OS1B-04

推論領域分割のためのグループベース方策最適化における知覚強化の統合

田邊 光†, 馬場 雄大†, 柳井 啓司†

†電気通信大学大学院 情報理工学研究科

2026/08/04 MIRU2026@出島メッセ長崎

研究背景：推論領域分割タスク (Reasoning Segmentation)

テキストが表す目標物体のマスクを画像の中から推定する

- **テキストには目標物体が直接的に表されないため、**
  モデルは目標物体を**推論 (reasoning)** する必要がある

問題：鏡の中に映っている人はどこ？



推論



研究背景：推論領域分割タスク (Reasoning Segmentation)

テキストが表す目標物体のマスクを画像の中から推定する

- **テキストには目標物体が直接的に表されないため、**
  モデルは目標物体を**推論 (reasoning)** する必要がある

問題：鏡の中に映っている人はどこ？



推論



ロボティクスや画像編集などにおける
システムの柔軟性や自律性の改善に応用が期待

研究の目的と方針

目的

推論領域分割における領域推定の正確性向上

方針

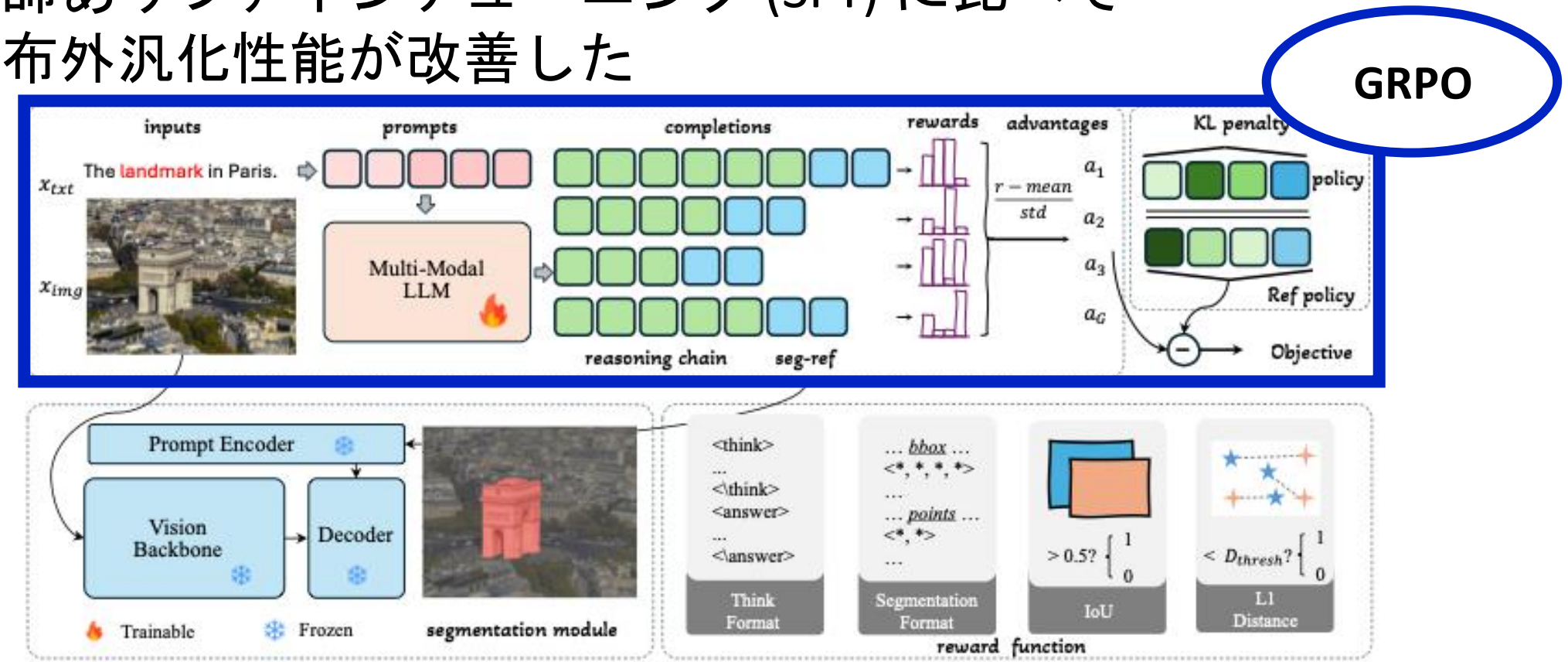
① グループベース方策最適化における学習効果を改善し、**推論性能を効果的に高める**

② 効果的な知覚損失を導入し、**知覚性能も同時に強化する**

関連研究

Seg-Zero [2]

