

Player-Centric Multimodal Prompt Generation for Large Language Model Based Identity-Aware Basketball Video Captioning

Zeyu Xi¹, Haoying Sun¹, Yaofei Wu¹, Junchi Yan², Haoran Zhang¹, Lifang Wu^{1,✉}, Liang Wang³, and Changwen Chen⁴

1 Beijing University of Technology. 2 Shanghai Jiao Tong University. 3 Chinese Academy of Sciences. 4 The Hong Kong Polytechnic University

<https://github.com/Zeyu1226-ml/LLM-VC>

Introduction

Video:	
Captions:	
(a): A group of people are playing the basketball	
(b): [PLAYER] (✗) makes 2-pt jump shot (✗) (assist (✓) by [PLAYER] (✗))	
(c): D. Murray (✗) makes 2-pt jump shot (✗) (assist (✓) by W. Carter (✓))	
(d): T. Ross (✓) makes 3-pt jump shot (✓) (assist (✓) by W. Carter (✓))	

- Traditional video captioning methods provide only a broad summary of the visual content, which is far from real-world applications.
- Anonymous methods replace entities in descriptions with special tokens (e.g., [PLAYER] and [TEAM]), still inadequate for real-world applications.
- Knowledge-enhanced methods introduce extra information, such as game news, player statistics, and the candidate player list, to generate descriptions with player identities. These methods do not truly recognize player identities from the video clips, but rather make inferences based on extra information.
- The proposed method utilizes the pretrained player identification network to recognize player identities from visual content.

