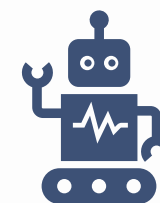


日本語ロボット基盤モデルの開発



RAMEN (Team035)

VLA (Vision-Language-Action) モデルの開発に挑戦

VLA開発の流れ

LLM

VLM

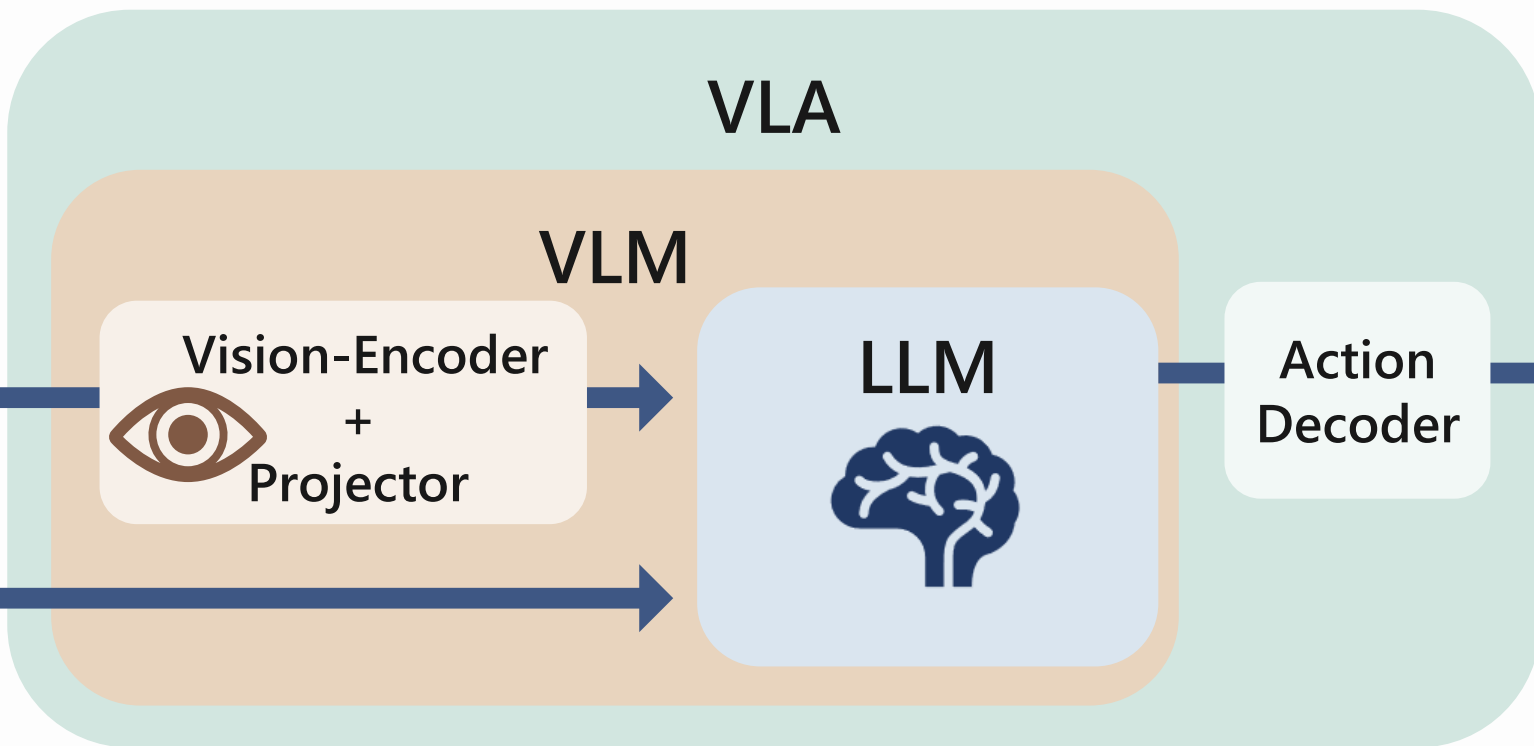
VLA

視覚を追加

行動を出力

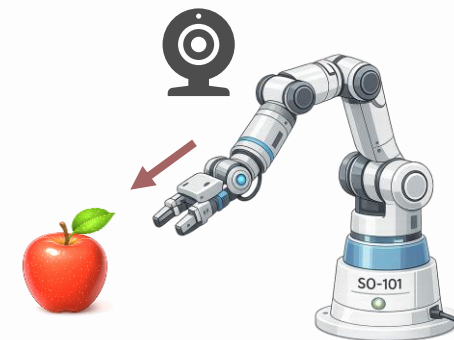


リンゴを掴んで



Actions

グリッパー : Open
手首 : 回転 +0.6
曲げ -0.4



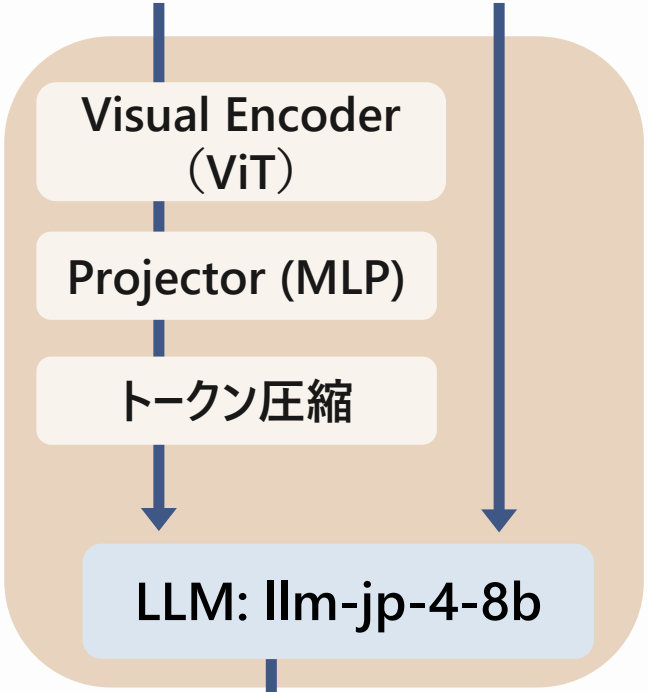
llm-jp-4-8b-instruct と llm-jp-3.1-1.8b-instruct4 をベースに開発
(8BはVLAとしてはモデルが大きいため)

