CONFIDENTIAL EXTENDED ABSTRACT.
DO NOT DISTRIBUTE ANYWHERE.

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[1,a)]　馬場 雄大[1,b)]　柳井 啓司[1,c)]

概要

推論領域分割は，画像と複雑なテキストクエリが与えられたときに，高度な推論を通じて対象物を特定し，その領域を推定するタスクである．従来手法は，マルチモーダル大規模言語モデルに対して強化学習を適用する際，多くの場合 GRPO を単純に適用した手法に依存している．このため，推論能力の学習効率に改善の余地が残るとともに，推論領域分割において重要となる視覚的接地の正確性を支える知覚能力の向上が限定的である．そこで本研究では，推論能力と知覚能力の安定した同時最適化を可能にする新たなグループベース方策最適化アルゴリズムを提案する．本手法で学習されたモデルは，同等のモデル規模において既存の強化学習ベースラインである Seg-Zero を大きく上回る性能を示し，ReasonSeg の validation セットで +5.5 gIoU，test セットで +4.5 cIoU の性能向上を達成した．

1. はじめに

推論領域分割は，目標物体を暗示するテキストと画像が与えられたときに，目標物体を推論し画素粒度でその領域を推定するタスクである．本タスクは，ロボティクス，異常検知，画像編集などの分野において，システムの自律性やユーザビリティの向上に貢献が期待されている．近年，大規模言語モデル (LLMs) の訓練においてグループベース方策最適化に基づく強化学習が導入され，特に推論系タスクでの性能向上が報告されていることを背景に，推論領域分割に強化学習の適用が進められている [8]．しかし，既存手法は標準的な GRPO アルゴリズム [14] の適用に留まり，学習効果の改善や，推論領域分割に不可欠な知覚性能の強化は十分に検討されていない．そこで本研究では，推論領域分割における領域推定の正確性向上を目的として，グループベース方策最適化手法における学習効果と知覚性能を効果的に高める方法を提案し，その有効性を検証する．

[1] 電気通信大学
[a)] tanabe-h@mm.inf.uec.ac.jp
[b)] baba-y@mm.inf.uec.ac.jp
[c)] yanai@cs.uec.ac.jp

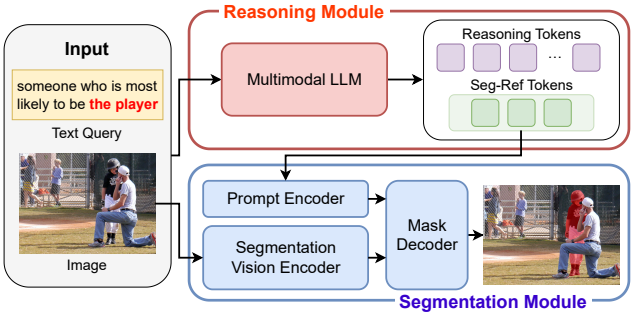


図 1: 推論領域分割モデルの構造．

2. 関連研究

Seg-Zero [8] は，推論領域分割に対して Group Relative Policy Optimization (GRPO) [14] に基づく強化学習を導入し，教師ありファインチューニング (SFT) と比較して分布外設定における汎化性能を改善した．しかし，GRPO には学習の不安定性や探索性の低下に関する課題が存在し，目的関数の観点から学習効果に改善余地が残る．Perception-Aware Policy Optimization (PAPO) [17] は，推論系視覚言語タスクの誤りを分析し，その主因が推論性能でなく画像内物体の知覚失敗にあることを観察した上で，知覚性能を高める Implicit Perception Loss (IP Loss) を導入した．推論領域分割においては，物体の種類や位置に関する正確な把握が視覚的接地の正確性を左右するため，知覚性能改善の工夫は領域推定の正確性向上に寄与すると考えられる．

3. 提案手法

本節では，本研究で提案する推論領域分割フレームワークとして，導入するモデル構造 (節 3.1) および新たに提案するグループベース方策最適化手法 (節 3.2) を説明する．最後に，Section 3.3 では，非反復報酬，フォーマット報酬，および複数物体に対応した精度報酬を含む報酬設計について説明する．

