

教本に基づいて学習者の動画に対してコーチングを行うAI

1. 研究背景・概要

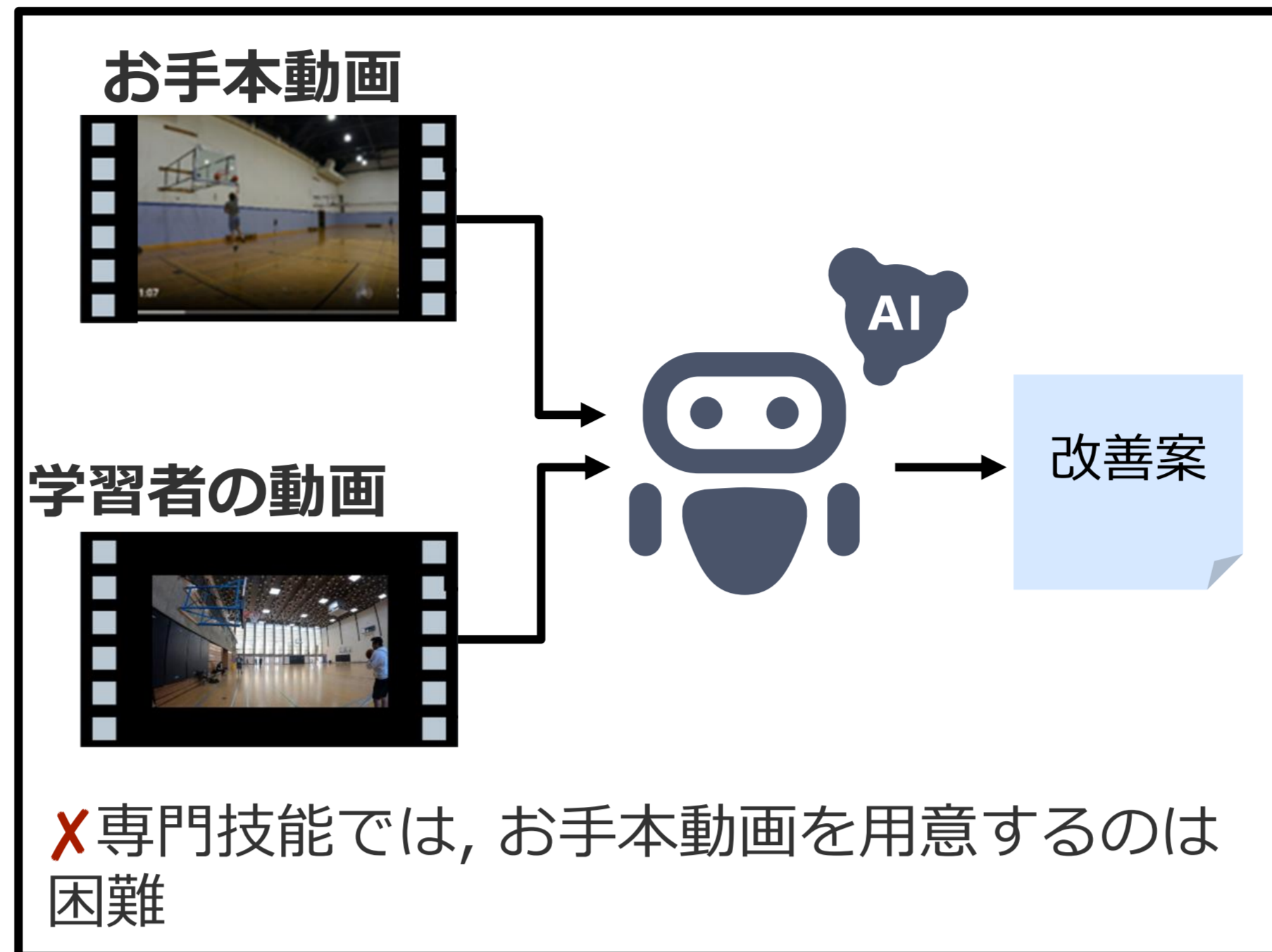
コーチングとは？

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 [1]