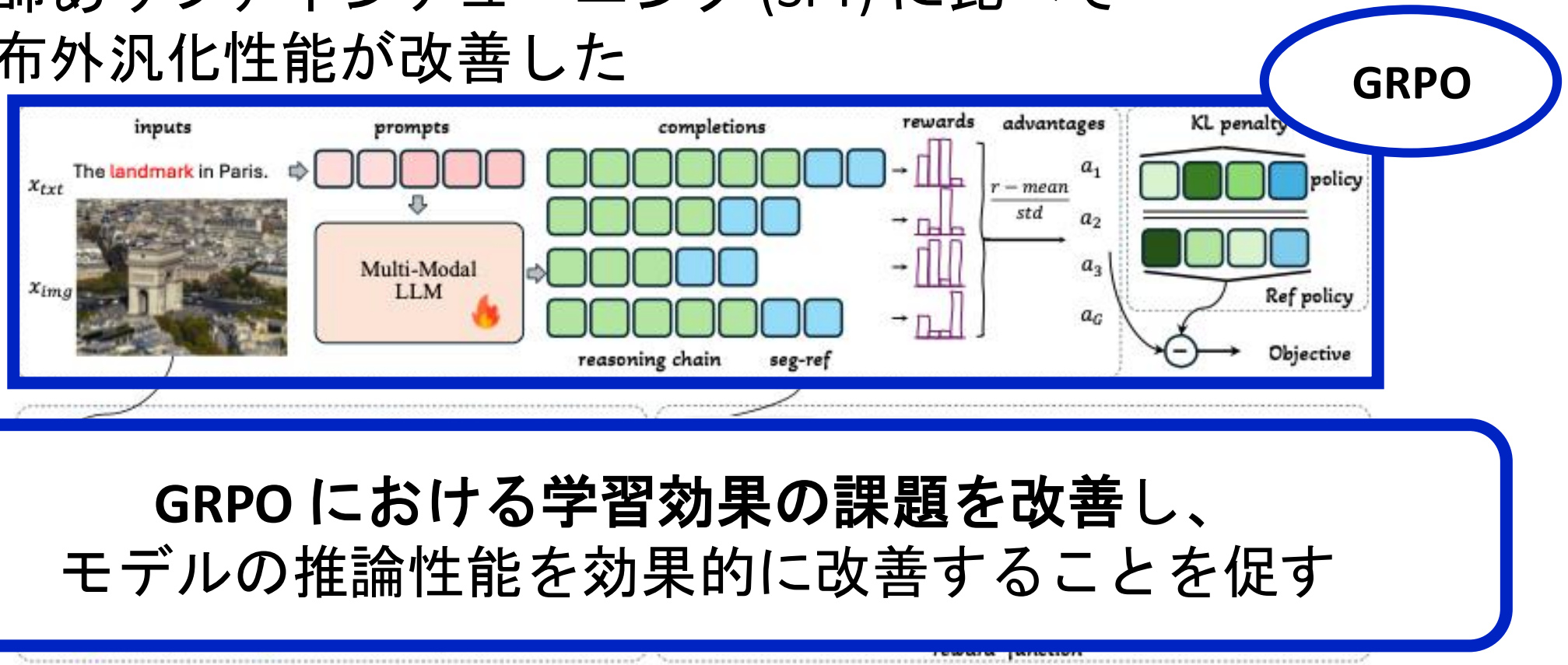
- **GRPO [1] に基づくグループベース方策最適化を推論領域分割に導入し、推論性能の改善を促す**
  - 教師ありファインチューニング (SFT) に比べて分布外汎化性能が改善した



[1] Zhihong Shao et al. DeepSeekMath: Pushing the limits of mathematical reasoning in open language models. arXiv preprint arXiv:2402.03300, 2024.
[2] Yuqi Liu et al. Seg-Zero: Reasoning-chain guided segmentation via cognitive reinforcement. arXiv preprint arXiv:2503.06520, 2025.

Seg-Zero [2]

- **GRPO [1] に基づくグループベース方策最適化**を推論領域分割に導入し、推論性能の改善を促す
  - 教師ありファインチューニング (SFT) に比べて分布外汎化性能が改善した

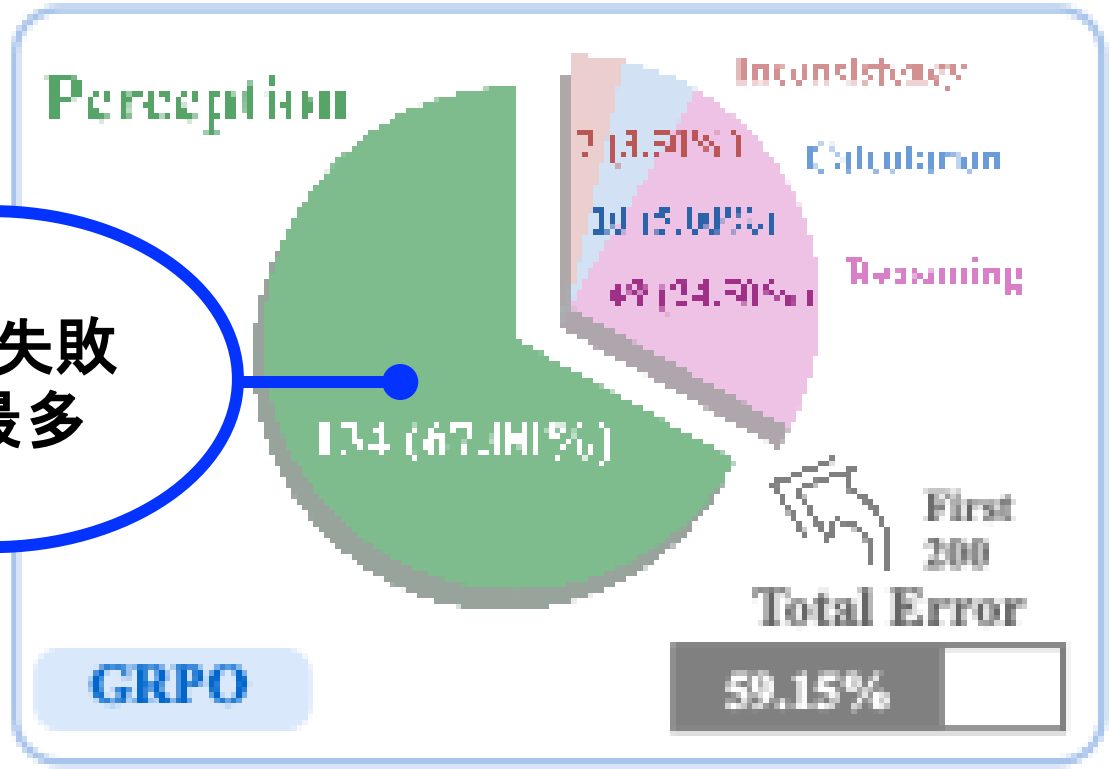


GRPO における学習効果の課題を改善し、
モデルの推論性能を効果的に改善することを促す

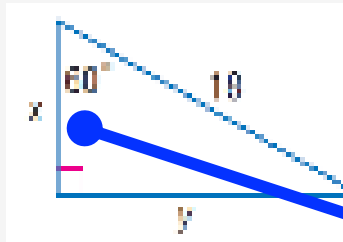
[1] Zhihong Shao et al. DeepSeekMath: Pushing the limits of mathematical reasoning in open language models. arXiv preprint arXiv:2402.03300, 2024.
[2] Yuqi Liu et al. Seg-Zero: Reasoning-chain guided segmentation via cognitive reinforcement. arXiv preprint arXiv:2503.06520, 2025.

