JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Poster ID: THU-AM-004

Post-Processing Temporal Action Detection




Sauradip Nag[1,2] **Xiatian Zhu**[1,3] **Yi-Zhe Song**[1,2] **Tao Xiang**[1,2]

[1]CVSSP, University of Surrey, UK
[2]IFlyTek-Surrey Joint Research Centre for Artificial Intelligence, UK
[3]People Centered AI, University of Surrey, UK

CVSSP Centre for Vision, Speech and Signal Processing

UNIVERSITY OF SURREY

People-Centred AI UNIVERSITY OF SURREY

iFLYTEK

Task Description : Revisiting Temporal Action Detection

Untrimmed Video

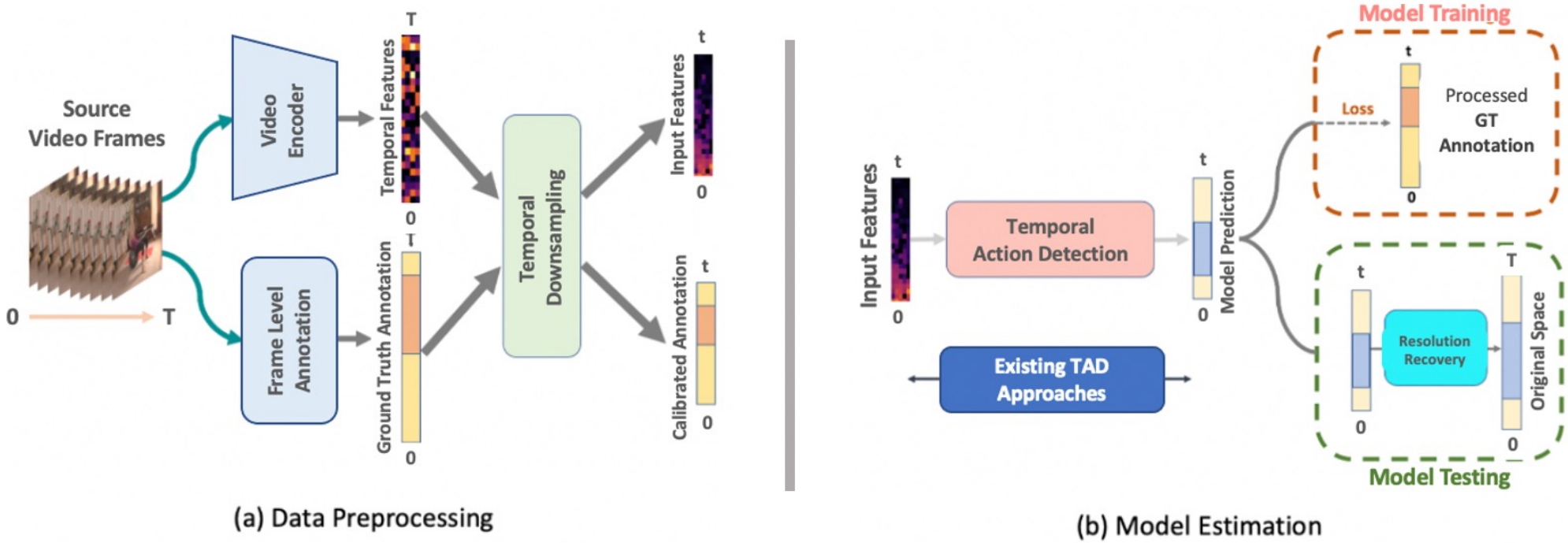
What is the Activity ?
("Playing Ice Hockey")

When is the Activity Occuring ?
(13 s – 28 s)

