

# LLM を用いた UGV の行動決定における 物体検知処理のオフロードによる性能向上

## Performance Improvement of LLM-based UGV Decision-Making via Offloading of Object Detection

○西尾昂世\*, 岡部豪太\*, 田中大介\*, 張山昌論\*

○Kosei Nishio\*, Gota Okabe\*, Daisuke Tanaka\*, Masanori Hariyama\*

\*東北大学

\*Tohoku University.

**キーワード** : 分散型ロボット制御 (Distributed Robot Control), エッジ AI (Edge AI), 言語ベースの意思決定 (Language-Based Decision Making), モバイルロボティクス (Mobile Robotics)

**連絡先** : 〒 980-8579 仙台市青葉区荒巻字青葉 6-3-09 電気・情報系 3 号館 308 号室  
西尾昂世, Tel.: (022)795-7155, E-mail: nishio.kosei.p2@dc.tohoku.ac.jp

### 1. はじめに

近年, 大規模言語モデル (LLM) の急速な発展により, 従来のテキスト生成にとどまらず, 画像や動画の生成・理解を含む幅広いタスクの実行が可能になっている. こうした進展に伴い, 自律的な意思決定を実現する新たなアプローチとして, LLM をロボット制御に応用する研究が大きな注目を集めている. 従来のロボット制御システムがあらかじめプログラムされた命令に依存しているのに対し, LLM ベースのシステムは自然言語による指示を柔軟に解釈し, 実行可能なタスクへと変換することができる<sup>1, 2, 3)</sup>.

Fig 1 に, LLM を用いたロボット制御の一例を示す. このシステムでは, 検出された物体情報に基づいて, LLM を用いてロボットの次の行動を決定する. 一方で, LLM や物体検出のような計算負荷の高いモデルを扱うためには, 膨大

な計算資源が必要となる. そのため, クラウドベースの推論が広く採用されている<sup>1)</sup>. この方式は有効である一方で, 通信遅延, 安定したネットワークインフラへの依存, およびデータプライバシーに関する懸念といったいくつかの問題を抱えている. 動的環境や通信制約下で動作する自律移動ロボットにおいては, 知覚と計画の両方をエッジデバイス上でローカルに実行することで, システム性能の向上が期待できる.

大規模モデルを用いたロボット制御に関する研究は, これまでに数多く報告されている. LLM やマルチモーダルモデルを活用することで, これらの研究は高水準の自然言語指示を解釈し, 実行可能なロボット動作へ変換することを可能にしている.

LLM をロボット制御に適用した代表的な研究として, RT-2 が挙げられる. RT-2 は, 視覚観測および自然言語による指示を入力とし, ロボット

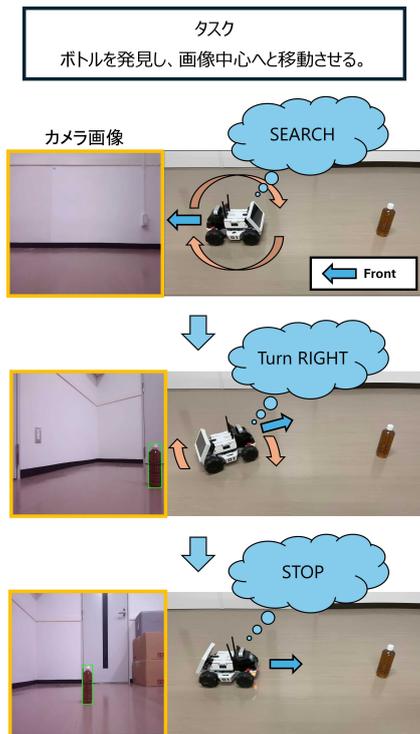


Fig. 1 LLM と物体検出を用いたロボット制御の例

の動作を直接出力することで、これまで見たことのない新規タスクに対しても高い汎化性能を示している<sup>1)</sup>。他の関連研究としては PaLM-E がある。これは、画像やロボットのセンサ観測を、PaLM と呼ばれる LLM の埋め込み空間に統合するマルチモーダルモデルである。インターネット規模の視覚・言語データとロボットタスクデータを用いた共同学習により、PaLM-E は高いデータ効率とロボットタスクにおける優れた汎化性能を実現している<sup>2)</sup>。さらに、SayCan は、大規模言語モデルによって推定されるタスクレベルでの有用性と、ロボットスキルの物理的実行可能性とを確率的に統合することで、抽象的な自然言語指示を実行可能な動作系列へ分解することを可能にしている<sup>3)</sup>。

