

動画に基づく教本参照型 コーチングエージェントの構築

川田 拓朗¹, 藤若 雅也², JI XIAOTONG², 劉 健全²

1. 法政大学, 2. 日本電気株式会社

はじめに

コーチングとは？

- 観測された対象者の行動を参照基準を比較し、その差分に基づき、目標達成へ向けて改善を促す教育的支援 [[Atkinson+, Eur J Pediatr'22](#)]
- 近年の LLM, VLM の発展に伴い、自動コーチング AI の構築が期待される (スポーツ・運転・料理・製造現場・建設現場・医療現場)

コーチング AI の現状と課題

- スポーツ指導の分野において、学習者の動画を入力とするコーチング AI の研究が盛ん
- 既存手法は大きく2種類に分類される:
 1. LLM の事前知識を参照基準とするコーチング [[Khasentino+, Nature Medicine'25](#)]
 - ✓ Zero-shot, Few-shot で様々な技能に適用可能
 - ✗ 根拠が曖昧. ハルシネーションの恐れ
 2. お手本動画を参照基準とするコーチング [[Ashutosh+, CVPR'25](#); [Yeh+, ACL'25](#)]
 - ✓ 動作差分を直接比較可能
 - ✗ 専門技能では、お手本動画を用意するのは困難

はじめに

参照基準として書籍・教本に着目

- 技能習得に有用な、テクニクや手順、ノウハウが集積されている
- 多様な技能に対応
- 大規模に収集可能

ユニット
02

基本姿勢と ボールコントロール

プレーの土台となる姿勢と、ボールの持ち方、パスの準備姿勢を身につけよう。

4つのステップで覚えよう!

1 まっすぐ立つ

足を肩幅くらいに開き、つま先は少し外へ。背すじをのぞいて、肩は首を見る。

2 ひざを曲げて腰を落とす

ひざを軽く曲げて、腰を落とす。重心は少し前に。すばやく動く姿勢をつくる。

3 ボールを胸の前で持つ

両手でボールを持ち、胸の前に置く。手のひらはボールの側面に、指は後ろにかかえる。

4 パスの準備姿勢

進む方向に体を向け、相手や味方を見る。ボールは胸の前でしっかりキープ。

1 はじめに意識したいこと

- リラックスして、力を入れすぎない。
- 常にまわりを見て、次のプレーを考えよう。

姿勢のポイントをチェック!

- 目線** まっすぐ前を見て、膝や両手の動きをチェック。
- ボールの位置** 胸の前でしっかり持つ。ひざは自然に下げ、おく。
- ひざの向き** ひざもまっすぐで、足の近くで安定させる。
- ひざの角度** 軽く曲げて、いつでも動くようにする。
- 足の向き** 肩幅くらいに開き、つま先は少し外へ。

ポイント

- ✓ ボールを落とさないように両手でしっかり持つ。
- ✓ 体の中心でボールをコントロールする。
- ✓ いつでもパスやドリブルにつながる姿勢をつくる。

よくあるミス

- ✗ ボールが膝から離れている
- ✗ コントロールがうまくいかない
- ✗ ひざが伸びたまま

練習のコツ

- 鏡の前や動画で、自分の姿勢をチェックしてみよう。
- 壁に向かってパスをして、フォームを確認しよう。
- 家で、姿勢を崩さずボールを持つ練習を続けよう。

やってみよう!

友だちや家族と向かい合い、1〜2m離れたパスを10回つなげてみよう。

第2章

運転の基本を身につけよう

基本姿勢とハンドル操作

安全運転の第一歩は、正しい姿勢と確実なハンドル操作から。落ちて、ひとつずつ身につけよう。

4つのステップで覚えよう!