Sub-Task 1 :
Action **Classification**

Sub-Task 2 :
Temporal **Regression**

Current Approaches: Data Pre-Processing and Model Estimation



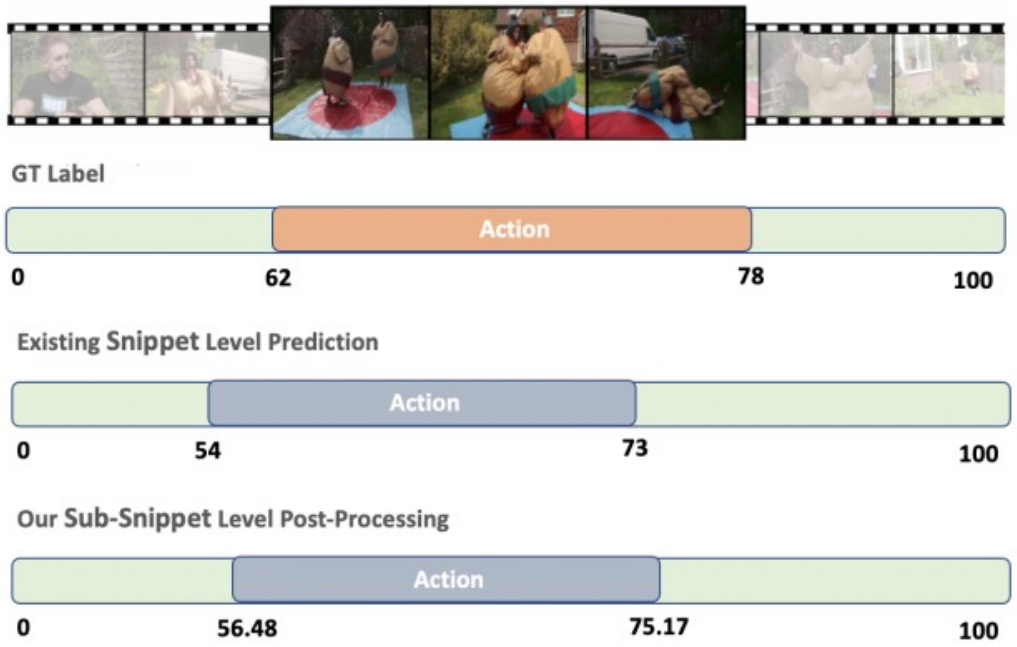
Downsample the Temporal Resolution ; Detect Actions in downsampled resolution ; Inference in original resolution

Quantization Error : How does it Arise ?

Quantization Error (in red) is introduced at start/end points of action due to **ceiling/floor/round operations**.

We solve this by **fitting Gaussian Distribution** at start/end points.

Quantization Error : How to Solve ?

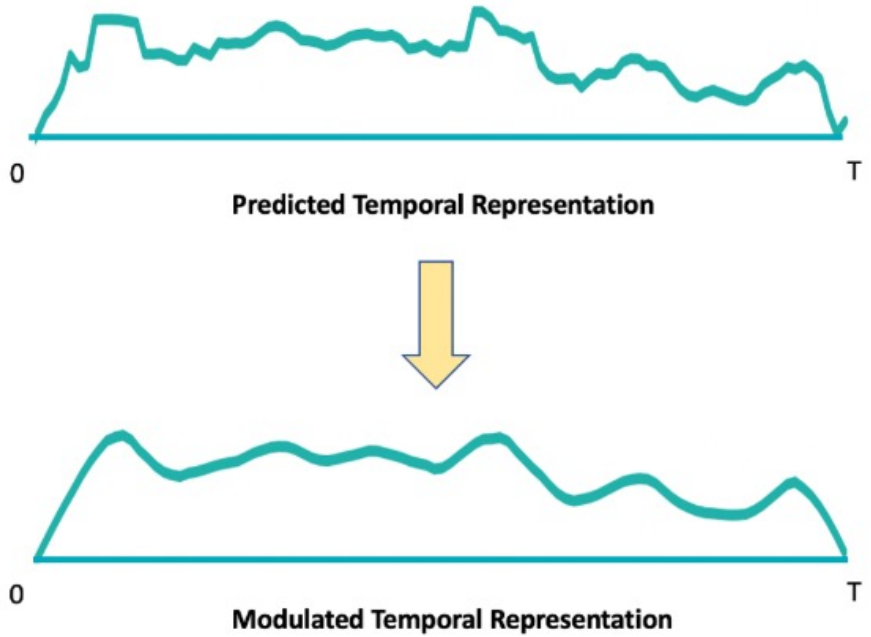


Instead of **Standard Snippet Level Prediction** we predict at **Sub-Snippet Level**