しかし、これらの手法はいずれも、大規模なデータ収集やモデル学習のために多大な計算資源を必要とする。また、モデル規模が非常に大きくなると、組込みシステムへの展開が困難になる。こうした背景から、比較的軽量なモデル

構成でありながら、自然言語指示に基づく柔軟な意思決定を可能にするロボット制御手法が求められている。

これに対し、我々は先行研究において、計算負荷の高い物体検知処理を別デバイスにオフロードすることで、リソースの限られたエッジ環境でも LLM ベースの意思決定を可能にするヘテロジニアスエッジベース制御システムを提案した<sup>4)</sup>。本システムでは、視覚知覚と LLM に基づく意思決定の両方を、すべてエッジデバイス上で実行する。提案システムは、視覚知覚と言語ベースの行動計画を分離し、計算負荷の高い処理を異なるエッジデバイス間に分散することで、計算資源の競合を緩和する。その結果、知覚と計画を単一のエッジデバイス上で同時実行する場合と比較して、意思決定のためにより大規模な LLM を搭載可能にするとともに、外部サーバへの依存を排除することができる。本稿では、提案する分散型システムの有効性を詳細に検証するため、対象物体の検出から画像中心への誘導に至る一連の動作挙動について詳細な解析を行った結果を報告する。

## 2. 分散型ヘテロジニアスエッジベース制御システム

Fig 2 に、提案システムの処理フローを示す。まず、LIMO に搭載されたカメラから画像を取得し、Renesas RZ/V2H を用いて周囲の物体を検出する。次に、検出された物体のカテゴリおよびその位置情報を、対象タスクの情報とともに LLM に入力し、ロボットの行動を決定する。その後、LLM は高水準の意思決定結果を生成し、決定された行動に基づいて ROS を介して LIMO へ制御入力を送信する。最後に、ROS 上で動作する制御ノードによってモータ制御が実行され、ロボットの移動が実現される。

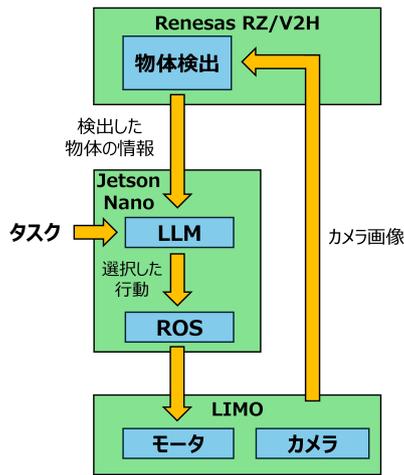


Fig. 2 提案システムの全体構成



Fig. 3 AgileX Robotics LIMO

## 2.1 移動ロボットプラットフォーム

AgileX Robotics LIMO (Fig. 3) は、LiDAR, IMU, 深度カメラなどの複数のセンサを搭載した小型無人地上車両 (UGV) である<sup>5)</sup>。本プラットフォームは Robot Operating System (ROS) をサポートしており、センサデータ取得, 自己位置推定, 経路計画, モータ制御といったロボットの基本機能を統合的に扱うことができる。

## 2.2 言語ベース意思決定モジュール

本研究では、ロボット制御における言語ベースの意思決定モジュールとして、NVIDIA Jetson Nano を用いる。Jetson Nano は、128 基の CUDA コアを備えた GPU と ARM Cortex-A57 プロセッサを搭載する、コンパクトかつ低消費電力なエッジ AI 向け組み込みハードウェアプラッ



Fig. 4 ユリ電気商会 Kakip

トフォームである<sup>6)</sup>。この構成により、自然言語処理のような計算負荷の高いタスクを、エッジデバイス上で直接実行することが可能となる。

## 2.3 視覚ベース知覚モジュール

ロボット周辺環境の認識には、Renesas RZ/V2H に基づく視覚ベース知覚モジュールを用いる。RZ/V2H は、Renesas Electronics が産業機器およびロボティクス用途向けに開発した AI MPU であり、Cortex-A55 や Cortex-R8 コアなどの複数のプロセッサに加えて、独自の AI アクセラレータ (DRP-AI3) を備えている<sup>7)</sup>。このアーキテクチャにより、画像認識に必要な中核計算を低消費電力かつ低遅延で実行できる。本システムでは、この知覚モジュールを、Renesas RZ/V2H を搭載したシングルボードコンピュータ Kakip (Fig. 4) 上に実装する。

## 3. 実装

本節では、提案システムの実装詳細について述べる。本システムは、物体検出を行う視覚ベース知覚モジュールと、その知覚結果に基づいてロボットの行動を決定する意思決定モジュールからなる 2 層構成を採用している。