VLM開発：NVILAアーキテクチャを3ステージで学習



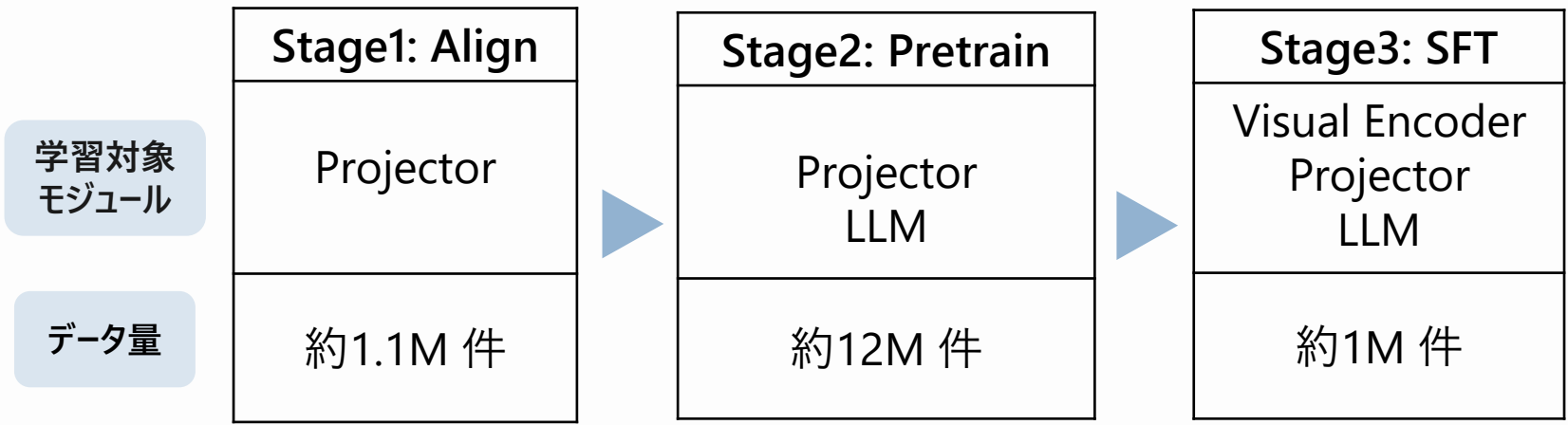
この写真には何が映っていますか？

タイル化

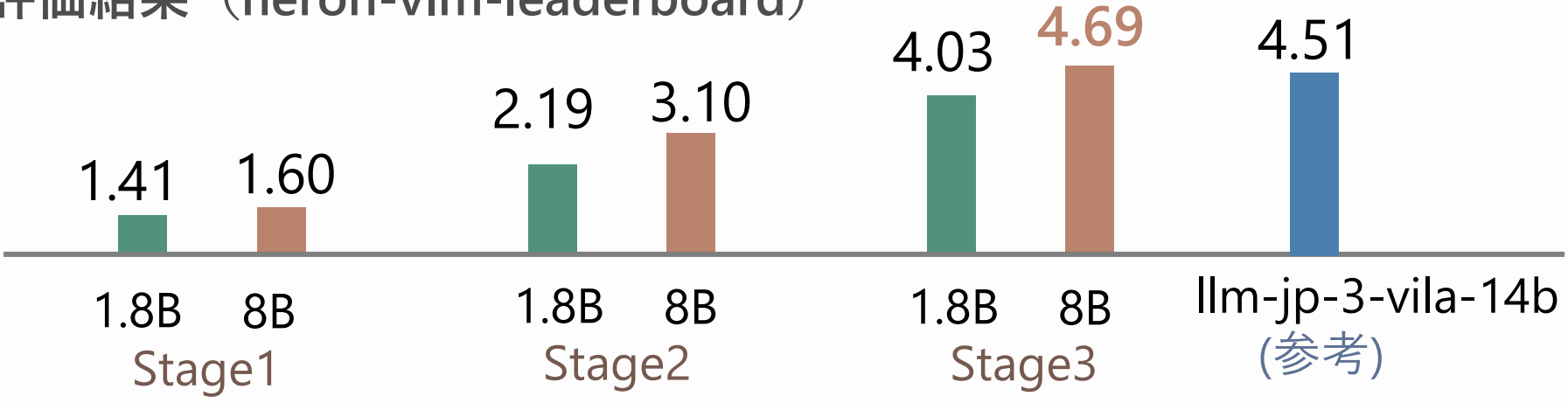


LLM: llm-jp-4-8b

富士山です。



評価結果 (heron-vlm-leaderboard)