1 シートの位置を合わせる

ペダルがしっかり踏め、ひざが少し曲がる位置に調整します。前後の動きが確認できるようにしましょう。

2 背もたれと視線を調整する

背中が自然に背もたれにつき角度に調整します。目線は遠くの信号や標識が見える高さを目安です。

3 ミラーを確認する

ルームミラーとドアミラーを調整し、後方が見えやすいようにします。死角がないようにしましょう。

4 ハンドルを正しく握る

9時15分位置を目安に握ります。親指はハンドルの内側に軽く添え、力を入れすぎないようにします。

正しい運転姿勢のポイント

- 目線** 遠くの信号や標識が自然に見える高さを保ちます。
- 背もたれの角度** 背中がしっかりとつく角度に調整。少し後ろに傾けるのがコツです。
- ハンドルの位置** 胸のやや下になる位置が普通。腕が自然に曲がる高さになります。
- ひざのゆとり** ひざが軽く曲がり、楽に操作できるゆとりをもちます。
- 足の置き方** フレームを踏んではき、ひざが伸びず、ペダルをしっかり踏みぬくこと。
- シートベルト** 肩にしっかりとかけ、腰ベルトは骨盤にかかるように着用します。

ポイント

- 力を抜いてリラックスすると、操作がスムーズになります。
- 運転前に姿勢を整えるだけで、疲れにくく、安全性も高まります。
- 正確な姿勢を見直すことが大切です。

よくあるミス

- △ 背もたれを倒しすぎて前方が見えにくい。
- △ ハンドルが遠く、腕が伸びすぎてしまう。
- △ ミラーの角度が合っていないと、後方が見えにくい。

練習のコツ

- 駐車場などで、姿勢を整える練習を始めよう。
- 鏡やスマホのカメラで姿勢を確認すると効果的です。
- 教習所職員にチェックしてもらいながら練習しましょう。

やってみよう! 姿勢チェック エクササイズ

エンジェルを対して、次のことを確認してみましょう。

① シート調整 ② 背もたれの角度 ③ ミラー調整 ④ ハンドルの握り方

チェックしたら、気づいたことをメモしてきましょう。

42

43

3

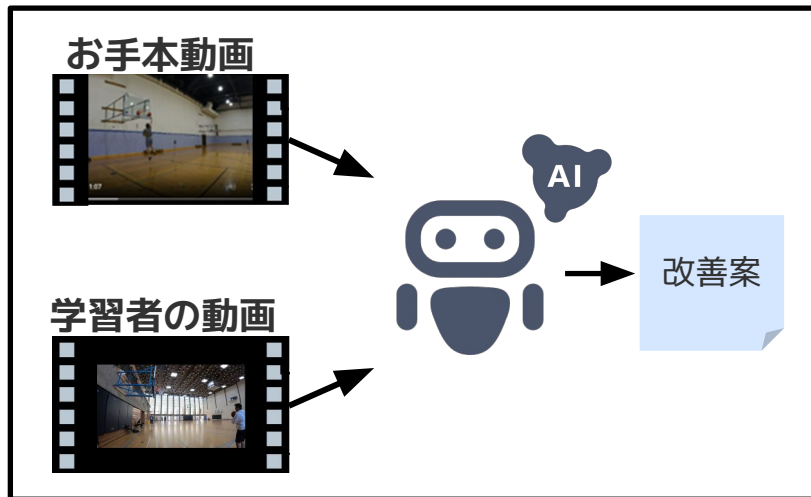
はじめに

参照基準として書籍・教本に着目

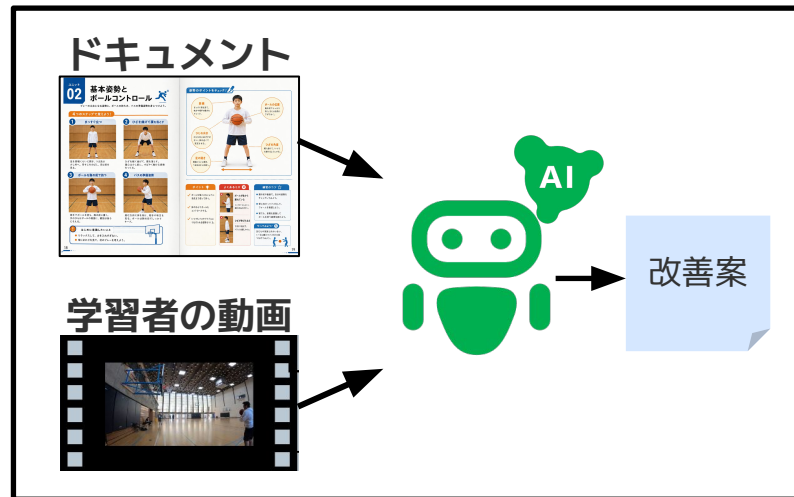
- 技能習得に有用な、テクニックや手順, ノウハウが集積されている
- 多様な技能に対応
- 大規模に収集可能

