

Limitations of Similarity-Based Context Utilization

Similarity-Based methods still suffer from cluttered background information and limited contextual feature learning.

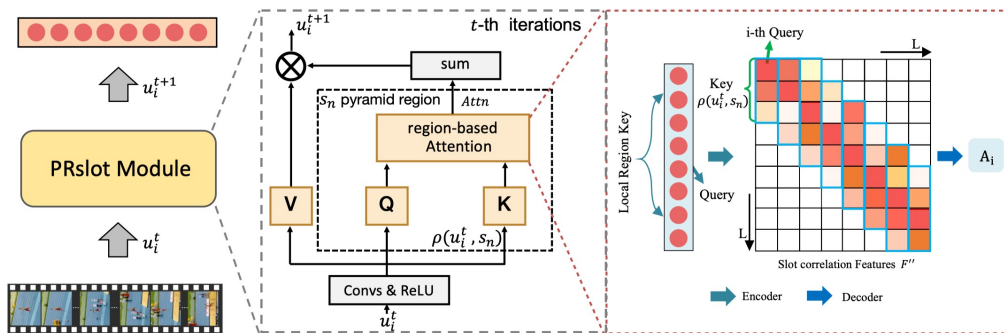


Our PRSA-Net selectively contextualize local semantic information instead of pairwise similarity.

Slot Attention for action detection

We refer to the idea of Slot attention^[1], which takes the snippet-level features as input and maps them to a set of output vectors by aggregating local region context that we refer to as slots.

Region-based Attention Mechanism.



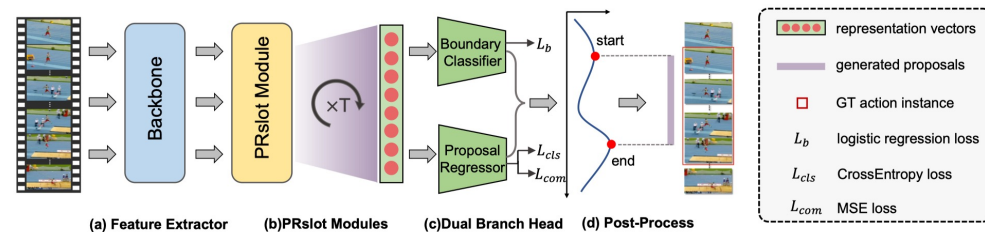
$$\rho(u_i, s) = \{u_{i-s}, \dots, u_{i+s}\} \quad F''_{i,j} = \begin{cases} F'_{i,j}, \\ 0, \end{cases} \quad \begin{matrix} u_i \in \rho(u_j, s), \\ \text{Otherwise,} \end{matrix}$$

$$A_{i,j} = \begin{cases} f_{\text{region}}^{\text{dec}}(F''_{i,j}), \\ 0, \end{cases} \quad \begin{matrix} u_i \in \rho(u_j, s), \\ \text{Otherwise,} \end{matrix}$$

$$A_i = f_{\text{region}}(u_i, \rho(u_i, s)).$$

Directly estimates the snippet interactions by the content of the snippet and its local surroundings instead of similarity-based.

The architecture of PRSA-Net



Experiments

Proposal generation on THUMOS14 and ActivityNet-1.3 in terms of AR@AN(%)

Method	Backbone	@50	@100	@200	@500
MGG [19]	TSN	39.93	47.75	54.65	61.36
BSN [17] + SNMS	TSN	37.46	46.06	53.21	60.64
BMN [18] + SNMS	TSN	39.36	47.72	54.70	62.07
BC-GNN [1] + NMS	TSN	41.15	50.35	56.23	61.45
BU-TAL [34]	I3D	44.23	50.67	55.74	-
BSN++ [25] + SNMS	TSN	42.44	49.84	57.61	65.17
RTD-Net [26]	I3D	41.52	49.32	56.41	62.91
Ours + NMS	TSN	47.49	55.14	60.18	63.53
Ours + SNMS	TSN	44.11	52.52	59.19	65.12
Ours + NMS	I3D	49.06	56.12	61.30	63.20
Ours + SNMS	I3D	45.81	53.13	59.32	66.32

Method	AR@1	AR@50	AUC
BSN [17]	32.17	74.16	66.17
MGG [19]	-	74.54	66.43
BMN [18]	-	75.01	67.10
BC-GNN [1]	-	76.73	68.05
BU-TAL [34]	-	75.27	66.51
RTD-Net [26]	33.05	73.21	65.78
Ours	35.37	76.90	69.21

Iteration and update strategy.

The recurrent function (GRU) is time-consuming and achieves limited performance boost. we apply a variety of local regions to exploit the slot representation completely.

Iteration Strategy	@50	@100	@200
GRU	42.9	52.8	55.6
Ours	49.1	56.1	61.3

Iteration times	@50	@100	@200	@500
1	48.4	55.3	60.1	61.6
2	49.1	56.1	61.3	63.2
3	48.9	55.7	61.9	62.3

Models and code are publicly available:

<https://github.com/handhand123/PRSA-Net>

[1] Francesco Locatello, et al, Object-centric learning with slot attention., NIPS2020