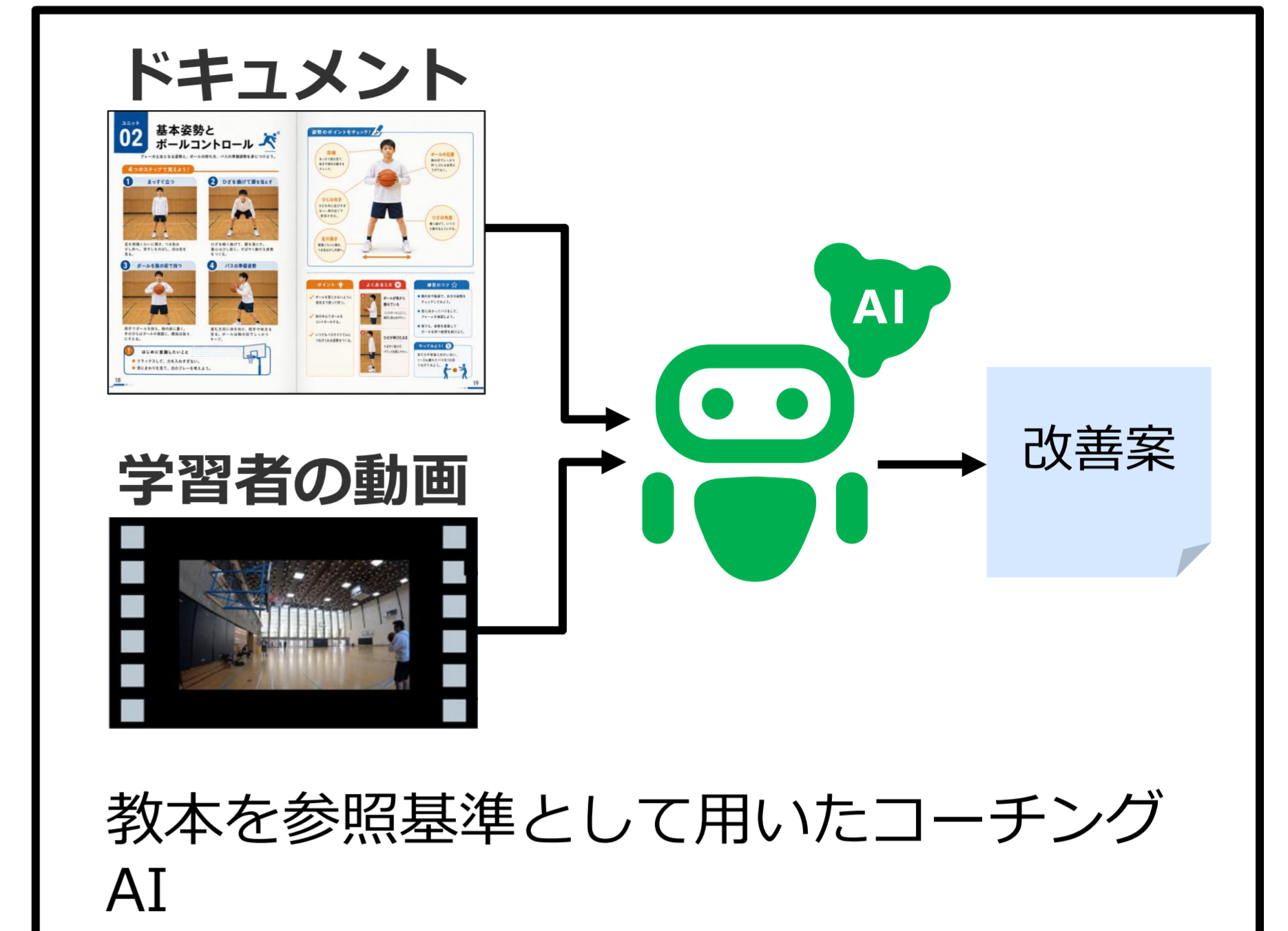
従来のコーチング AIの課題

- LLM の事前知識を参照基準とするコーチング [2]
- ✓ 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
- お手本動画を参照基準とするコーチング [3, 4]
- ✓ 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