→ **教本を参照基準として用いたコーチング AI を提案**

従来のコーチングAI

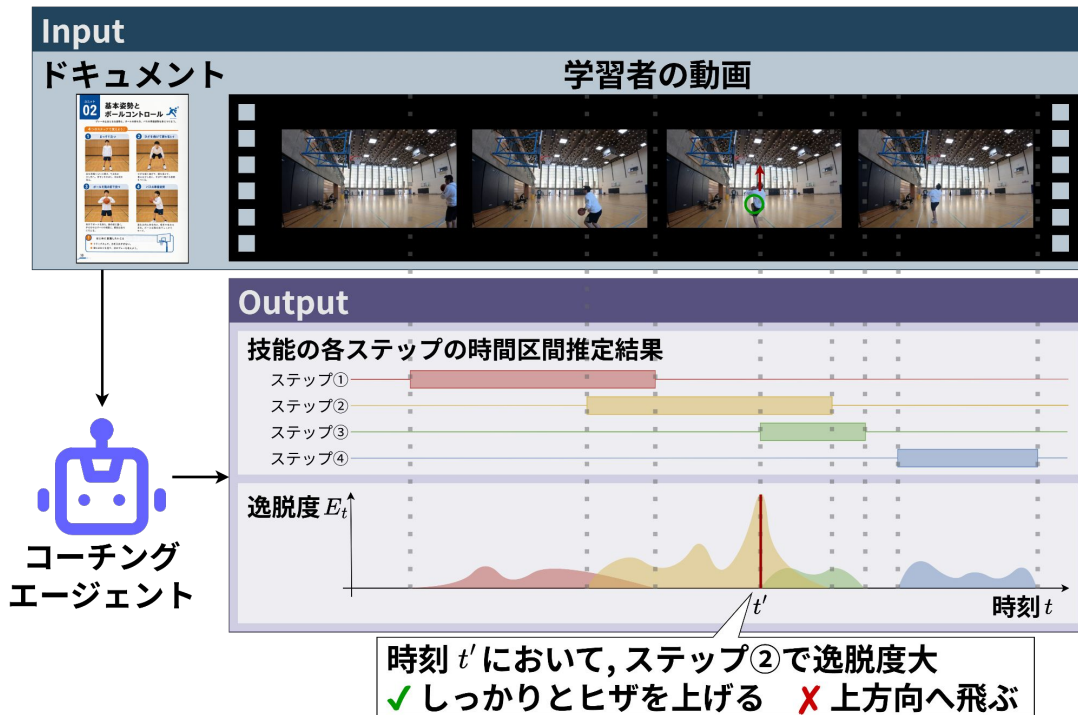


提案手法



提案手法

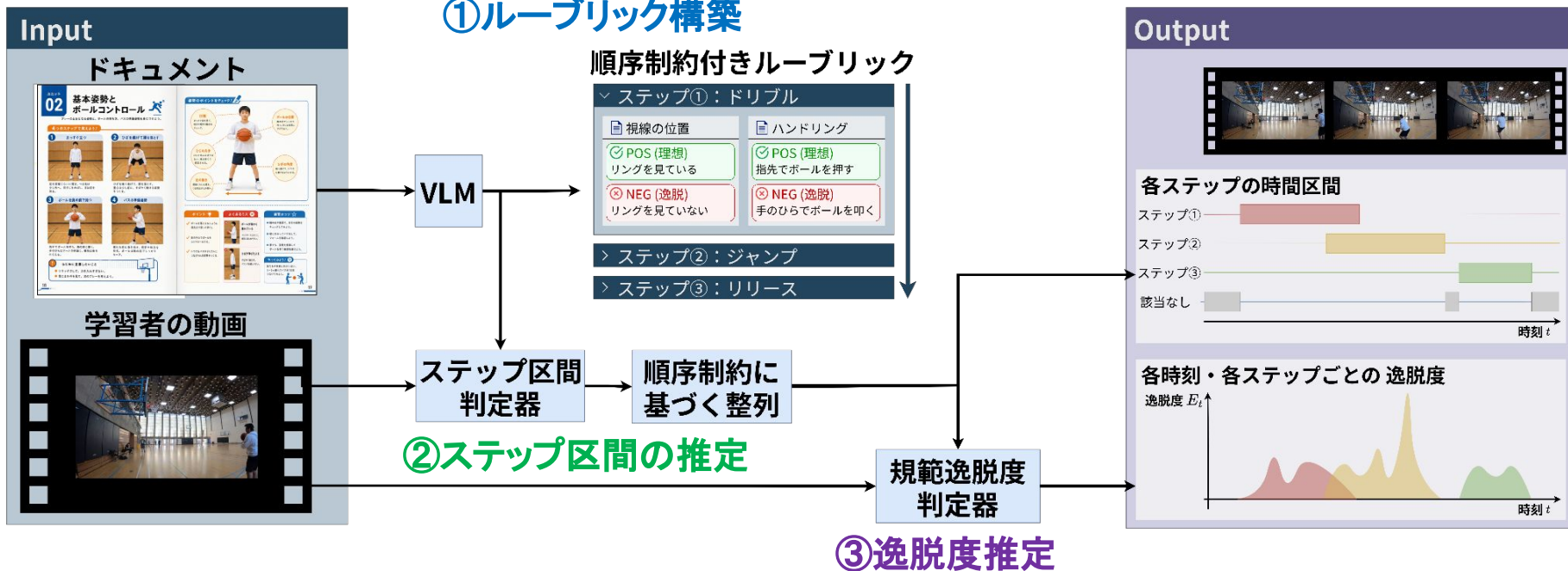
- 教本と動画を入力とし、「いつ・何を・どの程度」逸脱しているかを可視化
- お手本動画や追加学習を必要とせず、様々な技能にスケール可能



提案手法

1. ルーブリック構築 : 教本を評価可能な構造化データに変換. 技能をステップに分解
2. ステップ区間の推定: 各ステップの時間区間を検索 (Video Moment Retrieval)
3. 逸脱度推定 : 各時刻において, 対応するステップの規範項目を満たすかを判定

①ルーブリック構築



提案手法 | ① ループリック構築

1. **ステップ抽出:** VLM を用い、**教本の説明文を手順ステップごとに分解・構造化**
 - ドキュメント内の記述を原文のまま抽出し、知識の追加や編集は行わない
2. **極性付与・補完:** 各ステップの**説明文を規範項目に変換**
 - 規範項目は補集合となる 2 つの極性が異なる命題から成る
 - ① POS: 満たすべき状態 ② NEG: 満たすべきでない状態 (禁止事項など)
 - どちらか一方の極性の命題のみが記載されている場合は、VLM で補完する

ドキュメント



VLM

順序制約付きループリック



提案手法 | ② ステップ区間の推定

- 各ステップが動画のどの区間に対応するかを推定する Video Moment Retrieval タスク
 - クエリとなるステップは以下の特性・制約を持つ：
 - テキストの説明が曖昧で、図やイラストでの理解を前提としている場合がある
 - 技能特有の動作や、局所的で画面上では小さな動作を含む
 - 順序の時間的単調性が存在する
- Zero-shot で従来の Video Moment Retrieval モデルの適用は困難



