BASKET: A Large-Scale Video Dataset for Fine-Grained Skill Estimation

Supplementary Material

Our supplementary material consists of Additional Dataset Details (Section 1), Additional Implementation Details (Section 2), Additional Human Evaluation Details (Section 3), and Additional Results Analysis (Section 4).

1. Additional Dataset Details

In this section, we provide additional details of our BASKET dataset.

Visualization of Player Highlight Videos. Figure 1 showcases snapshots from five randomly selected player highlight videos used as inputs to the skill estimation models. For each video, we visualize eight uniformly sampled frames from the full highlight video.

Additional Details on Constructing Skill Labels. We also note that different leagues (e.g., college vs. professional) may exhibit skill-level differences. To account for these variations and avoid player comparisons across different leagues, we construct ground truth skill labels separately for each league. During model training/inference, the model then has to implicitly learn to predict skill levels that are specific to each basketball league.

Dataset Characteristics Analysis. Table 1 presents the distribution of players by seasons, player genders, and geographic locations. These numbers highlight the richness and diversity of our proposed dataset.

2. Additional Implementation Details

In this section, we provide additional details of model implementation.

Model Hyperparameters. In Table 2, we provide additional details about the hyperparameters used for all of our tested models.

LLaVA-OneVision Implementation. To fine-tune LLaVA-OneVision, we reformat our BASKET dataset into a text-based instruction-tuning format, commonly used by modern VLMs. Figure 2 provides a sample prompt used as input to the LLaVA-OneVision model. Specifically, we prompt the model to assign a numerical category to each of the twenty skills. We then evaluate the model's performance using top-1 accuracy based on the generated outputs.

3. Human Evaluation Details

Our user study is conducted online. To improve the reliability of the results and incentivize users, we used a two-level compensation scheme. Specifically, participants would receive a base compensation for completing each session with a bonus if their performance exceeded a certain accuracy threshold (i.e., 60%).

To optimize the study design, we conducted a preliminary test study with three participants to determine the appropriate number of player videos to review and the skills to evaluate. Based on feedback from this initial study, we finalized the study design to include the assessment of five players across five skills, each categorized into three levels of skills. We observed that evaluating a single-player video required approximately 10-12 minutes, and participants demonstrated consistent scoring accuracy for skills within the same coarse category. We also observed a noticeable decline in participant performance after one hour of video review, hence the motivation for the study design, where a single session could be completed within one hour.

To recruit participants for the study, we distributed advertisements in computer science department channels and online basketball group chats. Participants were asked about their basketball experience, including the number of years they spent watching or playing the sport. Based on their responses, all participants were categorized into novice, average, or expert groups.

4. Additional Results Analysis

In Table 3, we provide a detailed breakdown of how the best-performing VideoMamba model generalizes across different skill categories included in our BASKET dataset. Our results suggest that the average accuracy for coarse skill categories such as shooting, rebounding, defense, and playmaking are relatively similar. However, we also observe that offensive skills generally exhibit higher accuracy. We believe that the differences in accuracy may be attributed to the fact that many of the video clips in the player highlight videos are offensive plays. In contrast, defensive players are rarer and therefore more challenging for the model to learn to assess.

Figure 1. **Snapshots from the Player Highlight Videos.** Each row represents the video of a particular player. For these visualizations, we uniformly sample eight frames from the input video.