3.1 モデル構造

本研究で用いる推論領域分割モデルの構造を**図 1** に示す．本モデルは，画像内容を知覚しテキストに基づいて目標物

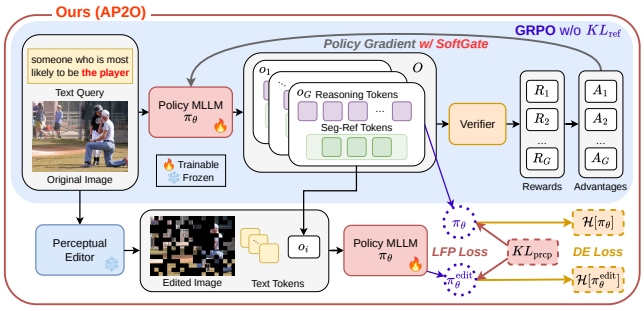


図 2: 提案するグループベース方策最適化手法.

体を特定する推論モジュールと，目標物体に対して画素粒度のマスクを推定するセグメンテーションモジュールから構成される．推論モジュールにはマルチモーダル大規模言語モデル (Multimodal Large Language Model, MLLM) を用いて，画像とテキストの情報を統合して目標物体のおおよその位置を特定し，各物体のバウンディングボックスと代表点を領域参照として推定する．セグメンテーションモジュールにはプロンプタブル領域分割モデル (Promptable Segmentation Model, PSM) を用いて，MLLM が出力した目標物体の領域参照を手がかりとして，画素粒度のマスクを推定する．

3.2 Adaptive Perception-Aware Policy Optimization (AP2O)

推論領域分割において重要な推論能力と知覚能力を同時に高めるために, 新たなグループベース方策最適化手法として AP2O (Adaptive Perception-Aware Policy Optimization) を提案する (**図 2**).

3.2.1 目的関数

AP2O では，GRPO に対して推論性能の向上を促進するために，SAPO に基づくソフトゲート型代理目的関数 [2] を導入する．また，有効勾配の消失を防ぐために dynamic sampling を組み合わせるとともに，過度な正則化を避けるために KL 正則化項を除去する [22]．さらに，視覚的証拠への注意と正確な位置同定を促すために，新たに提案する Low-Frequency Patch Perception Loss (LFP Loss) を導入することで，モデルの視覚的接地を効果的に改善する．

本研究の目的関数は，GRPO [14] によるグループベース方策最適化の定式化に基づき，式 (1) で表される．

$$\mathcal{J}(\theta) = \mathbb{E}_{(q,I)\sim\mathcal{D},\{o_i\}_{i=1}^{G}\sim\pi_{\theta_{\text{old}}}(O|q,I)} \frac{1}{G}\sum_{i=1}^{G}\frac{1}{|o_i|}\sum_{t=1}^{|o_i|} \Bigg\{ f_{i,t}(r_{i,t}(\theta))\widehat{A}_{i,t} + \gamma\,\mathbb{D}_{\text{KL}}\big[\pi_\theta \,\|\, \pi_\theta^{\text{edit}}\big] - \eta_1\,\mathcal{H}[\pi_\theta] - \eta_2\,\mathcal{H}\big[\pi_\theta^{\text{edit}}\big] \Bigg\} \tag{1}$$

ただし，π_θ は更新時方策, $\pi_{\theta_{\text{old}}}$ はロールアウト方策，π_θ^{edit} は後述する編集時方策，$\widehat{A}_{i,t}$ はグループ内アドバンテージ，$r_{i,t}(\theta)$ は更新時方策とロールアウト方策の重要度比である．γ, η_1, η_2 は各項の重みを制御するハイパーパラメータである．ここで，$f_{i,t}(x)$ はソフトゲート関数であり，式 (2) で定義される．

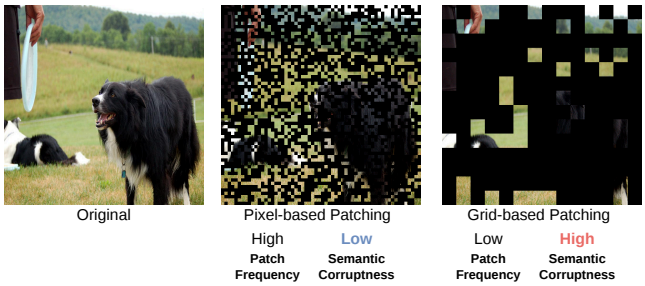


図 3: 低周波パッチの構成と意味的破壊の比較.