Perception-Aware Policy Optimization (PAPO) [3]

- マルチモーダル数学タスクにおけるエラーを分析
- 推論の失敗でなく**物体の知覚失敗が主因**であることを観察



知覚失敗が最多



Question: Find x.

Ground Truth: 9

GRPO: Perception

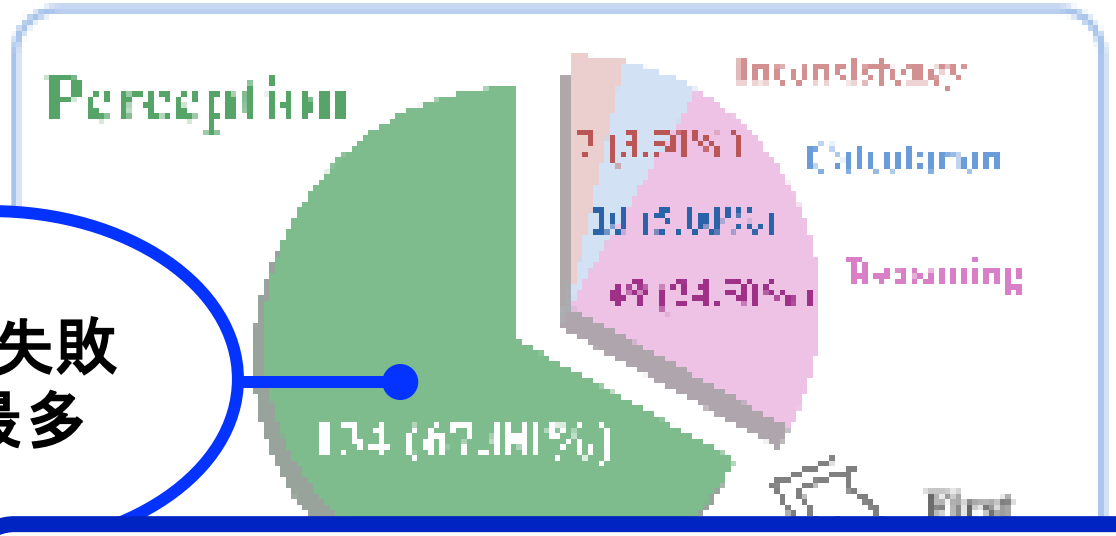
<think> To find the side x of [illegible] triangle, we can use trigono-metric ratios. The triangle has a 60-degree angle, and the hypotenuse is 18. Since x is the side opposite the 60-degree angle, we can use the sine function, which is defined as: ... Therefore, the value of x is $9\sqrt{3}$ </think> \boxed{ $9\sqrt{3}$ }

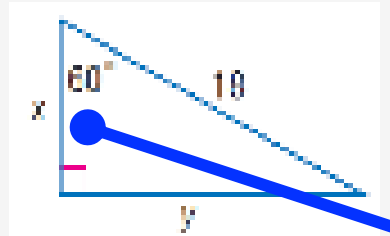
誤った知覚に基づく推論

[3] Zhenhailong Wang et al. Perception-aware policy optimization for multimodal reasoning. arXiv preprint arXiv:2507.06448, 2025.

Perception-Aware Policy Optimization (PAPO) [3]

- マルチモーダル数学タスクにおけるエラーを分析
- 推論の失敗でなく**物体の知覚失敗が主因**であることを観察





Question: Find x.

Ground Truth: 9

GRPO: Perception

<think> To find the side x of ... triangle, we can use trigono-metric ratios. The triangle has a 60 degree angle and the

知覚失敗が最多

誤った知覚に基づく推論

グループベース方策最適化に対して
知覚性能を改善する枠組みを効果的に統合する

[3] Zhenhailong Wang et al. Perception-aware policy optimization for multimodal reasoning. arXiv preprint arXiv:2507.06448, 2025.