### 3.1 使用したモデル

視覚ベース知覚モジュールには、物体検出モデルとして事前学習済みの YOLOv6-L<sup>8)</sup> を用い

Table 1 異なる構成における性能の比較

構成	モデル	YOLOv6-L (fps)	LLM (s)	タスク達成度
Jetson Nano + Kakip	<b>Qwen3-1.7B-Q8</b>	<b>8.2</b>	<b>15.1</b>	<b>10/10</b>
(with RZ/V2H)(提案手法)	Qwen3-0.6B-Q8	8.3	10.9	0/10
Jetson Nano only	Qwen3-1.7B-Q8	N/A	N/A	0/10
	Qwen3-0.6B-Q8	0.08	11.1	0/10

た. 言語ベース意思決定モジュールには, 事前学習済み言語モデル Qwen3-0.6B および Qwen3-1.7B の事前 8 ビット量子化版を用いた<sup>9)</sup>.

### 3.2 ハードウェア構成

Jetson Nano と Kakip 間の通信は Wi-Fi を用いた. カメラは Jetson Nano に接続し, 取得した映像ストリームを GStreamer を用いて Kakip へ送信した. Kakip では, 受信した映像ストリームに対して物体検出を行い, その検出結果を UDP 通信により Jetson Nano へ送信した. Jetson Nano は, 受信した検出結果に基づいてロボットの行動を決定した.

### 3.3 対象タスクの仕様

本研究では, 画像平面上で以下の条件を満たすようにボトルを探索することを対象タスクとして定義した. ロボットは, 画像中のボトル中心の水平方向座標  $x$  が概ね 320 px となるように回転する. この条件が満たされたとき, タスクは完了したものとみなす.

### 3.4 LLM プロンプト設計

LLM による自律的な意思決定を可能にするため, 本研究では以下の 4 要素から構成されるプロンプトを設計した.

- **タスク記述:** ロボットが達成すべき目標状態を記述したもの.
- **意味対応付け:** LLM に説明する座標データの物理的意味.

- **現在の状態:** Kakip から送信される最新の物体検出結果 (座標値).
- **行動空間:** 実行可能なロボット行動の一覧 (SEARCH TARGET, GO LEFT, GO RIGHT, STOP).

## 4. 評価

本節では, 提案システムを実機ロボット上で評価し, 各モジュールの処理時間およびタスク実行時の座標遷移に着目して検証を行う. システム性能を評価するため, 対象タスクを 10 回繰り返し行った.

表 1 に, Jetson Nano 単体と Jetson Nano と Kakip を組み合わせた場合における YOLO および Qwen3 の性能比較を示す. Qwen3-0.6B-Q8 モデルを用いた場合は, Jetson Nano + Kakip と Jetson Nano only の両方の構成において, 対象タスクを完了することができなかった. これは, 使用した LLM のパラメータ数が少なく, 与えられたタスクを遂行するための十分な理解能力が得られなかったことに起因すると考えられる. また, Jetson Nano 単体構成では, Qwen3-1.7B-Q8 を用いた場合には, メモリ制約のため両者を同時に実行することができなかった. これに対し, Jetson Nano と Kakip を組み合わせた提案構成では, YOLO と LLM を同時実行し, タスクを完了することができた. これらの結果は, 物体検出と言語ベース意思決定を異なるエッジデバイス間に分散することで, 限られた計算資源においても効果的なロボット制御が可能であることを示している.

