Learning Group Activity Features Through Person Attribute Prediction

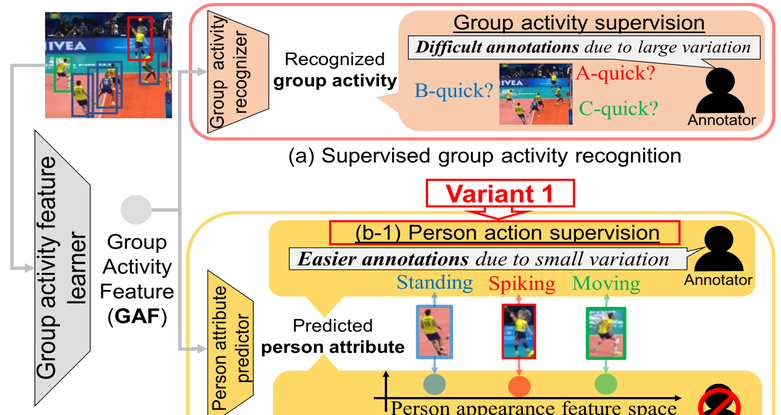
Chihiro Nakatani[1], Hiroaki Kawashima[2], Norimichi Ukita[1]

[1] Toyota Technological Institute, Japan [2] University of Hyogo, Japan

Introduction

What is the novelty of our method?

Group Activity Feature Learning (GAFL) **without group activity labels**.

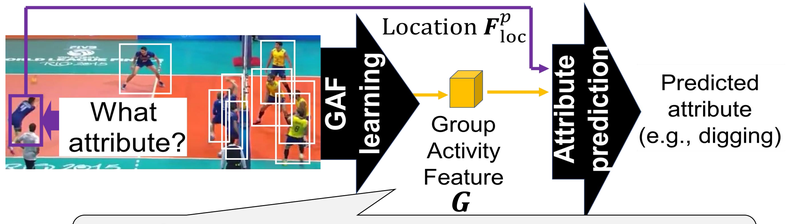


Two variants are **independent** and can be used individually.

(b) Ours (group activity feature learning with person attribute)

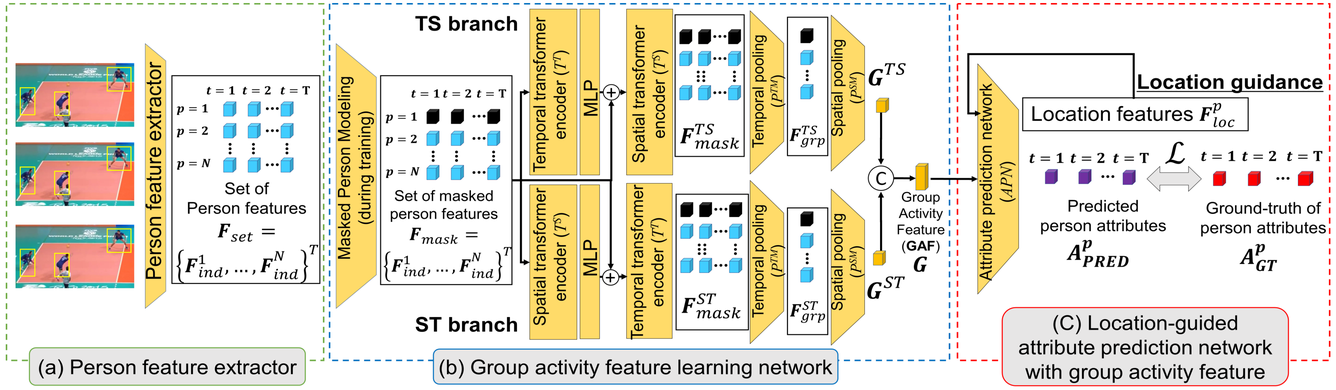
How is the Group Activity Feature (GAF) learned?

Location-Guided Person Attribute Prediction Using GAF.



G can be used for various downstream tasks (e.g., retrieval, clustering, and recognition)

Proposed Method (Group Activity Feature Learning: GAFL)



Experiments

Experimental Setting

Variant 1 GAFL-PAC (**supervision** with Person Action Classes)

Variant 2 GAFL-PAF (**self-supervision** with Person Appearance Features)

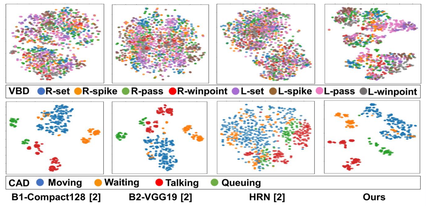
Dataset

VolleyBall Dataset (VBD), Collective Activity Dataset (CAD)

Ablation study

	Retrieval type	Action set (IoU)	Action set (AF-IDF)	Group activity
	Method	Hit@1	Hit@1	Hit@1
GAFL-PAC	Ours w/o F_{loc}	80.0	75.6	69.5
	Ours	83.0	80.1	84.8
GAFL-PAF	Ours w/o F_{loc}	64.1	51.7	53.2
	Ours	64.8	52.3	61.1