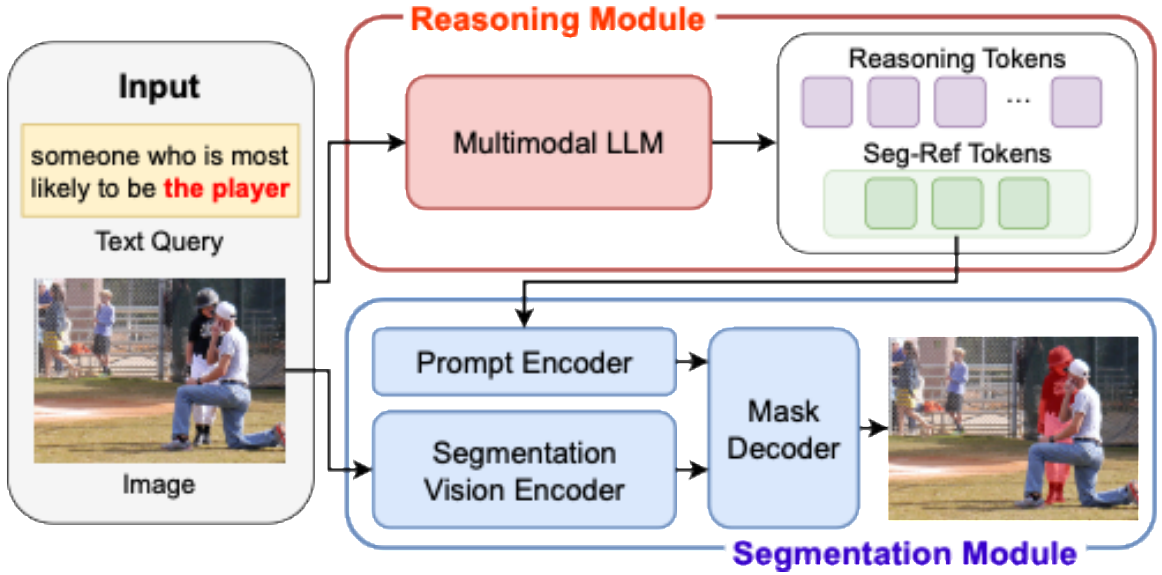
提案手法

提案手法の目次

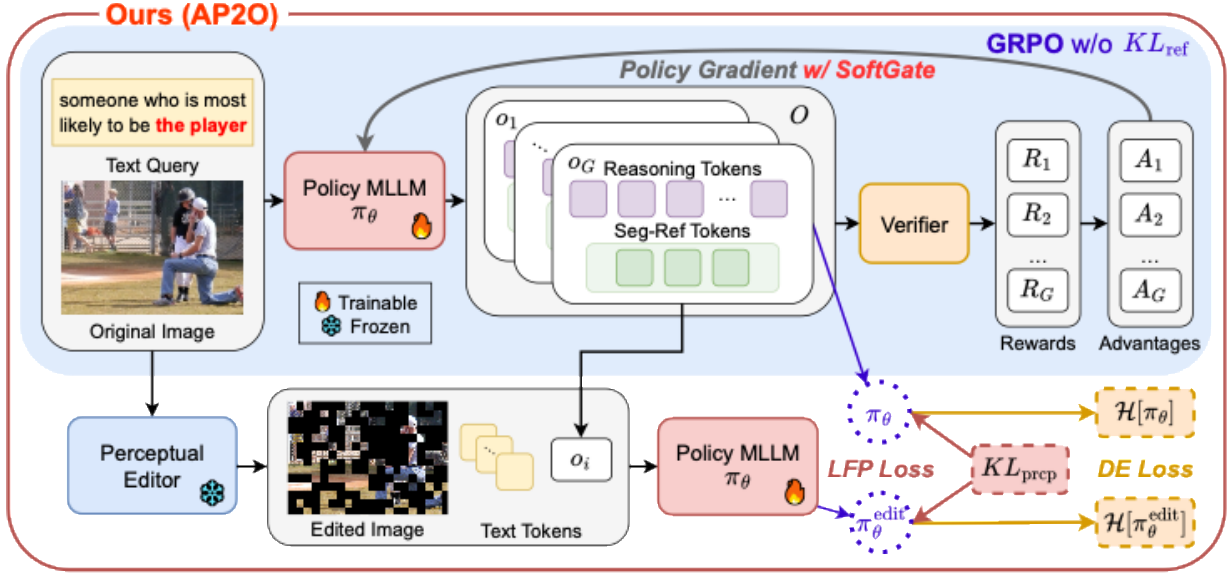
① 推論領域分割モデルの構造

② **適応的知覚方策最適化 (AP2O) (新規提案)**

③ 報酬関数



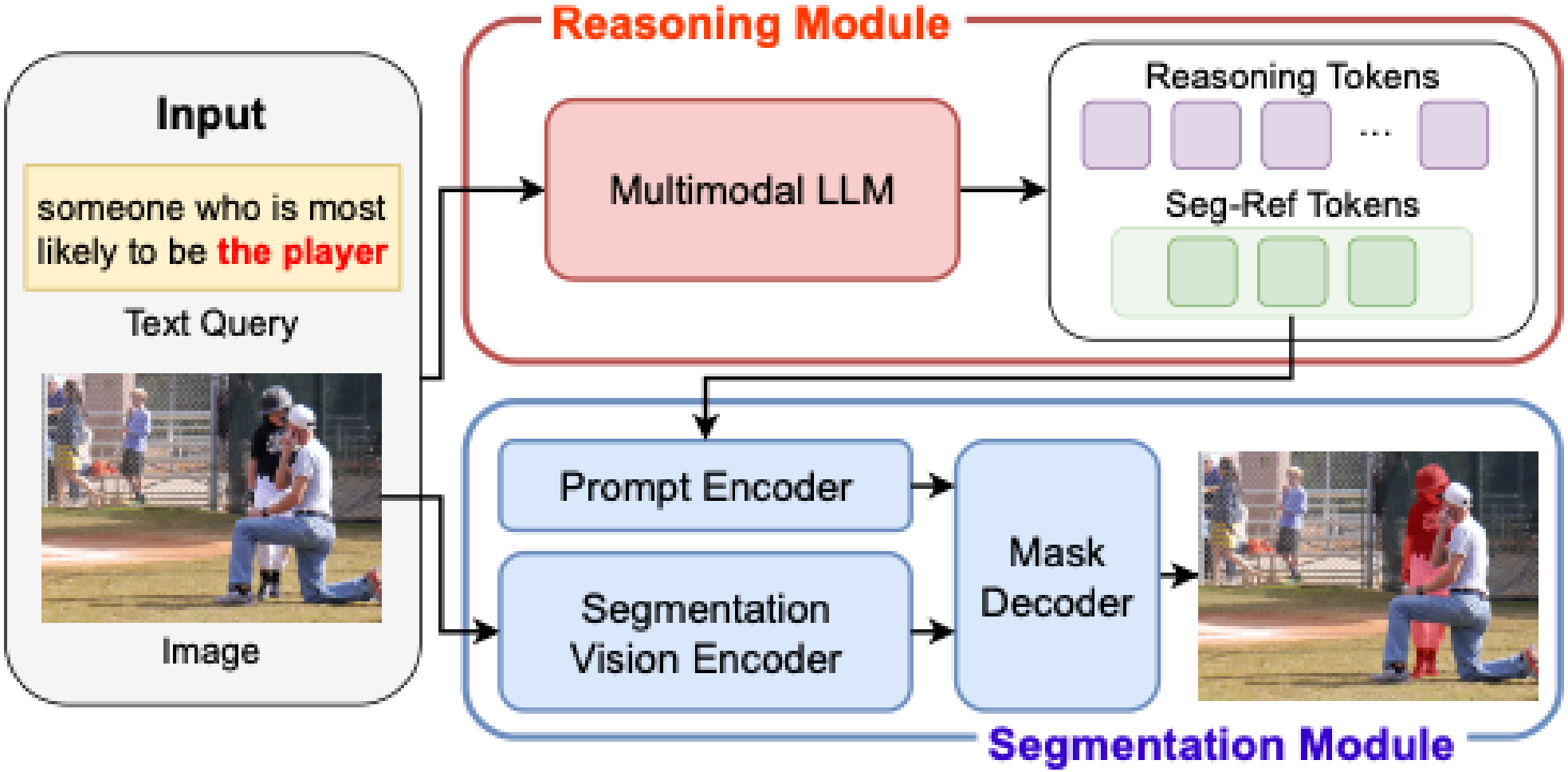
