FIFAWC: A Dataset with Detailed Annotation and Rich Semantics for Group Activity Recognition

Duoxuan PEI, Di HUANG, Yunhong WANG

Frontiers of Computer Science, ID: 10.1007/s11704-024-40027-3

Problems & Ideas

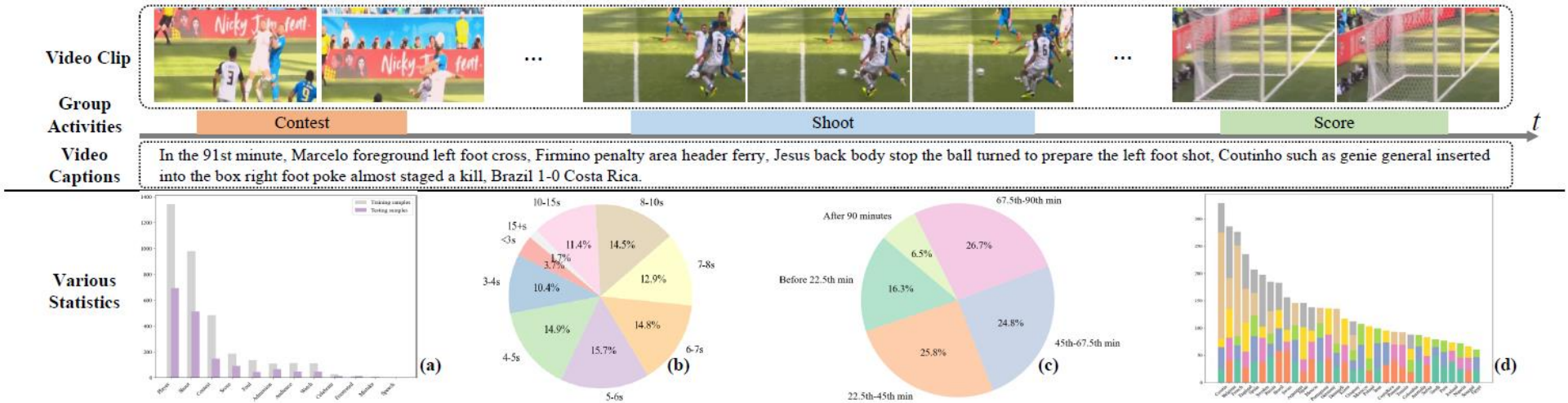
- Problems of existing group activity recognition (GAR) datasets:
  - Typically annotating only a single Group Activity (GA) instance per sample, carefully selected from original videos, diverges significantly from real-world contexts, which often involve multiple GA instances.
  - Single word-level annotations are insufficient to encapsulate the complex semantic information in GA, thereby constraining the expansion and research of other GA-related tasks.
- Ideas: Presenting a novel dataset for GAR to mitigate these limitations

Dataset	#Videos	#Instances	Scenario	Anno. Type	Labeling Level
CAD [1]	44	≈ 2,500	Pedestrian	Single GA	Word
VD [4]	55	4,830	Volleyball	Single GA	Word
VT [5]	12	4,960	Volleyball	Single GA	Word
NBA [6]	55	9,172	Basketball	Single GA	Word
Ours	64	5,196	**Soccer**	**Multi GAs**	**Word & Sen.**

Comparison of the existing datasets for GAR. 'Anno.' is the abbreviation for 'Annotation' and 'Sen.' is the abbreviation for 'Sentence'.

Main Contributions

- Contributions:
  - We propose the FIFAWC dataset for GAR characterized by three notable distinctions: comprehensive annotation, semantic description, new scenario.



The upper figure showcases a FIFAWC Dataset sample, depicting three group activities: 'Contest', 'Shoot', and 'Score', with English description. Subsequent figures detail FIFAWC's statistics: (a) sample distribution across categories, (b) duration distribution across samples, (c) timing distribution of sample occurrences during matches, and (d) team-wise sample distribution with distinct matches of the same team marked in varying colors.

GAR benchmark. Mean accuracy of all samples (MSA) indicates the recognition accuracy at sample level. We select 5 common categories to calculate Mean Per Class Accuracy (MPCA).

Method	MSA↑	Shoot	Score	Foul	Contest	Celebrate	MPCA↑
ARG [8]	51.0	69.3	87.6	93.9	80.1	98.2	85.8
DFWSGAR [9]	57.0	92.1	87.4	93.8	84.2	98.1	91.1

GA video captioning results. 'FT' is short for 'fine-tune'.

Method	FT	BLEU4↑	CIDEr↑	METEOR↑	ROUGE-L↑
PDVC [7]	✗	0.9	0.7	4.6	13.3
PDVC [7]	✓	7.0	8.3	9.9	24.5
VTimeLLM [10]	✗	1.6	2.0	5.1	15.5