Each Snippet may contain action/background frame

Gaussian Approximated Post-Processing

Step 1: Temporal Smoothing

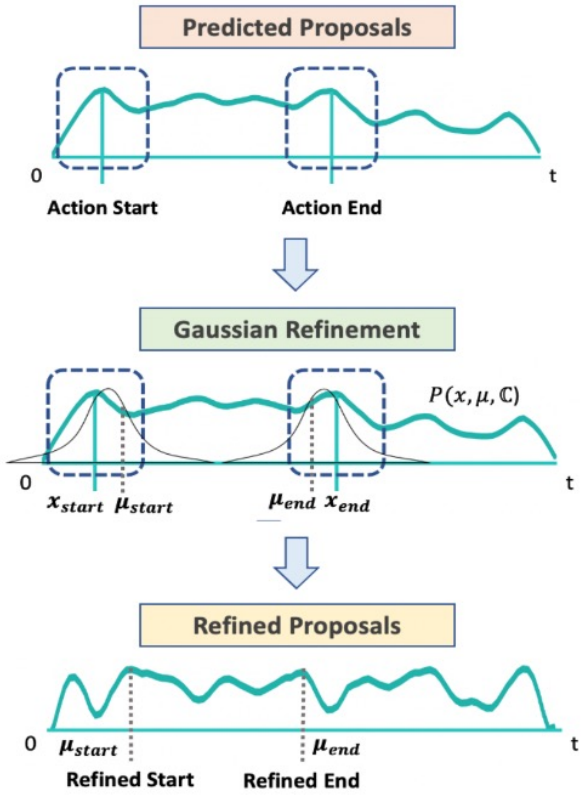


Given a Action **Temporal Snippet Representation** ,

We smooth the distribution using **Gaussian Kernel** to **avoid multiple peaks**.

Gaussian Approximated Post-Processing

Step 2: Gaussian Refinement and Calibration



Given a predicted boundary point (start/end) :

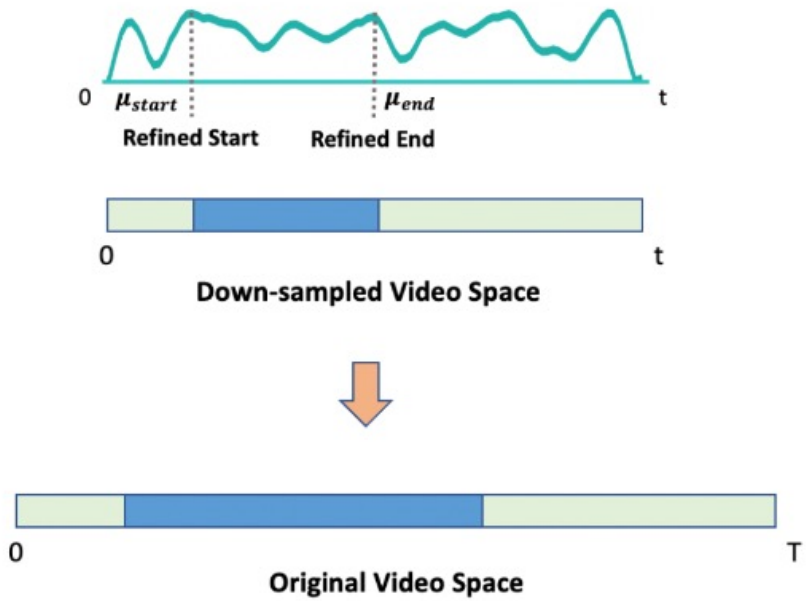
a) We fit **Gaussian Distribution** at the boundary point