① モデル構造



②, ③ 学習手法

推論領域分割モデルの構造

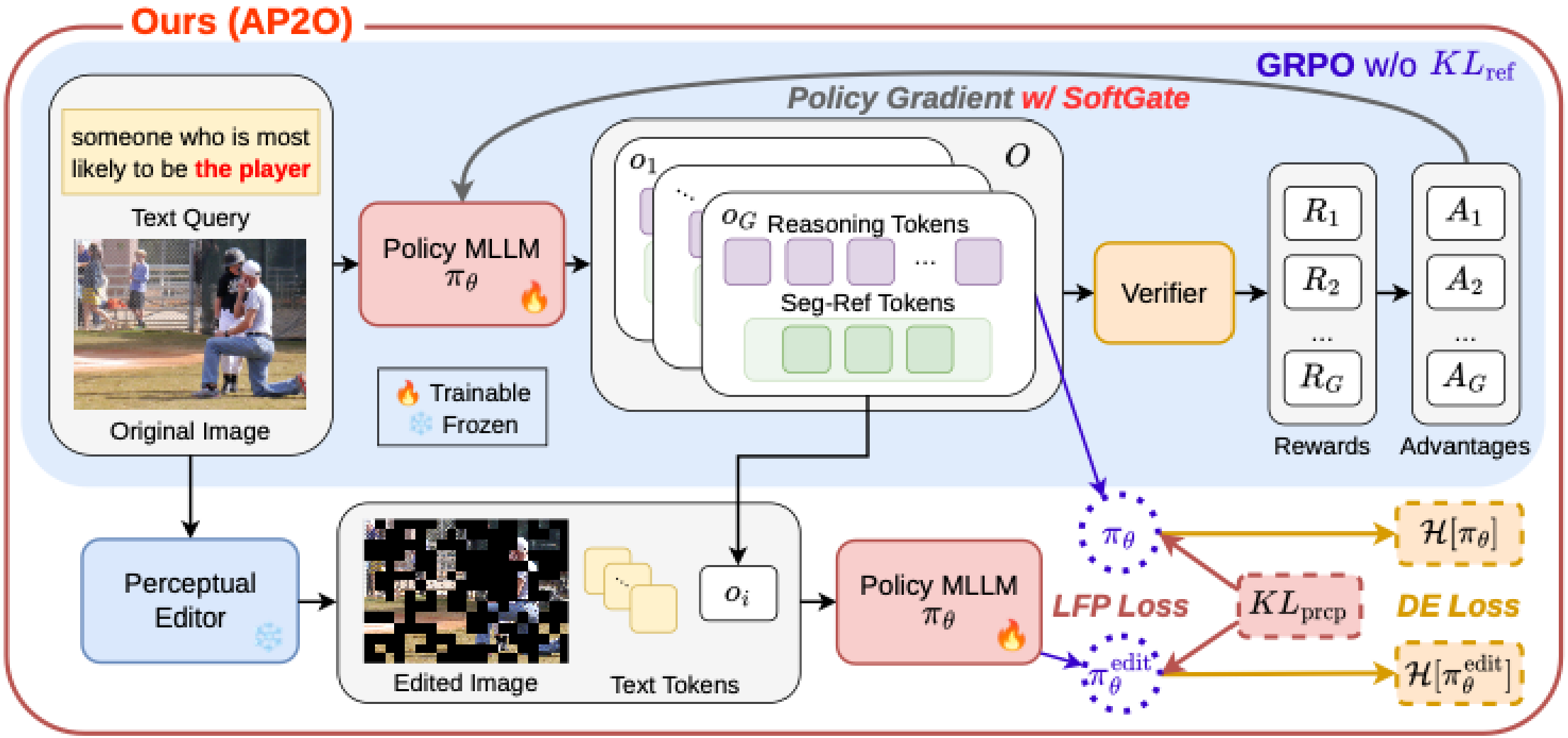
① **推論モジュール**で物体のおよその位置を特定 (ボックス・点)
② **セグメンテーションモジュール**で画素粒度のマスクを推定



適応的知覚方策最適化 (AP2O) (新規提案)