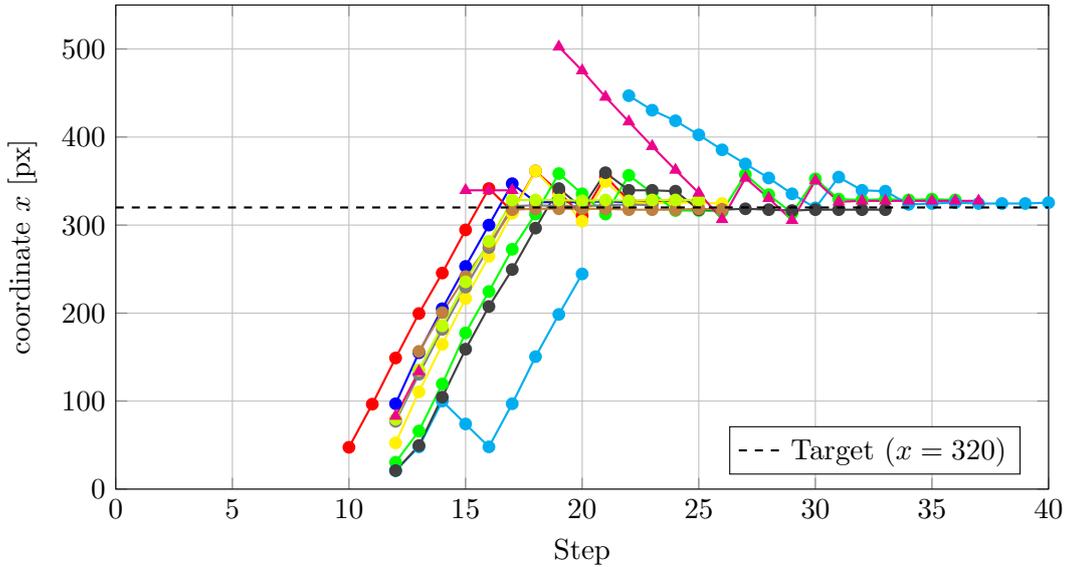


Fig. 5 提案システムによる対象タスク実行中のボトル中心座標  $x$  [px] の遷移

Fig 5 に、Jetson Nano + Kakip 構成において、Qwen3-1.7B-Q8 言語モデルを用いた提案システムによる対象タスク実行中のボトル中心座標  $x$  の遷移を示す ( $x = 320$  は画像中心を示す)。ここでは、例として三角形で示す試行の中心座標プロットの挙動を考察する。Step 0–11 ではボトルが検出されなかったため、 $x$  は未定義 (NaN) となり、データ点は表示されていない。初期状態では、対象が検出されていない場合、LLM は物体を探索するために SEARCH TARGET を選択する。その後、Step 12 でボトルの検出が開始されると ( $x = 83$ )、制御入力によって誤差は徐々に減少し、ボトルが画像中心へ移動していく様子が確認できる。一方で、Step 14 および Step 18 ではデータ点が表示されていないが、これは物体検出がボトルを認識できなかったためであり、これに伴い LLM が SEARCH TARGET を選択したことで、ロボットの動作により  $x$  座標が大きく変化している。座標値が大きく増加した後は、再び目標値に向かい着実に減少し、Step 25 付近で目標値近傍に到達した。Step 26 から Step 30 にかけてはわずかな振動が発生しているものの、Step 31 以降は  $x \approx 326\text{--}328$  付近で安定して推移しており、安定した目標追跡が実現されていることが分かる。

以上の結果から、LLM が SEARCH TARGET や GO LEFT, GO RIGHT といった行動を適切に選択することで対象物体の探索および画像中心への誘導を行い、定義されたタスクを完遂可能であることが示された。

## 5. 結言

本研究では、複数のオンボードエッジデバイスを用いて、LLM に基づく意思決定と視覚知覚を統合した、自律移動ロボット向けの分散型ヘテロジニアスエッジベース制御システムの有効性を詳細に検証した。

実験結果より、視覚知覚を RZ/V2H へ分離することで、物体検出モデルと LLM を同時に動作させた場合でも、計算資源の競合を効果的に緩和しつつ、安定したロボット制御を実現できることを示した。さらに、LLM が与えられたアクションリストから状況に応じた行動を適切に選択し、与えたタスクを完遂できることを示した。