$$f_{i,t}(x) = \frac{1}{1+e^{-\tau_{i,t}\,(x-1)}}\cdot\frac{4}{\tau_{i,t}}, \quad \tau_{i,t} = \begin{cases} \tau_{\text{pos}}, & \text{if } \widehat{A}_{i,t} > 0 \\ \tau_{\text{neg}}, & \text{otherwise} \end{cases} \tag{2}$$

τ_{pos} と τ_{neg} は，それぞれ $f_{i,t}(x)$ における正例トークンおよび負例トークンに対応する温度パラメータである．

3.2.2 Low-Frequency Patch Perception Loss (LFP Loss)

視覚的証拠への注意と正確な位置同定を促すために，本研究では Low-Frequency Patch Perception Loss（LFP Loss）を提案する．

Implicit Perception Loss (IP Loss) [17] では，テキストクエリと画像を更新時方策 π_θ に入力した際のトークン生成分布と，画像に対してパッチマスキングによる画像編集を付加した場合における方策 π_θ^{edit} のトークン生成分布を，KL ダイバージェンス最大化に基づき乖離する．これにより，モデルがトークンを生成する際に画像内の正確な知覚に依拠することを促し，モデルの知覚性能改善を促す．ここで，PAPO は，画像に対する高周波な摂動による編集では，画像内の意味的内容に対する破壊が十分に生じず，視覚的証拠への注意を強化するには不十分であることを報告した [17]．実際，物体形状や大域的なレイアウトは主として低周波成分によって表現されることが知られており [15]，高周波ノイズでは意味情報を効果的に除去することが難しい．しかし，IP Loss は通常，パッチマスキングにおけるパッチサイズを絶対的なピクセルサイズで固定しているため，画像解像度が大きい場合にはマスクパターンが実質的に高周波化してしまう．

そこで，画像解像度に依存せず低周波な破壊を維持するために，LFP Loss ではパッチサイズを画像サイズに対する相対値として設定する．具体的には，$u \times v$ グリッド上の各パッチ $p_{u,v}$ の一辺の長さ l_p を式 (3) で定義する．

$$l_p = \frac{\min(H, W)}{\alpha} \tag{3}$$

ここで，H と W は画像の高さおよび幅，α は短辺方向

に配置するグリッド数を表すハイパーパラメータである．l_p とランダムマスキング演算子 $\mathcal{E}$ を用いて，編集後画像 $I^{\text{edit}} = \mathcal{E}(I, l_p)$ を得る．

図 3 は，固定ピクセルサイズによるピクセルベース分割と，画像サイズに基づく相対的なパッチサイズによるグリッドベース分割の例を示している．ピクセルベース分割（中央）では，犬やフリスビーがかろうじて認識可能であり，物体の同一性や大域的レイアウトが十分に破壊されていないことがわかる．その結果，固定ピクセルサイズに基づく高周波マスキングでは，意味的破壊は限定的である．一方，グリッドベース分割（右）では，画像に 2 匹の犬と 1 つのフリスビーが含まれていることや，それらのおおよその位置を把握することが困難になっており，解像度を考慮したパッチサイズによる低周波マスキングの方が，より強い意味的破壊をもたらすことが示されている．

このような低周波マスキングを導入し，意味的に重要な視覚的証拠が遮蔽される割合を高めることで，本手法は方策が意味的に重要な視覚的証拠により強く注意を向けるよう促す．

LFP Loss $\mathbb{D}_{\text{KL}}[\pi_\theta || \pi_\theta^{\text{edit}}]$ は，式 (4) による KL ダイバージェンスの低分散近似値として計算する [13].

$$\mathbb{D}_{\text{KL}}[\pi_\theta || \pi_\theta^{\text{edit}}] = r_i^{\text{prcp}}(\theta) - \log r_i^{\text{prcp}}(\theta) - 1 \qquad (4)$$

ただし，知覚重要度比 r^{prcp} は式 (5) で定められる．

$$r^{\text{prcp}}(\theta) = \frac{\pi_\theta(o \mid q, I)}{\pi_\theta(o \mid q, I^{\text{edit}})} \qquad (5)$$