Adaptive Perception-Aware Policy Optimization

GRPO の枠組みを拡張し、推論と知覚を効果的に改善する

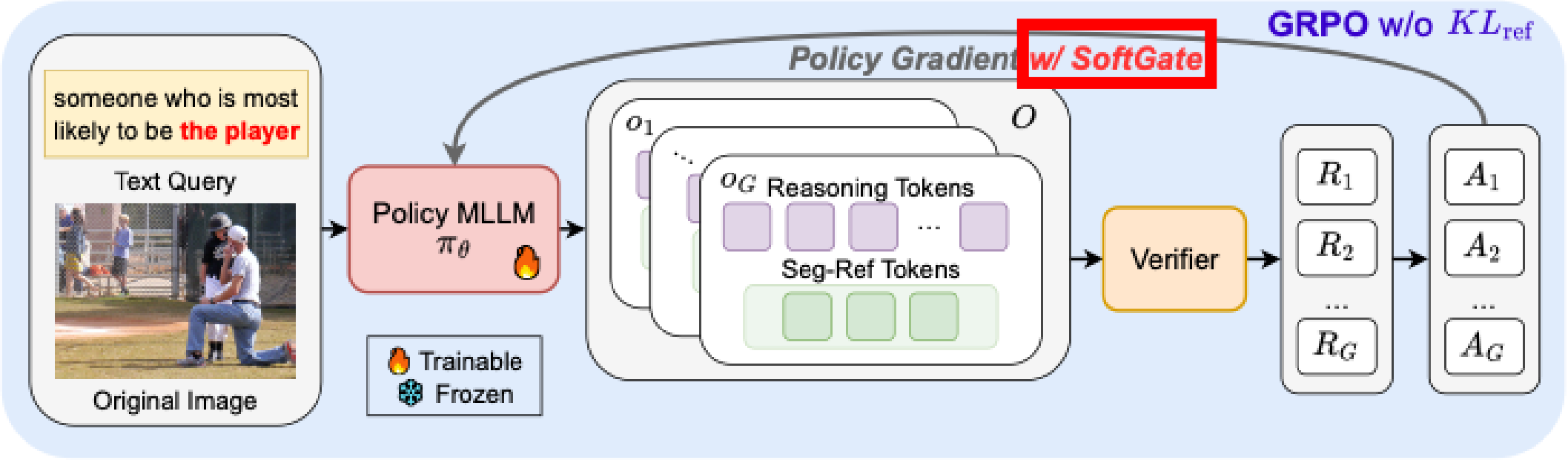


ソフトゲート制約付きグループベース方策最適化

- 重要度比 (importance ratio) に対して制約をかける
- GRPO [1] : 不連続な閾値制約 (Hard Clipping)
  - 閾値の外側で方策勾配が 0 になり、**学習が非効率**

➢ **シグモイド関数によるソフトゲート制約 [4] を導入**

  - 重要度比が大きいときの勾配消失を防ぎ、**学習効果を改善**



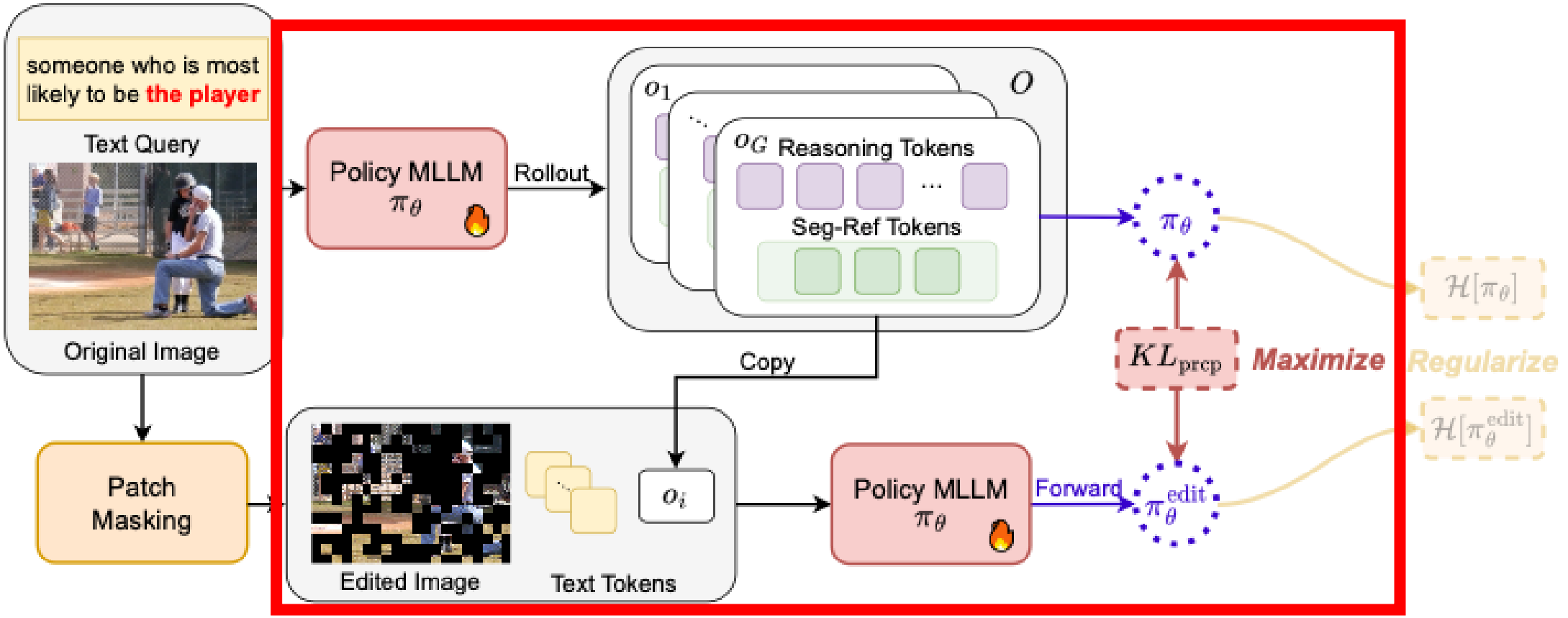
[4] Chang Gao et al. Soft adaptive policy optimization. arXiv preprint arXiv:2511.20347, 2025.

Implicit Perception Loss (IP Loss) [3]

① 元画像にパッチマスキングを適用して遮蔽画像を作成

② 各画像入力時の出力確率分布を乖離させる (KL 最大化)

➢ 画像内物体に対する依拠を強めることで、知覚の改善を促す



[3] Zhenhailong Wang et al. Perception-aware policy optimization for multimodal reasoning. arXiv preprint arXiv:2507.06448, 2025.