従来のコーチングAI

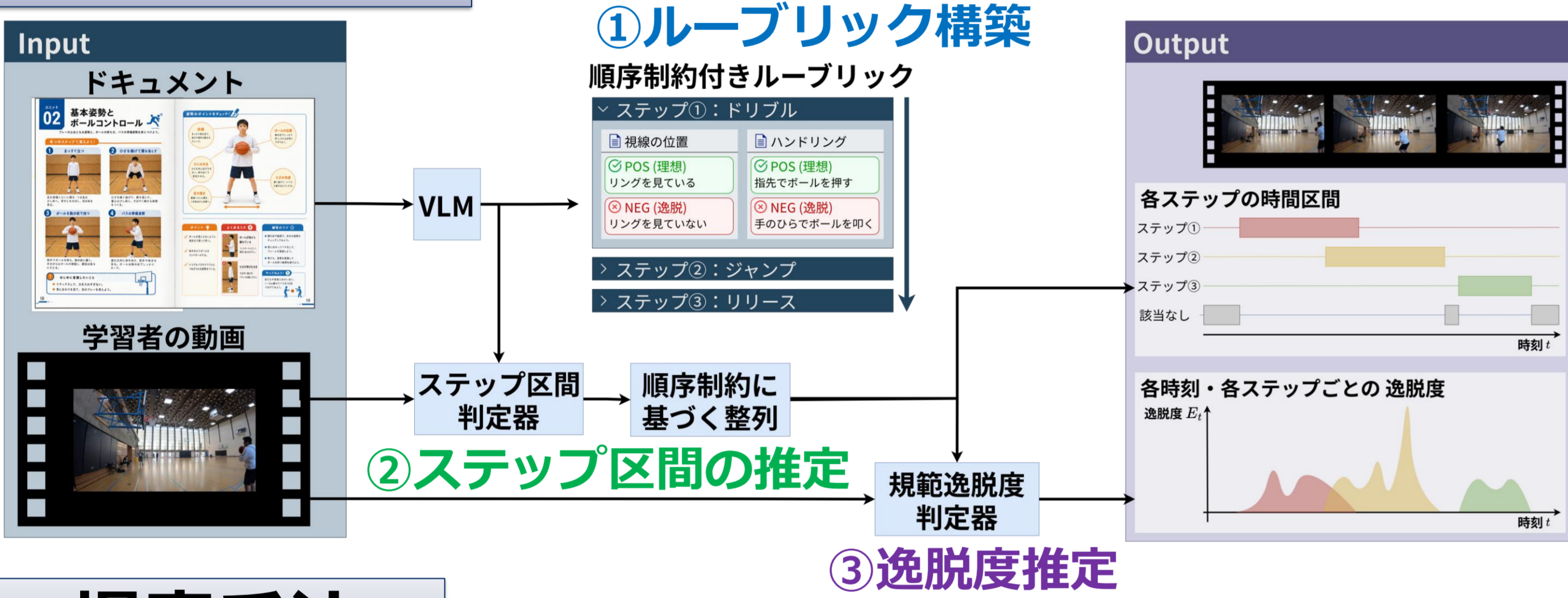


提案手法



2. 提案手法

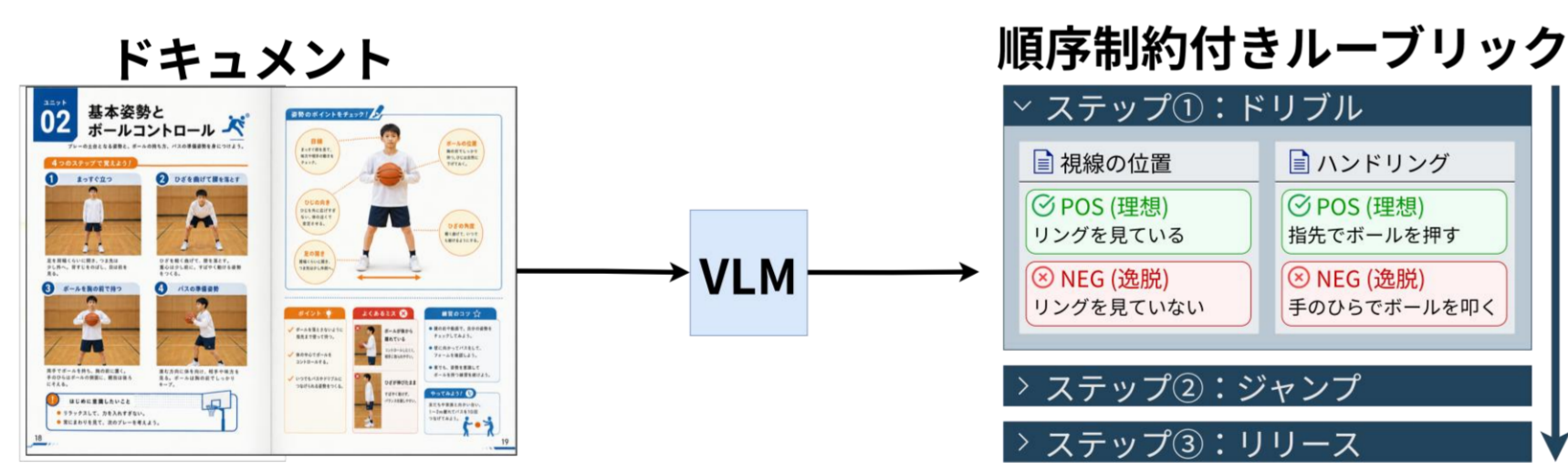
概要



提案手法

①ループリック構築：教本の分解・構造化

- ✓ **ステップ抽出**
 - VLMで教本を手順ステップ、各ステップの規範項目に分解, 構造化
 - ドキュメント内の記述を原文のまま抽出, 知識の追加や編集は行わない
- ✓ **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
 - POS: 満たすべき状態
 - 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



②ステップ区間の推定：各ステップの時間区間を検索

- ✓ **関連度行列の構築**: VLM で各時刻と各ステップの関連度行列を構築
 - クリップ、教本 (画像) を入力, 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
- ✓ **順序制約に基づく整列**: 各時刻に解釈性の高いステップを割当て
 - 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