3.2.3 Double Entropy Loss (DE Loss)

$\mathcal{H}[\pi_\theta] = \log \pi_\theta(o|q, I)$ および $\mathcal{H}[\pi_\theta^{\text{edit}}] = \log \pi_\theta(o|q, I^{\text{edit}})$ は，Double Entropy (DE) Loss [17] であり，LFP Loss に対する過剰適合を防ぐ正則化のために導入する．なお, 本研究では $\mathcal{H}[\pi_\theta], \mathcal{H}[\pi_\theta^{\text{edit}}]$ をそれぞれ Original Entropy (OE) Loss および Masked Entropy (ME) Loss と区別する．

3.3 報酬設計

Seg-Zero および VisionReasoner [9] に従い，本研究では出力形式への準拠，領域参照の幾何的正確性，および過剰な反復の抑制を同時に促進する統一報酬を用いる．

最終報酬 R は式 (6) の重み付き和で定義する．

$$R = \lambda_{\text{fmt}} R_{\text{fmt}} + \lambda_{\text{acc}} R_{\text{acc}} + \lambda_{\text{rep}} R_{\text{rep}} \qquad (6)$$

3.3.1 フォーマット報酬 R_{fmt}

応答が要求された構造化形式（推論部分と回答部分）に従うか，および回答から領域参照（バウンディングボックスと代表点）が正しく解析できるかを評価する．

3.3.2 正確性報酬 R_{acc}

推定参照と正解参照をハンガリー法で対応付けた後，バウンディングボックスの IoU，バウンディングボックスの L1 距離，および代表点の L2 距離に基づいて幾何的正確性を評価する (式 (7)).

$$R_{\text{acc}} = \lambda_{\text{IoU}} R_{\text{IoU}} + \lambda_{\text{L1}} R_{\text{L1}} + \lambda_{\text{pt}} R_{\text{pt}} \qquad (7)$$

3.3.3 非反復報酬 R_{rep}

応答内で同一文が過度に繰り返されることを防ぐため，文の重複回数が閾値を超えた場合にペナルティを与える．

4. 実験

4.1 実験設定

本実験では，MLLM および PSM として，それぞれ Qwen2.5-VL 3B [1] と SAM2 (Hiera-L) [11] を用いた．この選択により，Seg-Zero との公平な比較を維持しつつ，学習設定を実用的なメモリ制約内に収めた．学習時は MLLM のみを訓練可能とした．

ソフトゲート関数においては，正例トークンおよび負例トークンに対する温度パラメータを，それぞれ $\tau_{\text{pos}} = 1.0$ および $\tau_{\text{neg}} = 1.05$ に設定した．また，R_{acc} の計算においては，閾値を $\tau_{\text{IoU}} = 0.5$，$\epsilon_{\text{bbox}} = 10$，$\epsilon_{\text{pt}} = 30$ とした．R_{rep} に対しては $D_{\text{dup}} = 2$ を設定した．先行研究 [8], [9] に従い，報酬項に関するすべての係数はそれらの設定に合わせて 1 とした．最適化には AdamW を用い，学習率は 1.0×10^{-6} とした．更新バッチサイズは 32 とし，学習は 8 枚の NVIDIA RTX A6000 を用いて実施した．

4.2 学習データセット

VisionReasoner [9] で提供されている領域参照アノテーションを用いた．これらは，LVIS [3]，RefCOCOg [21]，gRefCOCO [7]，LISA++ [19] を含む複数の領域分割データセットから構築されたものである．各学習サンプルは，画像，参照表現（テキストクエリ），および各物体に対して 1 つのバウンディングボックスと 1 つの代表点で表される複数の領域参照から構成される．これらの領域参照は，インスタンスレベルのマスクから，それを囲むバウンディングボックスとマスク重心を計算することで生成される．本研究では約 7k 件の学習サンプルを用いた．

4.3 評価ベンチマーク

評価には推論領域分割の標準的なベンチマークである ReasonSeg [4] を用いて，検証セットとテストセットの評価結果を報告する．なお, 学習時には ReasonSeg のデータを一切使用しておらず，分布外設定に対する頑健性を検証する．また，先行研究に従い，評価指標には gIoU と cIoU を用いる．gIoU は画像ごとの IoU を平均する指標である．一方，cIoU はデータセット全体で交差領域と和集合領域を集約するため，大きな物体により大きな重みがかかる．先行研究と同様に，本研究では gIoU を主要指標とする．

表 1: ReasonSeg ベンチマークによる推論領域分割の結果.

