.7	-	-	-	-
	Hide-and-seek (2017)	36.4	27.8	19.5	12.7	6.8	-	-	-	-
	STPN (UNT) (2018)	45.3	38.8	31.1	23.5	16.2	9.8	5.1	2.0	0.3
	AutoLoc (2018)	-	-	35.8	29.0	21.2	13.4	5.8	-	-
	W-TALC (UNT) (2018)	49.0	42.8	32.0	26.0	18.8	-	6.2	-	-
	Liu et al. (UNT) (2019)	53.5	46.8	37.5	29.1	19.9	12.3	6.0	-	-
	Ours (UNT)	**56.2**	**50.3**	**42.8**	**34.7**	**25.1**	**17.1**	**9.3**	**3.7**	**0.5**
	STPN (I3D) (2018)	52.0	44.7	35.5	25.8	16.9	9.9	4.3	1.2	0.1
	W-TALC (I3D) (2018)	55.2	49.6	40.1	31.1	22.8	-	7.6	-	-
	MAAN (2019)	**59.8**	50.8	41.1	30.6	20.3	12.0	6.9	2.6	0.2
	Liu et al. (I3D) (2019)	57.4	50.8	41.2	32.1	23.1	15.0	7.0	-	-
	Ours (I3D)	58.2	**52.3**	**44.6**	**36.0**	**27.0**	**18.6**	**10.4**	**3.9**	**0.5**

04. Experiments

Thumos14

	Base branch	background class	Suppression branch	mAP@IoU 0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	AVG
baseline	✓			32.3	25.2	19.8	15.9	12.0	8.7	4.7	1.4	0.2	13.4
Base branch	✓	✓		28.5	23.0	18.1	13.7	9.2	5.8	2.7	0.8	0.1	11.3
Suppression branch		✓	✓	49.1	42.5	33.5	26.0	18.6	12.8	6.2	2.0	**0.5**	21.2
BaS-Net	✓	✓	✓	**58.2**	**52.3**	**44.6**	**36.0**	**27.0**	**18.6**	**10.4**	**3.9**	**0.5**	**27.9**

Table 5: Performances for detecting background frames on THUMOS'14 (F-measure).

	Base branch	Suppression branch	BaS-Net
F-measure	0.541	0.775	**0.846**

05. Conclusion

weak supervised temporal action localization

이전의 방법(supervised TAL)은 background가 action class로 잘못 분류된 문제를 다루지는 않음

(1) background에 대한 **auxiliary class** 소개

(2) BasNet 소개 : 비대칭적으로 학습하는 **2브랜치 가중치 공유** 모델

→ BasNet은 background frame에서 활성화를 억제하여 localization 성능 향상

(3) Thumos14, ActivityNet에서 SOTA