提案手法 | ② ステップ区間の推定

1. **関連度行列の構築**: VLM を用い, 各時刻と各ステップの関連度行列を構築
 - クリップ, 教本スクリーンショットを入力し, 該当ステップを答える QA を解く
 - 1トークン目に各選択肢が出力される連続的な事後確率を関連度とする

1. 関連度行列の構築

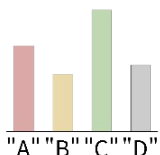
プロンプト
動画クリップ ドキュメント



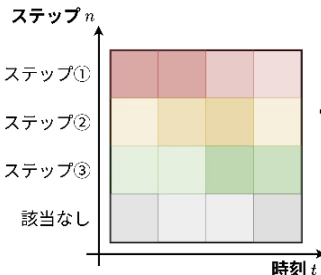
質問文 + 選択肢
行動はどのステップに属しますか？
(A) ステップ①：ドリブル
(B) ステップ②：ジャンプ
(C) ステップ③：リリース
(E) 該当なし

VLM
as a Classifier

各選択肢の
事後確率分布

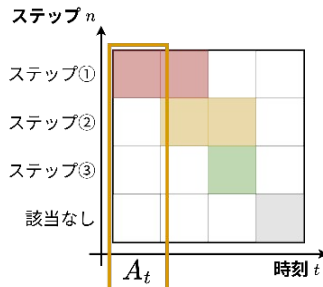


連続的な
関連度行列 P



2. 順序制約に基づく整理

離散的な
区間の割り当て A



提案手法 | ② ステップ区間の推定

2. 順序制約に基づく整列: 各時刻に解釈性の高い離散的なステップ割り当てを行う

- 時刻 t においてアクティブなステップ番号の集合 A_t を最適化
- ここで, 以下の制約を課す

$$\max A_t \leq \min A_{t+1}$$

各ステップは
単調に進行する

$$|A_t \cap A_{t+1}| \leq d$$

多数のステップが長時間にわたって,
同時に実行されることはない. 本研究では, $d = 1$

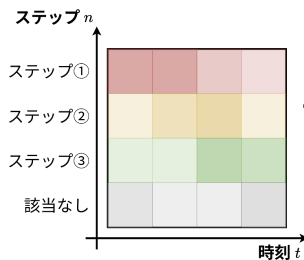
- ここで, 時刻 t におけるスコア ϕ_t を次のように定義する

$$\phi_t = \max_{n \in A_t} p_{t,n}$$

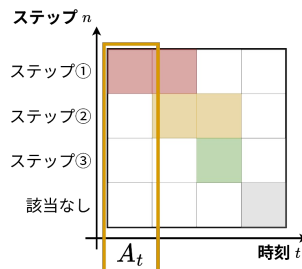
- 以下のスコア最大化問題を解き, 最適化

$$A^* = \operatorname{argmax}_A \sum_{t=1}^T \phi_t.$$

連続的な
関連度行列 P



離散的な
区間の割り当て A

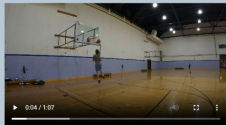


提案手法 | ③ 逸脱度推定

- 各時刻に割り当てられたステップの規範項目を学習者が満たすか判定する
- ステップ区間推定と同様に, VLM で QA を解き, 逸脱度を連続値で推定

プロンプト

動画クリップ

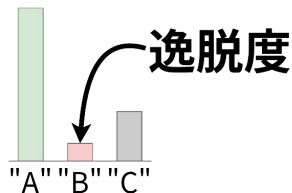


質問文 + 選択肢

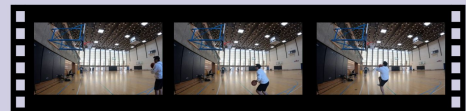
選手はどの状態ですか？
(A) リングを見ている
(B) リングを見ていない
(C) 判別不可能