Category	Method	val		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	-	-
	SegLLM 7B [ICLR25]	57.2	54.3	52.4	48.4
	HRSeg 7B [ACMMM25]	56.7	55.4	54.6	51.2
	HyperSeg 3B [CVPR25]	59.2	**56.7**	-	-
w/ MLLM RL	Qwen2.5VL 3B+SAM2	53.8	44.1	47.6	37.4
	Seg-Zero 3B [arXiv25]	58.2	53.1	56.1	48.6
	Seg-R1 3B [NeurIPSW25]	60.8	56.2	55.3	46.6
	Ours 3B	**63.7**	56.3	**60.0**	**53.1**

4.4 推論領域分割における結果

表 1 に，ReasonSeg による推論領域分割の結果を示す．比較対象として，MLLM を用いない Open-Vocabulary Segmentation ベースラインである OVSeg [5]，ReLA [7]，X-Decoder [23]，SEEM [24]，Grounded-SAM [12] を用いる．また，MLLM を用いる既存手法として，LISA [4]，GLaMM [10]，SegLLM [16]，HRSeg [6]，HyperSeg [18]，Seg-Zero [8]，Seg-R1 [20] と比較する．

MLLM を用いない手法は推論を必要とする参照表現による対象特定が困難であり，gIoU は 20% 台にとどまる．一方，MLLM を用いる手法は大幅に性能が向上し，gIoU は 50% 台後半に達する．提案手法は，訓練前の Qwen2.5-VL 3B + SAM2 ベースラインに対して大幅な改善を示し，validation gIoU を 53.8 から 63.7 (+9.9)，test gIoU を 47.6 から 60.0 (+12.4) へと向上させた．また，同一バックボーンを用いた既存の RL 手法である Seg-Zero と比較しても，validation gIoU を 58.2 から 63.7 (+5.5)，test cIoU を 48.6 から 53.1 (+4.5) へと改善した．

図 4 に，ReasonSeg のテストセットにおける定性結果を示す．GRPO で学習されたモデルが混同しやすい対象を誤って予測するのに対し，本手法は推論を通じて正しい対象を特定し，適切に視覚的接地を行うことに成功している．また，極めて細粒度な対象領域に対しても安定した推定を示す．これは，知覚損失によって視覚的証拠の利用が改善されたためと考えられる．以上の結果より，本手法は推論能力と知覚能力の両方を改善し，最終的な推論領域分割の正確性を高められることがわかる．

4.5 ソフトゲートと知覚損失の効果

表 2 に，各要素の寄与を示す．ソフトゲート (SG) を導入すると，GRPO ベースラインに対してスコアが向上し，

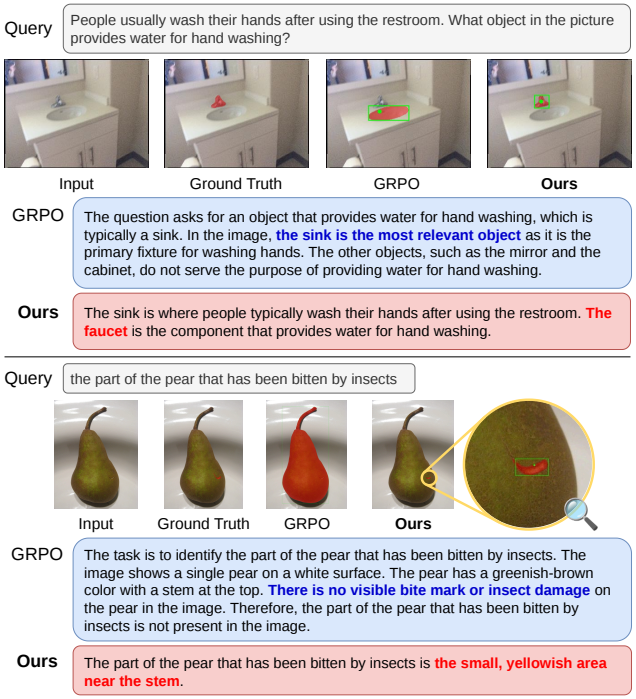


図 4: ReasonSeg におけるモデルの応答.

表 2: 主要設計要素の効果.