Visualization of learned GAF



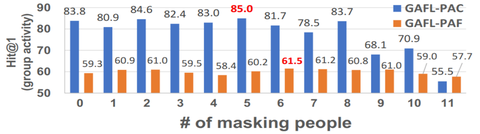
Quantitative results (VBD)

	Retrieval type	Action set (IoU [2])				Action set (AF-IDF)				Group activity			
	Method	Hit@1	Hit@2	Hit@3	mAP	Hit@1	Hit@2	Hit@3	mAP	Hit@1	Hit@2	Hit@3	
GAFL PAC	HiGCIN [3]	74.3	84.9	89.5	55.7	59.8	73.6	80.3	30.5	50.0	66.3	74.5	
	DIN [4]	79.7	90.1	93.4	60.2	74.5	85.2	88.3	39.3	57.0	73.1	81.1	
	Dual-AI [1]	67.6	84.7	91.6	56.9	72.6	83.7	88.6	53.0	64.4	76.5	82.0	
	Ours-ind	82.7	91.6	95.0	59.1	79.0	86.8	89.8	45.6	82.7	88.8	91.3	
	Ours-grp	83.0	92.7	95.5	64.2	80.1	88.4	91.5	59.9	84.8	89.6	91.8	
GAFL PAF	B1-Compact128 [2]	57.9	75.7	84.3	45.8	41.3	60.8	71.4	29.3	30.3	48.0	59.9	
	B2-VGG19 [2]	63.8	80.6	86.8	46.8	46.7	65.8	75.7	29.4	35.4	53.6	65.0	
	HRN [2]	60.9	78.6	86.0	46.9	40.8	60.9	72.9	28.7	31.2	47.0	57.6	
	Ours-ind	64.2	80.8	88.3	45.0	50.4	69.3	77.6	30.1	55.0	72.3	79.2	
	Ours-grp	64.8	82.7	90.3	46.4	52.3	71.4	81.0	31.4	61.1	75.1	82.4	

Quantitative results (CAD)

	Retrieval type	Action set (IoU [2])				Action set (AF-IDF)				Group activity			
	Method	Hit@1	Hit@2	Hit@3	mAP	Hit@1	Hit@2	Hit@3	mAP	Hit@1	Hit@2	Hit@3	
GAFL PAC	HiGCIN [3]	80.8	85.4	89.7	57.9	81.0	85.2	89.3	61.6	86.1	88.8	91.9	
	DIN [4]	71.4	74.1	74.9	51.5	90.1	92.7	94.0	52.8	90.8	92.5	93.2	
	Dual-AI [1]	61.0	72.5	76.7	61.5	85.5	86.9	88.1	82.7	82.1	84.1	84.7	
	Ours-ind	76.2	82.6	89.4	78.9	94.8	95.6	95.9	82.2	94.9	95.4	95.7	
	Ours-grp	81.8	90.7	93.5	69.9	96.1	96.5	96.6	93.9	94.9	95.6	96.3	
GAFL PAF	B1-Compact128 [2]	48.8	60.3	68.2	38.0	81.8	88.2	89.7	52.6	82.4	88.4	90.1	
	B2-VGG19 [2]	53.6	61.6	66.1	35.3	71.1	80.3	83.8	46.7	72.2	80.8	84.2	
	HRN [2]	37.1	50.1	58.6	22.2	53.2	64.8	72.5	34.2	54.0	64.8	72.4	
	Ours-ind	67.6	81.3	85.9	53.3	83.7	88.9	90.2	57.5	88.5	91.2	91.9	
	Ours-grp	52.7	70.3	74.1	46.4	74.0	80.5	82.6	60.1	79.2	81.0	82.0	

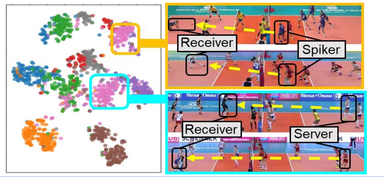
Masked Person Modeling effectiveness (VBD)



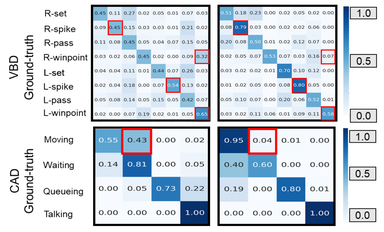
Future Work

Exploring other pretext tasks such as predicting the joint attention of a group (e.g., [5])

Fine granularity of our GAF



Group activity recognition (by nearest neighbor retrieval)



References

[1] Han et al. Dual-ai: Dual-path actor interaction learning for group activity recognition. CVPR, 2022

[2] M. Ibrahim et al. Hierarchical relational networks for group activity recognition and retrieval. ECCV, 2018

[3] Yan et al. Higcin: Hierarchical graph-based cross inference network for group activity recognition. TPAMI, 2023

[4] Yuan et al. Spatio-temporal dynamic inference network for group activity recognition. ICCV, 2021.

[5] Nakatani et al. Interaction-aware Joint Attention Estimation Using People Attributes. ICCV, 2023