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

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

- 時刻 t におけるスコア ϕ_t を次のように定義

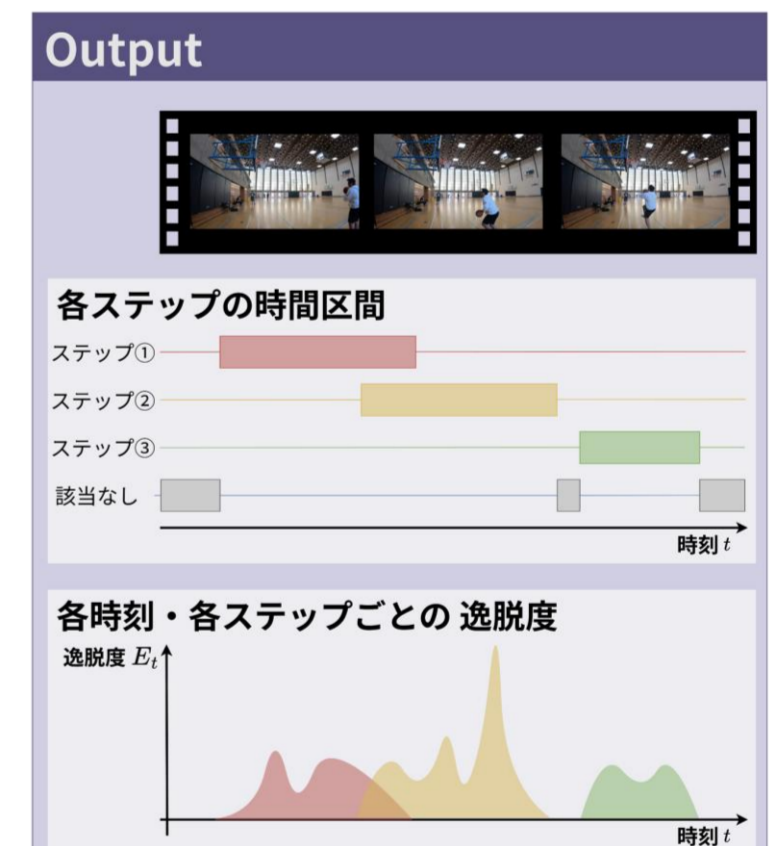
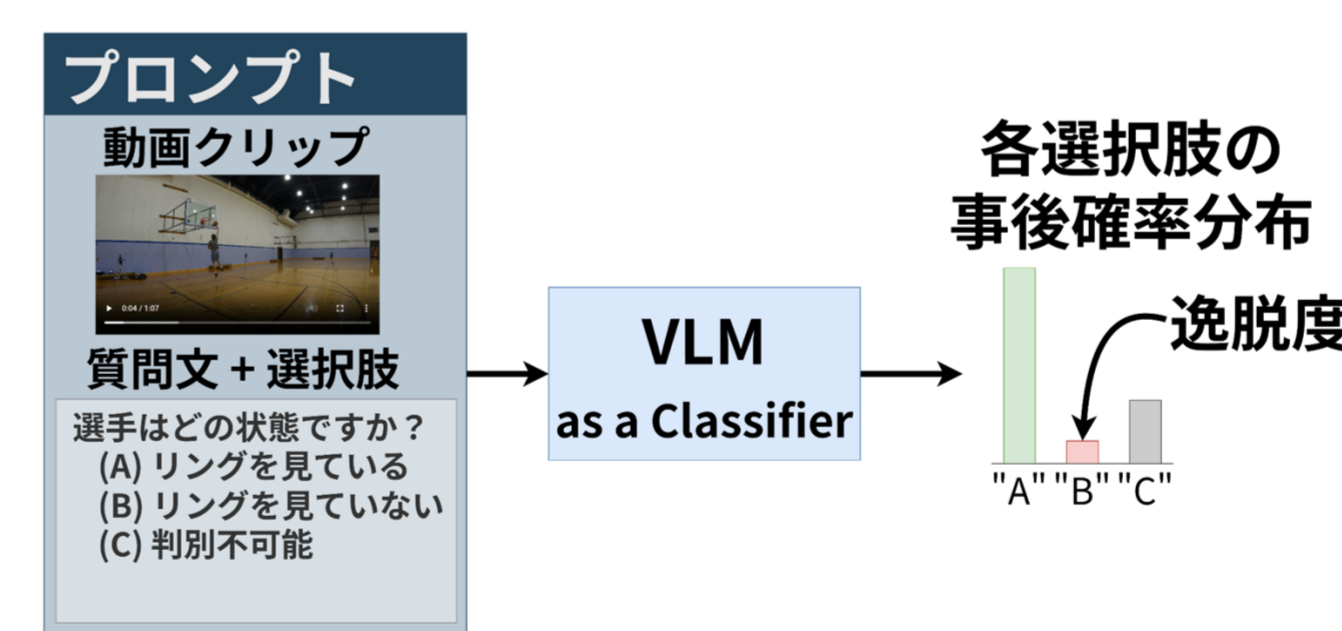
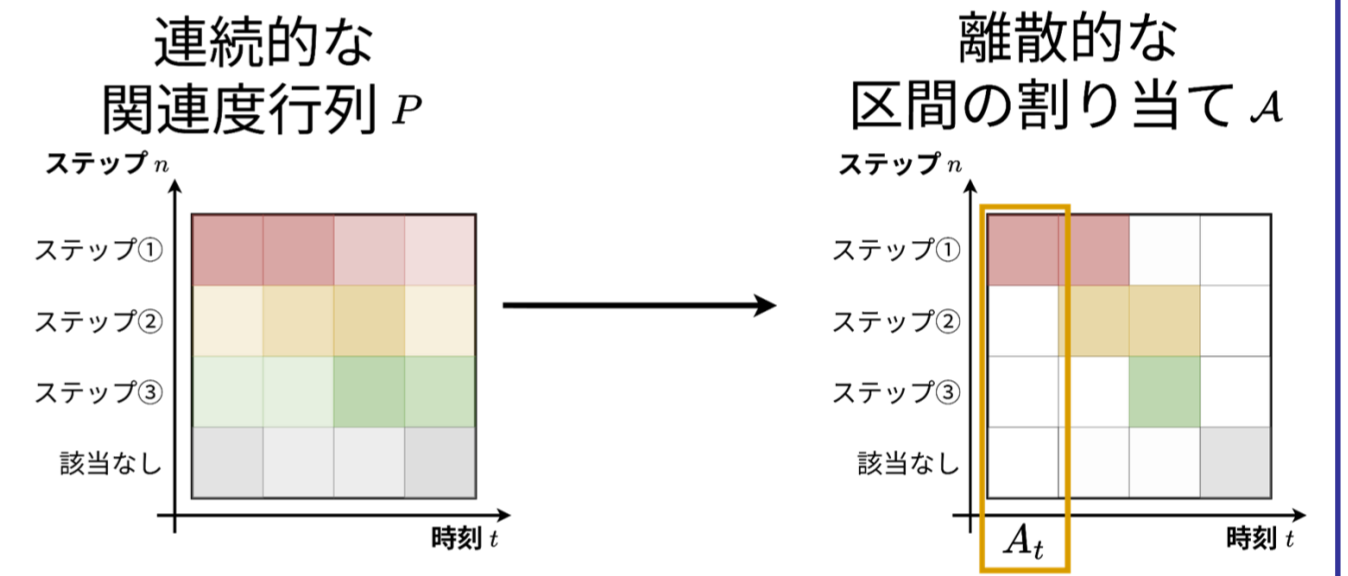
$$\phi_t = \max_{n \in A_t} p_{t,n}$$