VLM
as a Classifier

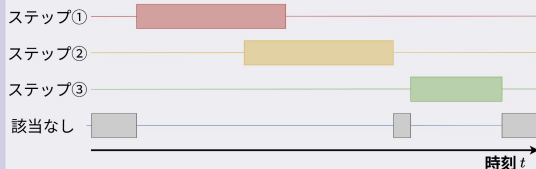
各選択肢の
事後確率分布



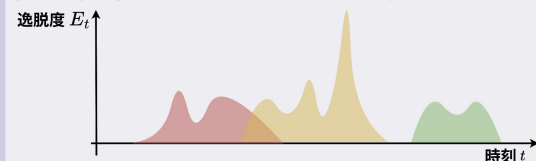
Output



各ステップの時間区間



各時刻・各ステップごとの逸脱度



実験

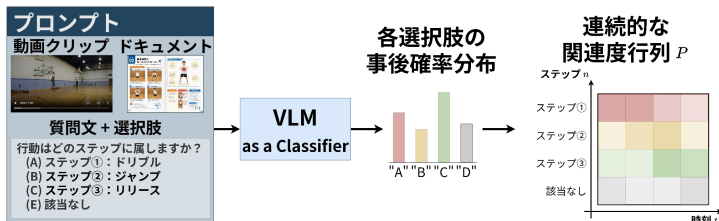
- 熟練度の異なるバスケの練習動画を対象に, 提案手法で逸脱度を可視化
 - Ego-Exo4D データセットのバスケの三人称動画 280 件 [[Grauman+, CVPR'24](#)]
 - バスケの教本『見るだけでうまくなる! バasketballの基礎』 [[森圭司, BBM'20](#)]
- 各処理における VLM は GPT-5.2 [[OpenAI, 2025](#)] を使用

ステップ区間推定のベースライン

- ステップ区間推定モデルとして, 既存の Video Moment Retrieval を用い, 比較
 - InternVideo2 [[Qang+, ECCV'24](#)]
 - R²-Tuning [[Liu+, ECCV'24](#)]

入力モダリティのアブレーション

- ステップ区間推定を行う際に, スクリーンショットの入力有りと無しで比較



実験

ループリック構築の妥当性評価

- 以下の観点に基づき、著者による人手評価を実施
 - 原文忠実性: 原文のまま文章を抽出できているか
 - ステップ混在の不存在: 異なるステップの文章が混在していないか
 - 手順的曖昧性の不存在: 手順分割が曖昧でないか
 - 意味逸脱の不存在: 各規範項目が教本と整合しているか
 - 極性の妥当性: 各規範項目の極性が教本と整合しているか
 - 動画判定可能性: 各規範項目が動画から判定可能か

ステップ区間推定の妥当性評価

- 各ステップと関連付けられた動画の正解時間区間と、推定された区間のIoUで評価

逸脱度推定の妥当性評価

- 熟練者と初心者の動画それぞれに対し、推定された逸脱度の分布の差を確認
 - 初心者の方が逸脱度が大きい場合、熟練度を反映していることを示唆する

結果

構築されたルーブリックの妥当性

評価する処理段階	評価観点	妥当な項目の割合
ステップ抽出	原文忠実性	1.000
	ステップ混在の不存在	1.000
	手順的曖昧性の不存在	1.000
極性付与・補完	意味逸脱の不存在	0.967
	極性の妥当性	1.000
	動画判定可能性	0.833

- 教本の記述に忠実な形で、手順と評価規範をルーブリックとして構造化できている
- 一方、「しっかり、○○する」「○○なイメージ」のような、曖昧な規範も存在した
 - 原文が曖昧な場合、Yes/Noで判定できない命題を生成してしまう