Setting	Modification			val		test		Avg.
	SG	Prcp.	DE	gIoU	cIoU	gIoU	cIoU	
Naive GRPO	✗	✗	✗	60.7	50.0	57.2	54.6	55.6
+IP+DE	✗	IP	✓	59.4	52.2	58.3	55.0	56.2
+SG	✓	✗	✗	61.7	58.5	57.3	50.7	57.0
SG+IP+DE	✓	IP	✓	61.8	52.4	59.0	50.4	55.9
SG+LFP+DE	✓	LFP	✓	63.7	56.3	60.0	53.1	58.3

重要度比制御による安定化が有効であることが分かる．

IP Loss と DE Loss の組み合わせ (+IP+DE) はテスト性能を改善するが，SG と併用した SG+IP+DE では平均性能は十分に向上しない．一方で IP Loss を提案する LFP Loss に置き換えた SG+LFP+DE が最良の結果を示し，ソフトゲートによる安定化と適切な知覚損失が相補的に機能することを示している．

5. おわりに

本研究では，推論領域分割における推定の正確性向上を目的として，新たなグループベース方策最適化手法 AP2O を提案した．AP2O は学習を安定化するソフトゲート機構と，画像の低周波情報を効果的に破壊する LFP Loss を統合することで，推論能力と知覚能力を同時に最適化する枠組みを実現する．実験の結果，提案手法は ReasonSeg において既存の RL 手法を一貫して上回ることを確認した．

今後は，推論過程の内容や画素粒度マスクを直接考慮した報酬設計の検討，および知覚損失の特性に関するより詳細な分析を通じて，推論領域分割における知覚性能のさらなる向上を目指す．

参考文献

[1] Bai, S., Chen, K., Liu, X., Wang, J., Ge, W., Song, S., Dang, K., Wang, P., Wang, S., Tang, J., Zhong, H., Zhu, Y., Yang, M., Li, Z., Wan, J., Wang, P., Ding, W., Fu, Z., Xu, Y., Ye, J., Zhang, X., Xie, T., Cheng, Z., Zhang, H., Yang, Z., Xu, H. and Lin, J.: Qwen2.5-VL Technical Report, *arXiv preprint arXiv:2502.13923* (2025).

[2] Gao, C., Zheng, C., Chen, X.-H., Dang, K., Liu, S., Yu, B., Yang, A., Bai, S., Zhou, J. and Lin, J.: Soft adaptive policy optimization, *arXiv preprint arXiv:2511.20347* (2025).

[3] Gupta, A., Dollar, P. and Girshick, R.: Lvis: A dataset for large vocabulary instance segmentation, *Proc. of IEEE Computer Vision and Pattern Recognition*, pp. 5356–5364 (2019).

[4] Lai, X., Tian, Z., Chen, Y., Li, Y., Yuan, Y., Liu, S. and Jia, J.: LISA: Reasoning Segmentation via Large Language Model, *Proc. of IEEE Computer Vision and Pattern Recognition* (2024).

[5] Liang, F., Wu, B., Dai, X., Li, K., Zhao, Y., Zhang, H., Zhang, P., Vajda, P. and Marculescu, D.: Open-vocabulary semantic segmentation with mask-adapted clip, *Proc. of IEEE Computer Vision and Pattern Recognition*, pp. 7061–7070 (2023).

[6] Lin, W., Ma, Y., Sun, X., He, S., Ji, J., Cao, L. and Ji, R.: Hrseg: High-resolution visual perception and enhancement for reasoning segmentation, *Proc. of ACM International Conference Multimedia*, pp. 3202–3211 (2025).

[7] Liu, C., Ding, H. and Jiang, X.: Gres: Generalized referring expression segmentation, *Proc. of IEEE Computer Vision and Pattern Recognition*, pp. 23592–23601 (2023).

[8] Liu, Y., Peng, B., Zhong, Z., Yue, Z., Lu, F., Yu, B. and Jia, J.: Seg-Zero: Reasoning-Chain Guided Segmentation via Cognitive Reinforcement, *arXiv preprint arXiv:2503.06520* (2025).

[9] Liu, Y., Qu, T., Zhong, Z., Peng, B., Liu, S., Yu, B. and Jia, J.: VisionReasoner: Unified Visual Perception and Reasoning via Reinforcement Learning, *arXiv preprint arXiv:2505.12081* (2025).

[10] Rasheed, H., Maaz, M., Shaji, S., Shaker, A., Khan, S., Cholakkal, H., Anwer, R. M., Xing, E., Yang, M.-H. and Khan, F. S.: GLaMM: Pixel Grounding Large Multimodal Model, *Proc. of IEEE Computer Vision and Pattern Recognition* (2024).