低周波パッチ知覚損失 (LFP Loss) (新規提案)

Low-Frequency Patch Perception Loss

- IP Loss のパッチ分割ではパッチサイズを画素数で指定
  - 画像が高解像度のとき、画像全体に対してパッチが高周波になる
    - 低周波成分に多い [5] **意味情報が十分に除去されない**
- **パッチ分割のサイズを画像サイズに応じて動的に指定**
  - 意味的情報を効果的に遮蔽し、**知覚改善の効果を高める**



Original

Pixel-based Patching

High Patch Frequency　Low Semantic Corruptness

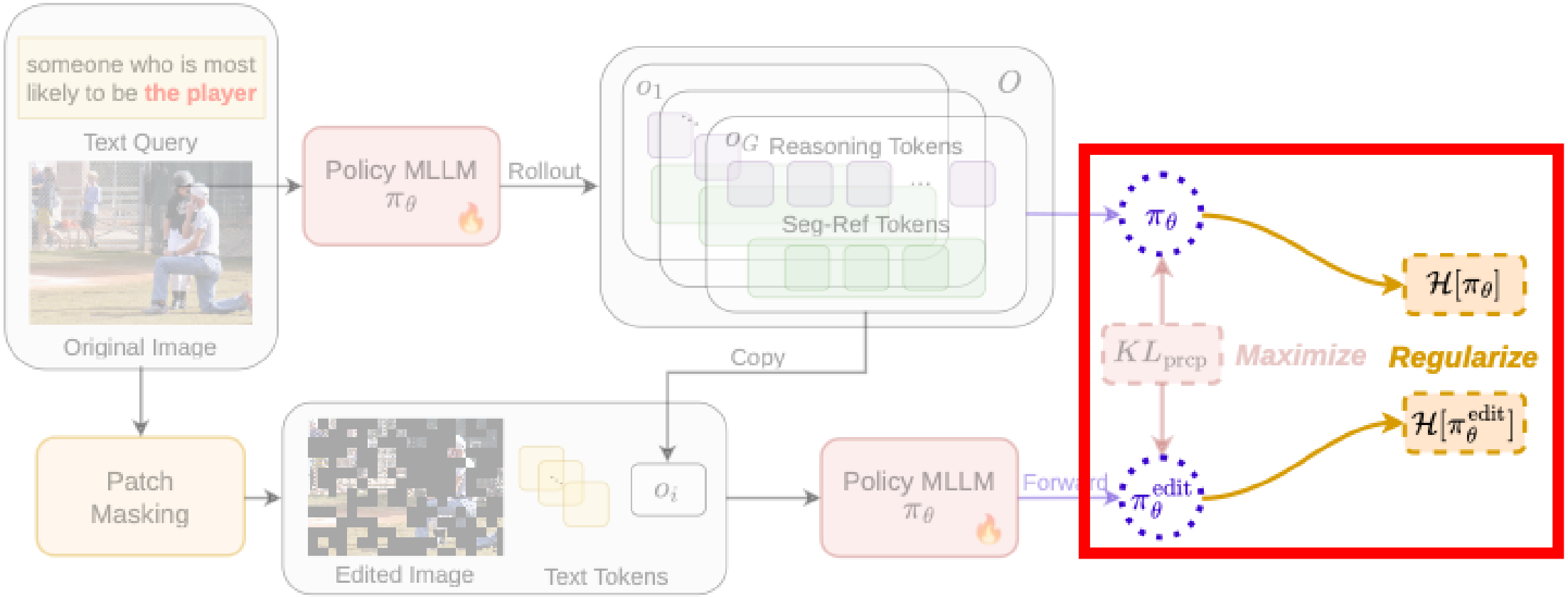
Grid-based Patching

Low Patch Frequency　High Semantic Corruptness

[5] Haohan Wang, et al. High-frequency component helps explain the generalization of convolutional neural networks. CVPR, 2020.

Double Entropy Loss (DE Loss) [3]

- LFP Loss には制約が存在せずモデルが**過剰適合しやすい**
  - ➢ 各画像入力時の方策について**エントロピーを最小化する**
  - ➢ LFP Loss に対する過剰適合を防ぎ、**学習を安定化する**



[3] Zhenhailong Wang et al. Perception-aware policy optimization for multimodal reasoning. arXiv preprint arXiv:2507.06448, 2025.

報酬関数（Seg-Zero [2] に基づく）

推論モジュールの応答に対して、次の３要素によって評価する

① **形式報酬**：指定した**応答形式の遵守**を検証する

② **正確性報酬**：視覚的接地の**幾何的正確性**を検証する
(ボックスの IoU, L1, および代表点の L2)

③ **非重複報酬**：応答内における**３文以上の繰り返し**を検証する

$$R = \lambda_{\text{fmt}} R_{\text{fmt}} + \lambda_{\text{acc}} R_{\text{acc}} + \lambda_{\text{rep}} R_{\text{rep}}$$

```
<think> The player is wearing a
helmet ... </think>
<answer>[
{"bbox_2d": [503, 41, 628, 765],
"point_2d": [573, 300]}
]</answer>
```

形式遵守



```
The player is wearing a helmet.
The player is wearing a helmet.
The player is wearing a helmet.
...
```