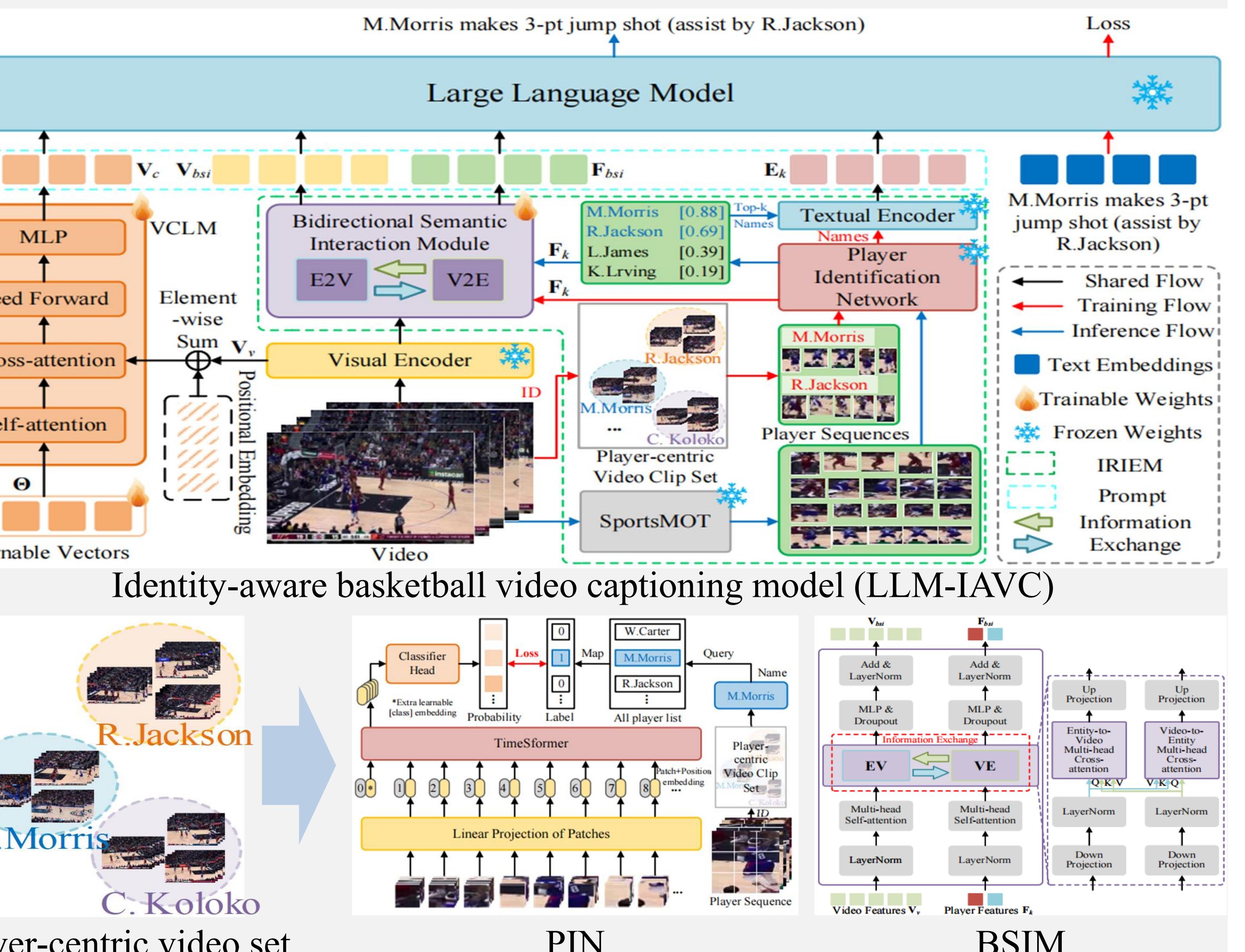
Contributions

- We propose a player-centric multimodal prompt generation network for LLM-based identity-action basketball video captioning (LLM-IAVC), which jointly utilizes player identity features and contextual video information to guide LLM in generating identity-aware descriptions.
- A large identity-aware basketball video captioning dataset NBA-Identity is constructed. This dataset contains the largest number of videos among identity-aware sports video captioning datasets. Its captions cover 9 major events in the basketball domain.
- Extensive experiments on our NBA-Identity and VC-NBA-2022 datasets demonstrate that the proposed model achieves state-of-the-art performance without the need for additional knowledge assistance.

Dataset

VideoID: 100001	
Captions:	Offensive rebound by C. LeVert
Bbox:	[[701, 349, 787, 500], [768, 348, 846, 499], [841, 345, 902, 510], [859, 358, 981, 517], [939, 381, 996, 524], [963, 410, 1028, 542], [953, 400, 1047, 542], [949, 395, 1059, 518], [933, 389, 1026, 526], [920, 372, 1013, 559], [864, 363, 990, 564], [850, 352, 924, 559]...]
Action:	Offensive rebound
Player:	C. LeVert
Start and end time:	[4, 7.6]

Method



Quantitative Results

Method	NBA-Identity				VC-NBA-2022			
	CIDEr	METEOR	Rouge-L	BLEU-4	CIDEr	METEOR	Rouge-L	BLEU-4
Traditional methods								
Clip4Caption [39]	21.7	21.0	37.6	6.5	70.4	26.7	51.2	28.8
SwinBERT [21]	23.5	20.7	37.1	7.5	69.1	26.5	49.0	28.4
CoCap [38]	22.7	20.0	37.7	7.5	70.5	27.4	50.7	28.9
OmniViD [44]	39.6	21.6	37.7	8.3	71.2	27.5	50.7	29.2
Extra information enhanced methods								
KEANet [47]	50.1	22.7	38.1	8.5	138.5	28.0	54.9	32.4
EIKA [48]	55.7	22.7	38.6	9.0	140.7	29.5	56.8	36.7
LLM-IAVC without prompt								
LLM-IAVC (G)	46.8	17.5	34.0	9.2	87.4	27.7	53.8	30.5
LLM-IAVC (Q-0.5B)	54.3	17.9	33.8	11.5	88.1	27.6	54.2	30.8
LLM-IAVC (Q-1.5B)	58.8	18.8	38.5	12.3	90.0	27.5	55.2	31.0
LLM-IAVC (Q-3B)	61.8	18.2	39.0	14.7	101.2	28.0	55.5	32.2
LLM-IAVC (L-1B)	60.3	18.0	33.2	14.0	90.0	27.3	55.3	31.4
LLM-IAVC (L-3B)	66.5	18.2	34.5	14.2	103.1	28.5	55.7	32.8
LLM-IAVC with prompt								
LLM-IAVC (G)	94.1	22.0	35.4	17.0	139.0	29.0	57.8	37.2
LLM-IAVC (Q-0.5B)	87.6	21.5	36.4	16.0	139.4	29.2	55.7	36.8
LLM-IAVC (Q-1.5B)	97.1	22.1	40.0	16.7	141.5	30.1	57.4	37.8
LLM-IAVC (Q-3B)	103.5	22.9	41.2	17.7	147.8	30.7	56.8	38.3
LLM-IAVC (L-1B)	88.3	21.6	40.1	15.3	139.3	30.0	57.3	37.7
LLM-IAVC (L-3B)	105.3	23.0	42.5	18.8	150.7	30.9	57.9	38.6