結果

ステップ区間推定手法ごとの, ステップ区間推定能・熟練度ごとの逸脱度の比較

ステップ区間推定手法	入力情報		ステップ区間 IoU	逸脱度の平均		
	説明テキスト	スクリーンショット		熟練者	初心者	Δ (初心者 - 熟練者)
InternVideo2 [8]	✓	✗	0.199 ± 0.157	2.704 ± 1.019	2.748 ± 1.007	0.044
R^2 -Tuning [5]	✓	✗	0.230 ± 0.095	6.400 ± 1.370	6.590 ± 1.378	0.190
ours (GPT-5.2 [6])	✓	✗	0.312 ± 0.154	0.732 ± 0.400	0.848 ± 0.321	0.116
	✓	✓	0.415 ± 0.230	0.695 ± 0.388	0.928 ± 0.440	0.233

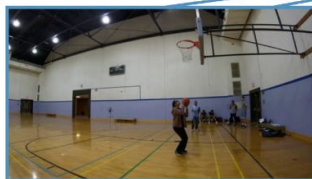
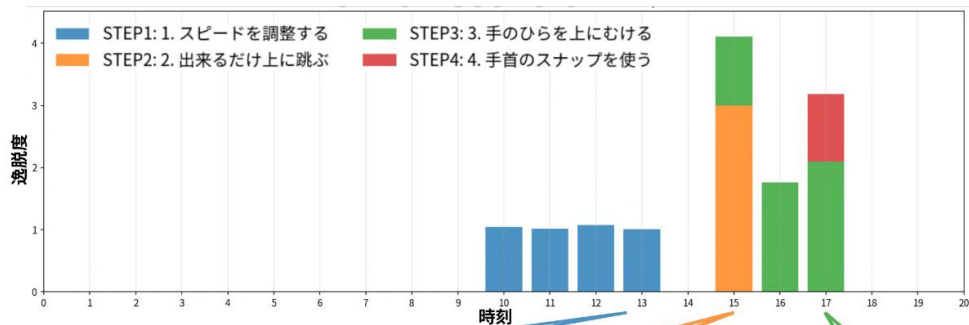
ステップ区間推定の性能

- スクリーンショットから図やレイアウトなどのリッチな文脈を活用できることを確認
- 教本をクエリとする設定において, VLM を分類器として定式化する有効性が示された

熟練度と逸脱度

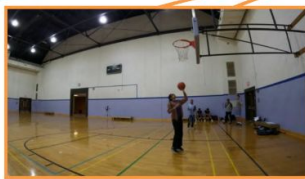
- 初心者に対し, 一貫して高い逸脱度を示し, その妥当性が示唆された
- ステップ推定が上手くいかないと, 逸脱度が大きくなった

結果



ステップ1：
スピードを調整する

- ✓ ボールを持って前進する際、歩数は2歩以内に収めている。
- ✓ スピードを適切に制御している。
- ✗ 片足で踏み切り、反対側の足を曲げて振り上げている。



ステップ2：
できるだけ高く飛ぶ

- ✗ 2歩目で十分な勢いをつけている。
- ✓ 膝をしっかりと曲げている。
- ✗ 上にまっすぐ高く飛んでいる。



ステップ3：
手のひらを上に向ける

- ✗ 手のひらが上を向いている。
- ✓ 手のひらでボールを持ち上げている。
- ✗ 腕が伸びている。

← 人手評価による、
満たしている規範と
満たしていない規範

- 逸脱度が高い時刻では、満たしていない規範の数が多い
- 解釈可能性の高い逸脱度を提供

結論

- 本研究では、教本を参照基準とし、学習者動画中の行動が「いつ・何を・どの程度」逸脱しているかを推定するコーチングエージェントを提案
 - 構築されたループリックは、教本の記述に忠実かつ評価可能な形で構造化できている
 - 既存の Video Moment Retrieval 手法と比較して高いステップ区間推定精度を実現
 - 推定された逸脱度は、熟練度の違いを一貫して反映する指標として機能した
- **様々な技能に拡張可能なコーチング支援の新たな枠組みを示した**

今後の展望

- バスケット以外の技能, 多様な書籍・教本への拡張
- 専門家による逸脱度の妥当性検証
- 各ステップの単調性が保たれない場合を明示的に検知する手法の検討
 - 順序が入れ替わった場合, 飛ばした場合