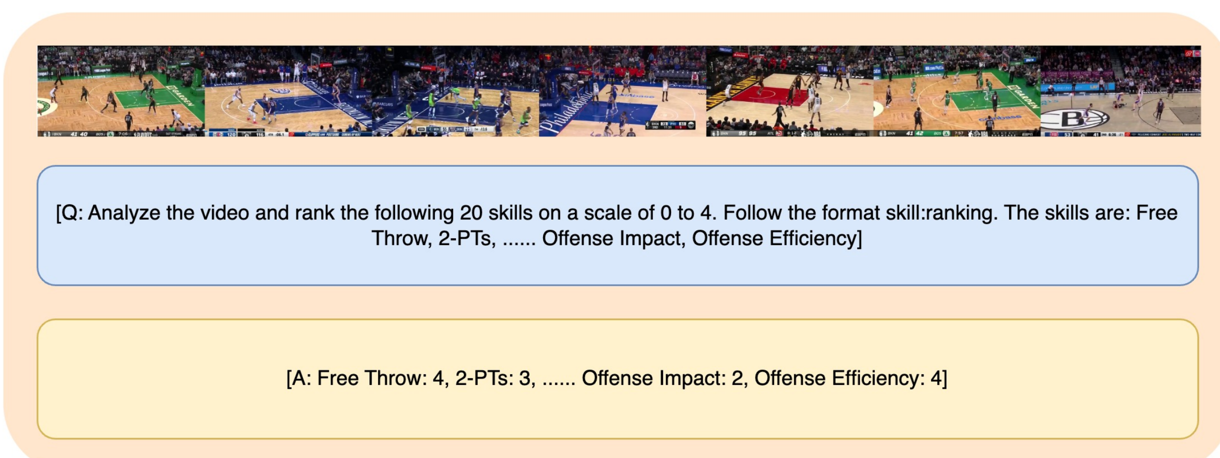

Figure 2. **A Sample Prompt Used to Train LLaVA-OneVision Model.** We convert our BASKET dataset into a text-based instruction tuning format commonly used by modern VLMs. The model takes the instruction to assign a numerical category to each of the twenty basketball skills and then generates a textual answer. We then evaluate the model's performance using top-1 accuracy based on the generated outputs.

Season	17-18	18-19	19-20	20-21	21-22	22-23
Num. Players	2360	3736	6047	5073	6879	8137

(a) Number of Players by Season

Gender	Male	Female
Num. Players	24669	7563

(b) Number of Players by Gender

Location	N. America	Europe	Asia	Australia
Num. Players	22632	8177	1045	378

(c) Number of Players by Location

Table 1. **Breakdown of Player Numbers.** We present the detailed player numbers of BASKET by (1) seasons, (2) gender, and (3) geographic locations, highlighting the diversity of our dataset.

Attribute	LLaVA-OV	MeMViT	SigLIP	VideoMAE2	X-CLIP	TSF	UMT	IV2	VideoMamba
LR	1e-5	5e-2	1e-5	7e-4	1e-6	1e-2	7e-3	6e-5	3e-4
Epoch	1	20	20	20	20	20	20	20	20
Warmup Epochs	0	0	0	2	0	0	2	2	2
Batch Size	1	16	8	1	8	16	2	4	4
Optimizer	AdamW	SGD	Adamw	Adamw	Adamw	SGD	Adamw	Adamw	AdamW
Drop Path Rate	0	0.5	0	0.1	0	0.5	0.1	0.3	0.5

Table 2. **Summary of Hyperparameters for Different Models**. **TSF**: TimeSformer, **UMT**: UnmaskedTeacher, **IV2**: InterVideo2

Skill	Test Acc. (%)
Shooting	
Free Throw	27.76
2-PTs	27.62
3-PTs	24.14
Contested-shots	26.62
Overall Shooting	26.19
Average	26.47
Rebounding	
Rebounds	27.11
Defensive	26.85
Offensive	25.89
Average	26.62
Defense	
Steals	24.76
Fouls	25.79
Points-allowed	27.76
Defensive Consistency	31.20
Average	27.38
Playmaking	
Assists	25.89
Passing Accuracy	24.69
Turnovers	25.59
Average	25.39
Offense	
Contribution	35.09
Offensive Consistency	34.34
Teamwork	30.17
Impact	37.24
Efficiency	36.14
Average	34.60

Table 3. **Accuracy Breakdown Across Skills.** We take our best-performing VideoMamba and breakdown its accuracy on the twenty fine-grained. We observe that the average accuracy for the four coarse skill categories of shooting, rebounding, defense, and playmaking are relatively similar. Additionally, we observe that the accuracy for offensive skills is slightly higher, suggesting that these skills might be easier to predict.