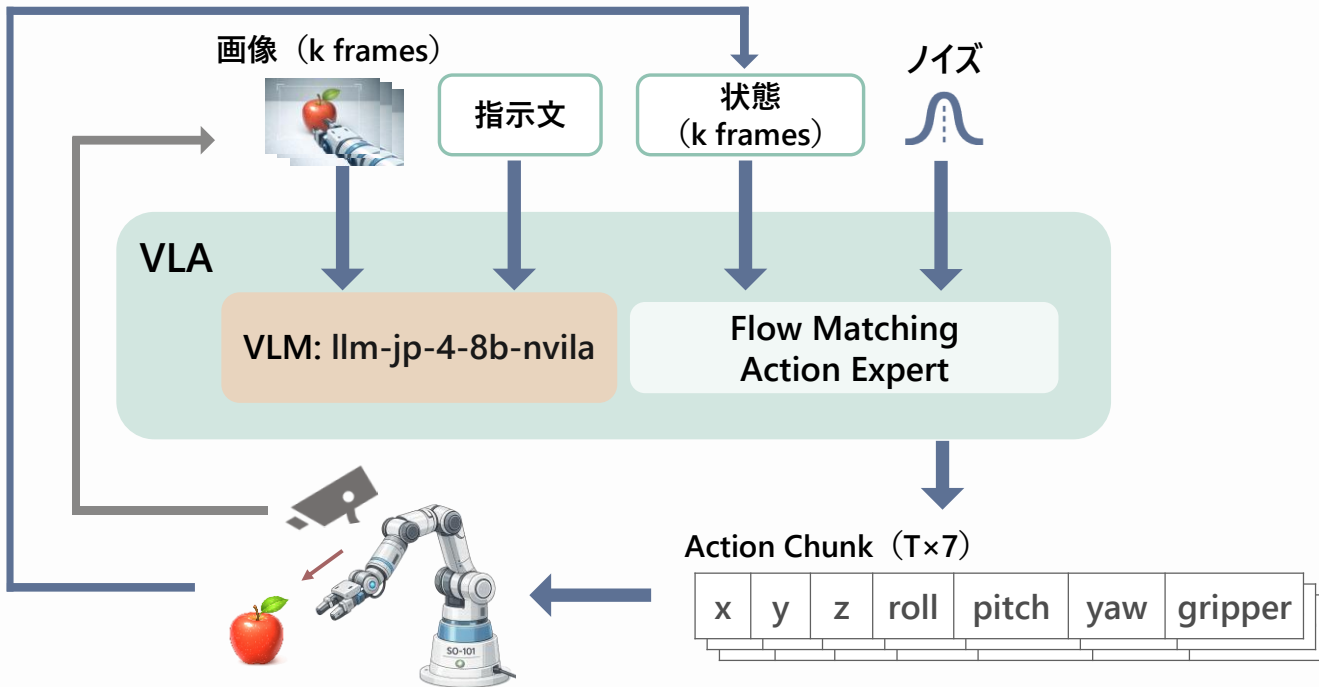


8Bモデルはllm-jp-3-vila-14bと比較して高スコア 3

VLA開発

VILAFlowPolicy

VILA系VLMベースのVLAポリシーをLeRobot上に実装



評価

SimplerEnv によるシミュレータ評価 (10ep)
日本語に翻訳した指示文を入力

タスク	成功率 (1.8B)
google_robot_close_top_drawer	60%
google_robot_open_top_drawer	10%
google_robot_pick_coke_can	20%
google_robot_move_near	20%
widowx_spoon_on_towel	10%
widowx_carrot_on_plate	0%

8Bは失敗：一般的なVLAモデルに比べてサイズが大きく、十分な学習量を確保できなかった？

VLA : 成功タスクの例

pick_coke_can



close_top_drawer



まとめ

- LLM→VLM→VLAまで一連の開発パイプラインを構築
- llm-jp-4-8b-instruct及びllm-jp-3.1-1.8b-instruct4を基盤とした日本語VLAを開発し、1.8Bモデルで複数のタスクに成功
- 今後の展望：性能向上、実機での検証

貴重な機会 & 計算資源の提供 ありがとうございます！

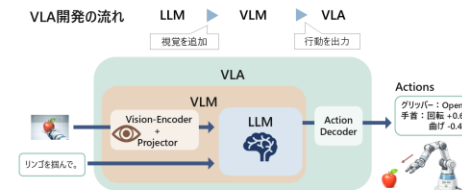
FT-LLM 2026 自由型タスク

日本語ロボット基盤モデルの開発

RAMEN (Team035)

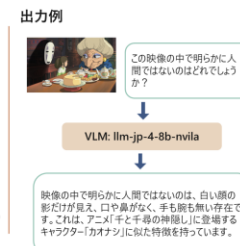
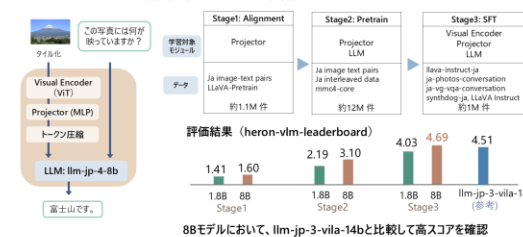
背景

- Vision-Language-Action (VLA) モデル：画像と言語指示を入力とし、ロボットの行動を生成するモデル
- 国内におけるVLA開発の知見はまだ少ない
→ 先進的な取り組みとしてVLA開発に挑戦