Qualitative Results

NBA-Identity		
GT: Turnover by J. Richardson (bad pass; steal by B. Hyland)		
OVD: Defensive rebound by T. Mann		
EIKA: Turnover by D. Garland (lost ball)		
G w/o: Defensive rebound by B. Brown		
Q w/o: J. Murray makes 3-pt jump shot from 21 ft		
L w/o: Turnover by J. McGee		
G w: Turnover by T. Mann (traveling)		
Q w: Turnover by J. Richardson (traveling)		
L w: Turnover by J. Richardson (offensive foul)		
VC-NBA-2022		
GT: Landry Shamet misses the 3pt jump shot and Jalen Smith gets the defensive rebound		
OVD: Chris Paul misses the 3pt jump shot and Tari Eason gets the defensive rebound		
EIKA: Landry Shamet misses the 3pt jump shot and Josh Okogie gets the defensive rebound		
G w/o: Kevinise III makes the 3pt jump shot with an assist from Josphine Winslow		
Q w/o: Kevin Porter Jr. makes the 2pt layup		
L w/o: Marcus Morris makes the 2pt layup		
G w: Landry Shamet misses the 3pt jump shot and Josh Okogie gets the defensive rebound		
Q w: Landry Shamet misses the 3pt jump shot and Jalen Smith gets the defensive rebound		
L w: Landry Shamet misses the 3pt jump shot and Jalen Smith gets the defensive rebound		
NBA-Identity		
GT: D. Finney-Smith misses 2-pt jump shot from 4 ft (block by K. Durant)		
OVD: C. Koloko misses 3-pt jump shot from 20 ft		
EIKA: D. Finney-Smith misses 2-pt jump shot from 15 ft		
G w/o: F. VanVleet misses 3-pt jump shot from 22 ft		
Q w/o: T. Maledon misses 3-pt jump shot from 14 ft		
L w/o: D. Smith misses 3-pt jump shot from 7 ft (block by C. Boucher)		
G w: D. Finney-Smith misses 2-pt jump shot from 2 ft (block by D. Vassell)		
Q w: M. Plumlee misses 2-pt jump shot from 7 ft (block by K. Durant)		
L w: D. Finney-Smith misses 2-pt jump shot from 6 ft (block by K. Durant)		
VC-NBA-2022		
GT: Jayson Tatum makes the 2pt layup		
OVD: Justin Jackson makes the 2pt layup		
EIKA: Jayson Tatum makes the 2pt jump shot		
G w/o: Al Horford makes the 2pt layup		
Q w/o: Justin Jackson makes the 2pt jump shot with an assist from Jayson Brown		
L w/o: Jamal Mitchell makes 3-pt jump shot		
G w: Jayson Tatum makes the 2pt jump shot		
Q w: Jayson Tatum makes the 2pt layup		
L w: Jayson Tatum makes the 2pt layup		