➢ 繰り返しを抑制

形式報酬 R_{fmt}

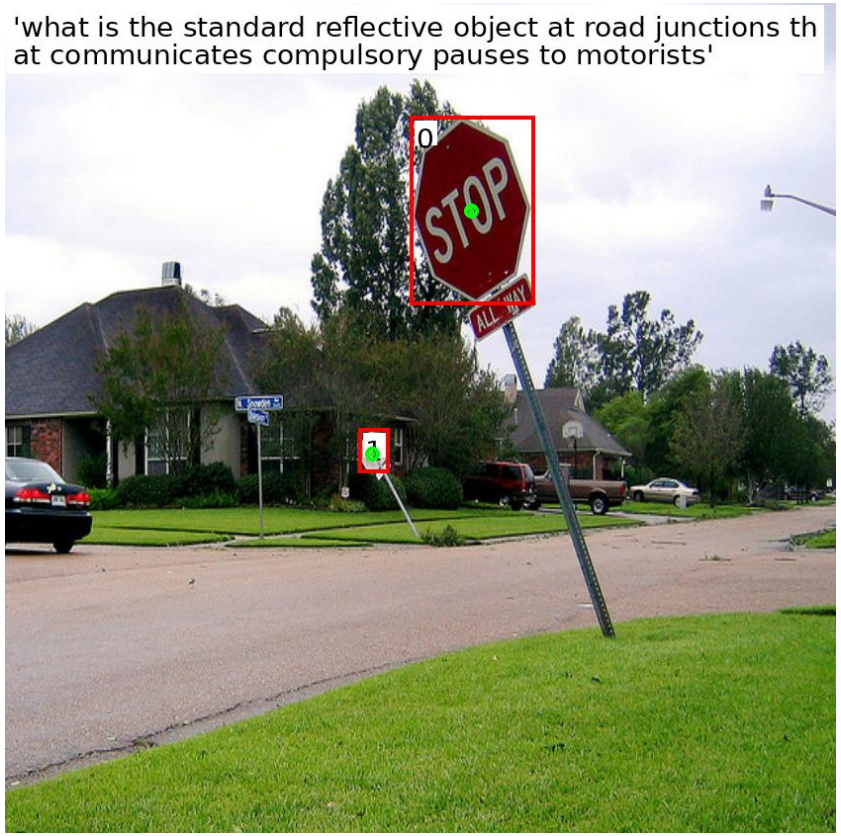
正確性報酬 R_{acc}

非重複報酬 R_{rep}

実験

訓練データセット

- 複数の領域分割データセット (LVIS, RefCOCOg, gRefCOCO, LISA++) から画像・マスク・テキストを収集 (約 7000 件)
- マスクを覆う**ボックスおよびマスクの重心**を訓練に使用





ReasonSeg [6] ベンチマーク

- 推論領域分割において最も標準的なベンチマーク
- テキストクエリは**正解物体を直接的に表さない**
- 訓練セットで学習せず、**分布外設定での評価結果を報告**



[6] Xin Lai et al. LISA: Reasoning segmentation via large language model. CVPR, 2024.

実験設定

- **MLLM**：Qwen2.5-VL 3B [7]
- **領域分割モデル**：SAM2 (Hiera-L) [8]
- **グループサイズ**：8
- **パッチ分割数**：14
- **パッチ遮蔽割合**：80%
- **LFP Loss の係数**：0.02
- **DE Loss の係数**：0.05

DAPO [9] に基づく設定

- **参照方策との KL 正則化**：無効
- **Clip-Higher**：（ソフトゲート制約による代替のため未定義）
- **動的サンプリング**：有効
- **トークンレベルの方策勾配損失**：有効
- **長期報酬成形**：無効

[7] Shuai Bai et al. Qwen2.5-VL Technical Report. arXiv preprint arXiv:2502.13923, 2025.
[8] Nikhila Ravi et al. SAM 2: Segment anything in images and videos. arXiv preprint arXiv:2408.00714, 2024.
[9] Qiying Yu et al. DAPO: An open-source llm reinforcement learning system at scale. arXiv preprint arXiv:2503.14476, 2025.

ReasonSeg ベンチマークにおける正確性評価

同モデルサイズ帯 (3B+) で
既存手法と比べて
高い正確性を達成

同バックボーンを用いた
RL ベースラインである
Seg-Zero から改善

Category	Method	ReasonSeg val		ReasonSeg test	
		gIoU ↑	cIoU ↑	gIoU ↑	cIoU ↑
w/o MLLM	OVSeg	28.5	18.6	26.1	20.8
	ReLA	22.4	19.9	21.3	22.0
	X-Decoder	22.6	17.9	21.7	16.3
	SEEM	25.5	21.2	24.3	18.7
	Grounded-SAM	26.0	14.5	21.3	16.4
w/ MLLM SFT	LISA-7B [CVPR24]	44.4	46.0	36.8	34.1
	LISA-13B [CVPR24]	48.9	46.9	44.8	45.8
	GLaMM 7B [CVPR24]	47.4	47.2	-	-
	SAM4MLLM 7B [ECCV24]	46.7	48.1	-	-
	SegLLM 7B [ICLR25]	57.2	54.3	52.4	48.4
	HRSeg 7B [ACMMM25]	56.7	55.4	54.6	<u>51.2</u>
	HyperSeg 3B [CVPR25]	59.2	**56.7**	-	-
w/ MLLM RL	Qwen2.5VL 3B+SAM2 (w/o RL)	53.8	44.1	47.6	37.4
	Seg-Zero 3B [arXiv25]	58.2	53.1	<u>56.1</u>	48.6
	Seg-R1 3B [NeurIPSW25]	<u>60.8</u>	56.2	55.3	46.6
	Ours 3B	**63.7**	<u>56.3</u>	**60.0**	**53.1**
		↑+5.5	↑+3.2	↑+3.9	↑+4.5

応答例②：微細領域の知覚

微細な領域を正確に知覚して暗示領域を推定することが可能

Query the part of the pear that has been bitten by insects



GRPO ... **There is no visible bite mark or insect damage** on the pear ...

Ours ... **the small, yellowish area near the stem.**

おわりに

おわりに

推論領域分割において適応的知覚方策最適化 (AP2O) を提案

① 推論の学習効果改善のために**ソフトゲート制約**を導入

② 効果的な知覚強化のために **LFP Loss** を提案

➢ SFT, RL ベースラインから**領域推定の正確性を改善**

今後の方針

- 推論過程の内容やマスクに対する明示的な報酬の検証
- 局所遮蔽や画像編集等の意味的編集を用いた知覚損失の検討