今後は、物体検出および言語推論の双方の処理速度を改善し、意思決定精度を向上させるとともに、より多様なタスクへ対応可能なシステムへ拡張していく予定である。

## 謝辞

本研究は、JSPS 科研費 JP23K28369 の助成を受けたものである。

## 参考文献

- 1) Brianna Zitkovich, Tianhe Yu, Sichun Xu, Peng Xu, Ted Xiao, Fei Xia, Jialin Wu, Paul Wohlhart, Stefan Welker, Ayzaan Wahid, Quan Vuong, Vincent Vanhoucke, Huong Tran, Radu Soricut, Anikait Singh, Jaspiar Singh, Pierre Sermanet, Pannag R. Sanketi, Grecia Salazar, Michael S. Ryoo, Krista Reymann, Kanishka Rao, Karl Pertsch, Igor Mordatch, Henryk Michalewski, Yao Lu, Sergey Levine, Lisa Lee, Tsang-Wei Edward Lee, Isabel Leal, Yuheng Kuang, Dmitry Kalashnikov, Ryan Julian, Nikhil J. Joshi, Alex Irpan, Brian Ichter, Jasmine Hsu, Alexander Herzog, Karol Hausman, Keerthana Gopalakrishnan, Chuyuan Fu, Pete Florence, Chelsea Finn, Kumar Avinava Dubey, Danny Driess, Tianli Ding, Krzysztof Marcin Choromanski, Xi Chen, Yevgen Chebotar, Justice Carbajal, Noah Brown, Anthony Brohan, Montserrat Gonzalez Arenas and Kehang Han: RT-2: Vision-language-action models transferred from web-knowledge to robotic control, Proceedings of Machine Learning Research (PMLR), 229, 2165/2183 (2023)
- 2) Danny Driess, Fei Xia, Mehdi S. M. Sajjadi, Corey Lynch, Aakanksha Chowdhery, Brian Ichter, Ayzaan Wahid, Jonathan Tompson, Quan Vuong, Tianhe Yu, Wenlong Huang, Yevgen Chebotar, Pierre Sermanet, Daniel Duckworth, Sergey Levine, Vincent Vanhoucke, Karol Hausman, Marc Toussaint, Klaus Greff, Andy Zeng, Igor Mordatch and Pete Florence: PaLM-E: An embodied multimodal language model, Proceedings of Machine Learning Research (PMLR), 202, 8469/8488 (2023)
- 3) brian ichter, Anthony Brohan, Yevgen Chebotar, Chelsea Finn, Karol Hausman, Alexander Herzog, Daniel Ho, Julian Ibarz, Alex Irpan, Eric Jang, Ryan Julian, Dmitry Kalashnikov, Sergey Levine, Yao Lu, Carolina Parada, Kanishka Rao, Pierre Sermanet, Alexander T Toshev, Vincent Vanhoucke, Fei Xia, Ted Xiao, Peng Xu, Mengyuan Yan, Noah Brown, Michael Ahn, Omar Cortes, Nicolas Sievers, Clayton Tan, Sichun Xu, Diego Reyes, Jarek Rettinghouse, Jornell Quiambao, Peter Pastor, Linda Luu, Kuang-Huei Lee, Yuheng Kuang, Sally Jesmonth, Nikhil J. Joshi, Kyle Jeffrey, Rosario Jauregui Ruano, Jasmine Hsu, Keerthana Gopalakrishnan, Byron David, Andy Zeng, and Chuyuan Kelly Fu: Do As I Can, Not As I Say: Grounding Language in Robotic Affordances, Proceedings of the 6th Conference on Robot Learning (PMLR), 205, 287/318 (2023)
- 4) Kosei Nishio, Gota Okabe, Daisuke Tanaka, Masanori Hariyama: An Edge-Based Robot Control System Using an LLM and Object Detection, 12th IEEEJ international workshop on Sensing, Actuation, Motion Control, and Optimization (SAMCON2026), 516/519 (2026)
- 5) AgileX ROS TEAM: LIMO Usage and Development Manual, <https://static.generation-robots.com/media/limo-standard-user-manual-en.pdf> (参照 2026-01-03)
- 6) NVIDIA: Jetson Nano, <https://www.nvidia.com/ja-jp/autonomous-machines/embedded-systems/jetson-nano/product-development/> (参照 2026-01-03)
- 7) Renesas Electronics: RZ/V2H, <https://www.renesas.com/ja/products/rz-v2h> (参照 2026-01-03)
- 8) Chuyi Li, Lulu Li, Hongliang Jiang, Kaiheng Weng, Yifei Geng, Liang Li, Zaidan Ke, Qingyuan Li, Meng Cheng, Weiqiang Nie, Yiduo Li, Bo Zhang, Yufei Liang, Linyuan Zhou, Xiaoming Xu, Xiangxiang Chu, Xiaoming Wei, Xiaolin Wei: YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications, arXiv:2209.02976, preprint, arXiv (2022).
- 9) An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, Chujie Zheng, Dayiheng Liu, Fan Zhou, Fei Huang, Feng Hu, Hao Ge, Haoran Wei, Huan Lin, Jialong Tang, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Yang, Jiayi Yang, Jing Zhou, Jingren Zhou, Junyang Lin, Kai Dang, Keqin Bao, Kexin Yang, Le Yu, Lianghao Deng, Mei Li, Mingfeng Xue, Mingze Li, Pei Zhang, Peng Wang, Qin Zhu, Rui Men, Ruize Gao, Shixuan Liu, Shuang Luo, Tianhao Li, Tianyi Tang, Wenbiao Yin, Xingzhang Ren, Xinyu Wang, Xinyu Zhang, Xuancheng Ren, Yang Fan, Yang Su, Yichang Zhang, Yinger Zhang, Yu Wan, Yuqiong Liu, Zekun Wang, Zeyu Cui, Zhenru Zhang, Zhipeng Zhou, Zihan Qiu: Qwen3 Technical Report, arXiv:2505.09388, preprint, arXiv (2025).