b) We find the **optimal Shift** using **Taylors Expansion**

c) The newly shifted start/end points are at **sub-snippet level**

Gaussian Approximated Post-Processing

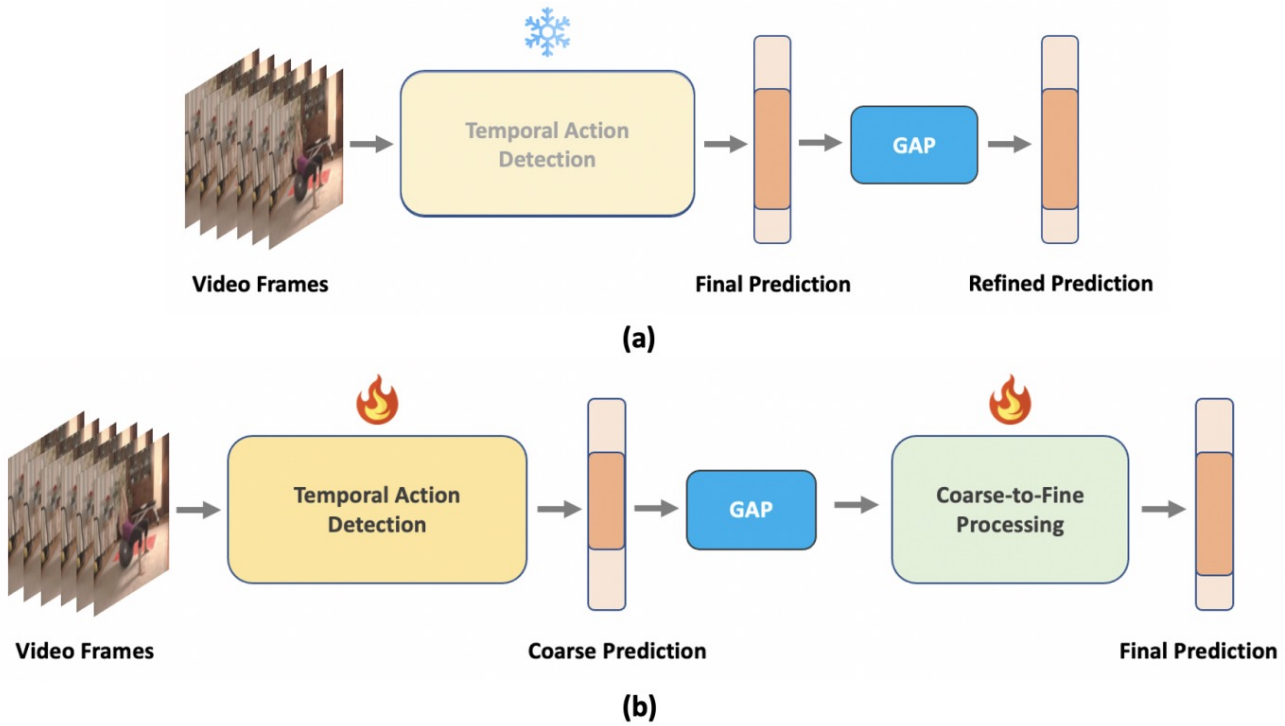
Step 3: Temporal Resolution Recovery



We **recover the temporal resolution** by multiplying the Video Duration

The **refined start/end point** reduces the **quantization error** at **sub-snippet level**

GAP Integration : Plug-And-Play Module



Can be used during both

(a) Model Retraining

(b) Inference Post-Processing

GAP Improves on Standard Supervised Action Detection

Category	Method	ActivityNet mAP				THUMOS14 mAP			
		0.5	0.75	0.95	Avg	0.3	0.5	0.7	Avg
Anchor-based	MUSES	50.0	34.9	6.5	34.0	68.9	56.9	31.0	53.4
	MUSES + **GAP**	**50.3**	**35.5**	**6.9**	**34.3**	**69.3**	**57.8**	**31.9**	**53.8**
	PBRNet	53.9	34.9	8.9	35.0	58.5	51.3	29.5	-
	PBRNet + **GAP**	**54.4**	**35.4**	**9.2**	**35.2**	**59.2**	**51.9**	**30.0**	-
Anchor-Free	BMN	50.1	34.8	8.3	33.9	56.0	38.8	20.5	38.5
	BMN + **GAP**	**50.5**	**35.2**	**8.6**	**34.3**	**56.6**	**39.4**	**21.0**	**38.9**
	GTAD	50.4	34.6	9.0	34.1	54.5	40.2	23.4	39.3
	GTAD + **GAP**	**50.8**	**34.9**	**9.2**	**34.4**	**55.0**	**40.5**	**23.8**	**39.6**
	DCAN	51.8	35.9	9.4	35.4	68.2	54.1	32.6	-
	DCAN + **GAP**	**52.4**	**36.4**	**9.6**	**35.8**	**68.6**	**54.6**	**33.0**	-
	RTDNet	47.2	30.7	8.6	30.8	68.3	51.9	23.7	-
	RTDNet + **GAP**	**47.7**	**31.1**	**8.8**	**31.2**	**68.8**	**52.3**	**24.2**	-
	AFSD	52.4	35.3	6.5	34.4	67.3	55.5	31.1	52.0
	AFSD + **GAP**	**53.0**	**35.9**	**7.1**	**34.8**	**68.0**	**56.1**	**31.5**	**52.5**
	ActionFormer	53.5	36.2	8.2	35.6	82.1	71.0	43.9	66.8
	ActionFormer + **GAP**	**53.9**	**36.4**	**8.5**	**36.0**	**82.3**	**71.4**	**44.2**	**66.9**
	React	-	-	-	-	69.2	57.1	35.6	55.0
	React + **GAP**	**-**	**-**	**-**	**-**	**69.5**	**57.3**	**35.7**	**55.2**
Proposal-Free	TAGS	56.3	36.8	9.6	36.5	68.6	57.0	31.8	52.8
	TAGS + **GAP**	**56.7**	**37.2**	**9.8**	**36.7**	**69.1**	**57.4**	**32.0**	**53.0**