- 以下のスコア最大化問題を解き, 最適化

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

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

- 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を判定する
- 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



3. 実験

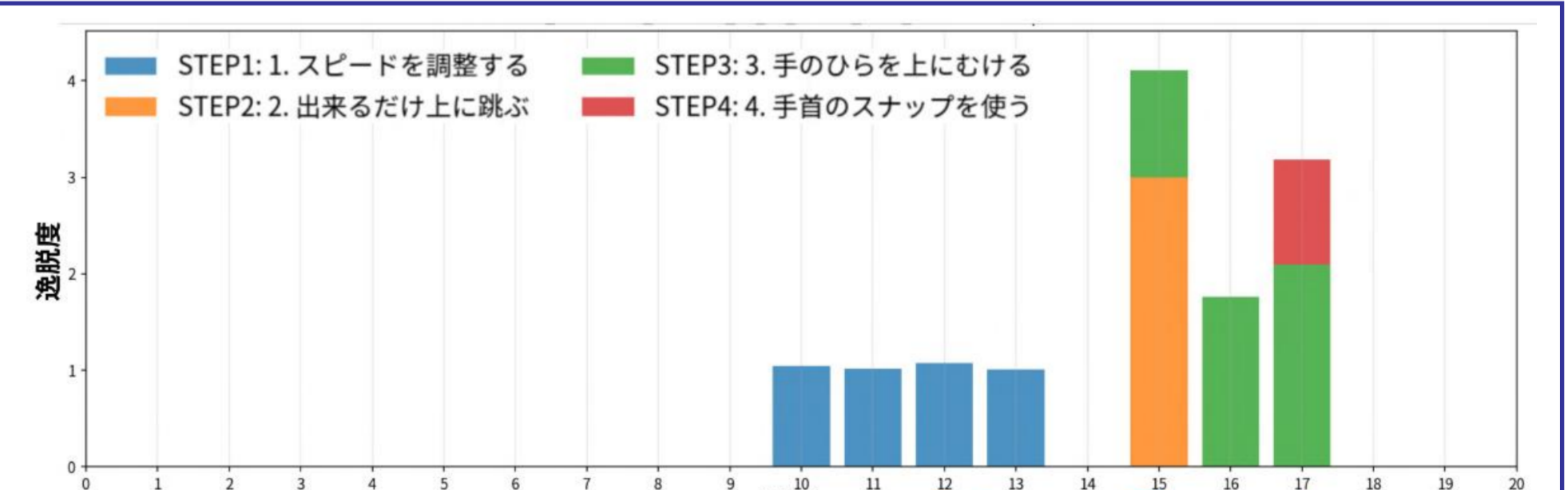
- 熟練度の異なるバスケの練習動画を対象に, 提案手法で逸脱度を可視化
 - 学習者動画: Ego-Exo4D データセットのバスケの動画280件 [5]
 - ドキュメント: バスケの教本『見るだけでうまくなる! バスケボールの基礎』 [6]
- 各処理における VLM は GPT-5.2 [2] を使用

構築されたループリックの妥当性

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
	手順の曖昧性の不存在	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

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。

4. 参考文献

[1] Adelle Atkinson et al. "Feedback and coaching." European Journal of Pediatrics, Vol. 181, pp. 441-446, 2022.
 [2] Justin Kshentino et al. "A personal health large language model for sleep and fitness coaching." Nature Medicine, Vol. 31, pp. 3394-3405, 2025.
 [3] Kumar Ashutosh et al. "ExpertAF: Expert actionable feedback from video." In proc. of CVPR, 2025.
 [4] Wei-Hsin Yeh et al. "CoachMe: Decoding Sport Elements with a ReferenceBased Coaching Instruction Generation Model." In proc. of ACL, 2025.
 [5] Kristen Grauman et al. "Ego-Exo4D: Understanding Skilled Human Activity from First- and Third-Person Perspectives." In proc. of CVPR, 2024.
 [6] 森圭司. "目で学ぶシリーズ 3 見るだけでうまくなる! バスケボールの基礎." ベースボール・マガジン社, 第1版. 第1刷, 2020.
 [7] OpenAI. GPT-5, 2025. <https://platform.openai.com/docs/models/gpt-5>.