・ 知見を共有し、国内におけるVLA開発の促進へつなげる
- 8BはVLAとしてはモデルが大きいため、llm-jp-3.1-1.8b-instruct4についても学習



VLM (Vision-Language-Model) 開発

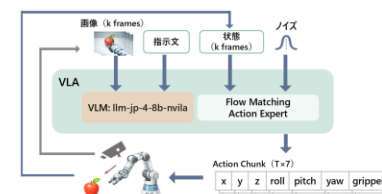
NVILAアーキテクチャを採用し、3ステージで学習



VLA (Vision-Language-Action) 開発

VILAFlowPolicy

VILA系VLMベースのVLAポリシーをLeRobot上に実装



学習

学習対象

- LLM (LoRA)
- Flow Matching Action Expert

以下のデータを日本語に翻訳して使用

- rail-berkeley/bridge_data_v2: Berkeley/Stanford が収集した実機ロボットデータ
- jxu124/OpenX-Embodiment: Google の RT-1 学習用データセット

評価

SimplerEnv によるシミュレータ評価 (10ep)

指示文を日本語に翻訳し、日本語VLAとしての挙動を確認

タスク	成功率 (1.8B)	成功率 (8B)
google_robot_close_top_drawer	60%	0%
google_robot_open_top_drawer	10%	0%
google_robot_pick_coke_can	20%	0%
google_robot_move_near	20%	0%
widowx_spoon_on_towel	10%	0%
widowx_carrot_on_plate	0%	0%

8Bは一般的なVLAモデルに比べてサイズが大きいため、十分な学習量を確保できなかった可能性がある

成功例: close top drawer



まとめ

- LLM→VLM→VLAまで一連の開発パイプラインを構築
- llm-jp-4-8b及びllm-jp-3.1-1.8b-instruct4を基盤とした日本語VLAを開発し、1.8Bモデルで複数のタスクに成功
- 今後の展望：性能向上、実機での検証