SOTA on standard benchmarks
Like **THUMOS14** and **ActivityNet**

GAP Also Improves on Any-Shot/ Any-Supervision Action Detection

GAP on Weakly Supervised Setting

Model	mAP					
	0.3	0.4	0.5	0.6	0.7	Avg
ASL [5]	51.8	-	31.1	-	11.4	32.2
ASL [5] + **GAP**	**53.0**	**-**	**31.7**	**-**	**11.5**	**32.4**
CoLA [13]	51.5	41.9	32.2	22.0	13.1	40.9
CoLA [13] + **GAP**	**51.8**	**42.2**	**32.4**	**22.2**	**13.2**	**41.0**
TS-PCA [4]	52.4	43.5	34.6	23.7	12.6	-
TS-PCA [4] + **GAP**	**52.9**	**44.0**	**34.9**	**24.0**	**12.8**	**-**
CO2-Net [3]	54.5	45.7	38.3	26.4	13.4	-
CO2-Net [3] + **GAP**	**54.9**	**46.0**	**38.8**	**27.1**	**14.0**	**-**
ASM-Loc [2]	57.1	46.8	36.6	25.2	13.4	45.1
ASM-Loc [2] + **GAP**	**58.1**	**47.5**	**37.1**	**25.6**	**13.8**	**45.5**
DELU [1]	56.5	47.7	40.5	27.2	15.3	46.4
DELU [1] + **GAP**	**57.0**	**48.1**	**40.9**	**27.6**	**15.5**	**46.6**

GAP on Few-Shot Setting

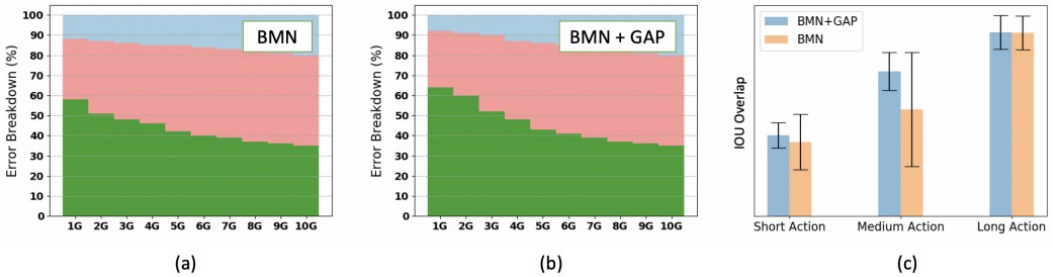
Shot	Models	mAP			
		0.5	0.7	0.9	Avg
1	QAT [10]	44.9	29.2	11.2	25.9
	QAT [10] + **GAP**	**45.8**	**30.0**	**11.8**	**26.5**
5	QAT [10]	51.8	32 .6	11.9	30.2
	QAT [10] + **GAP**	**52.2**	**32.9**	**12.1**	**30.4**

GAP on Zero-Shot Setting

Models	mAP			
	0.5	0.75	0.95	Avg
Baseline	28.0	16.4	1.2	16.0
Baseline† + **GAP**	**28.7**	**16.8**	**1.7**	**16.5**
Baseline + **GAP**	**28.2**	**16.6**	**1.3**	**16.2**
STALE [9]	32.1	20.7	5.9	20.5
STALE [9] + **GAP**	**32.4**	**21.1**	**6.2**	**20.8**

GAP is effective when Temporal Resolution is Small

Method	Temporal Resolution	mAP			
		0.5	0.75	0.95	Avg
BMN [9]	25	44.7	27.9	7.0	28.1
BMN+GAP	25	**45.5**	**28.4**	**7.3**	**28.5**
BMN	100	50.1	34.8	8.3	33.9
BMN+GAP	100	**50.5**	**35.2**	**8.6**	**34.3**
BMN	400	50.9	34.9	8.1	34.0
BMN+GAP	400	**51.1**	**35.0**	**8.2**	**34.1**



(a) (b) (c)

As **Temporal Resolution is Low**,
More Chances of Ambiguity among Action Boundaries

GAP is effective in such scenarios

As **Temporal Resolution Increases**,
Duration per snippet increases, **Ambiguity reduces**

GAP is less effective in such scenarios

Error-Sensitivity Analysis shows that

GAP does not bottleneck efficiency

Method	Inference Time	Speed
AFSD [8]	0.29 sec	1931 FPS
AFSD + **GAP**	0.31 sec	1792 FPS

GAP **does not** learn any extra parameters.

GAP may slightly **increase the training time** but negligible.

GAP has **almost negligible effect** on inference time.

UNIVERSITY OF SURREY

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Thank You for Listening

For any questions, contact : s.nag@surrey.ac.uk

Scan the QR Code for Code