[11] Ravi, N., Gabeur, V., Hu, Y.-T., Hu, R., Ryali, C., Ma, T., Khedr, H., Rädle, R., Rolland, C., Gustafson, L., Mintun, E., Pan, J., Alwala, K. V., Carion, N., Wu, C.-Y., Girshick, R., Dollár, P. and Feichtenhofer, C.: SAM 2: Segment Anything in Images and Videos, *arXiv preprint arXiv:2408.00714* (2024).

[12] Ren, T., Liu, S., Zeng, A., Lin, J., Li, K., Cao, H., Chen, J., Huang, X., Chen, Y., Yan, F. et al.: Grounded sam: Assembling open-world models for diverse visual tasks, *arXiv preprint arXiv:2401.14159* (2024).

[13] Schulman, J.: Approximating KL Divergence, http://joschu.net/blog/kl-approx.html (2020). Accessed: 2025-12-27.

[14] Shao, Z., Wang, P., Zhu, Q., Xu, R., Song, J., Bi, X., Zhang, H., Zhang, M., Li, Y., Wu, Y. and Guo, D.: DeepSeekMath: Pushing the Limits of Mathematical Reasoning in Open Language Models, *arXiv preprint arXiv:2402.03300* (2024).

[15] Wang, H., Wu, X., Huang, Z. and Xing, E. P.: High-frequency component helps explain the generalization of convolutional neural networks, *Proc. of IEEE Computer Vision and Pattern Recognition*, pp. 8684–8694 (2020).

[16] Wang, X., Zhang, S., Li, S., Kallidromitis, K., Li, K., Kato, Y., Kozuka, K. and Darrell, T.: Segllm: Multi-round reasoning segmentation, *arXiv preprint arXiv:2410.18923* (2024).

[17] Wang, Z., Guo, X., Stoica, S., Xu, H., Wang, H., Ha, H., Chen, X., Chen, Y., Yan, M., Huang, F. et al.: Perception-Aware Policy Optimization for Multimodal Reasoning, *arXiv preprint arXiv:2507.06448* (2025).

[18] Wei, C., Zhong, Y., Tan, H., Liu, Y., Zhao, Z., Hu, J. and Yang, Y.: HyperSeg: Towards Universal Visual Segmentation with Large Language Model, *arXiv preprint arXiv:2411.17606* (2024).

[19] Yang, S., Qu, T., Lai, X., Tian, Z., Peng, B., Liu, S. and Jia, J.: LISA++: An Improved Baseline for Reasoning Segmentation with Large Language Model, *arXiv preprint arXiv:2312.17240* (2023).

[20] You, Z. and Wu, Z.: Seg-R1: Segmentation Can Be Surprisingly Simple with Reinforcement Learning, *arXiv preprint arXiv:2506.22624* (2025).

[21] Yu, L., Poirson, P., Yang, S., Berg, A. C. and Berg, T. L.: Modeling context in referring expressions, *Proc. of European Conference on Computer Vision*, pp. 69–85 (2016).

[22] Yu, Q., Zhang, Z., Zhu, R., Yuan, Y., Zuo, X., Yue, Y., Dai, W., Fan, T., Liu, G., Liu, L., Liu, X., Lin, H., Lin, Z., Ma, B., Sheng, G., Tong, Y., Zhang, C., Zhang, M., Zhang, W., Zhu, H., Zhu, J., Chen, J., Chen, J., Wang, C., Yu, H., Song, Y., Wei, X., Zhou, H., Liu, J., Ma, W., Zhang, Y., Yan, L., Qiao, M., Wu, Y. and Wang, M.: DAPO: An Open-Source LLM Reinforcement Learning System at Scale, *arXiv preprint arXiv:2503.14476* (2025).

[23] Zou, X., Dou, Z.-Y., Yang, J., Gan, Z., Li, L., Li, C., Dai, X., Behl, H., Wang, J., Yuan, L. et al.: Generalized decoding for pixel, image, and language, *Proc. of IEEE Computer Vision and Pattern Recognition*, pp. 15116–15127 (2023).

[24] Zou, X., Yang, J., Zhang, H., Li, F., Li, L., Wang, J., Wang, L., Gao, J. and Lee, Y. J.: Segment everything everywhere all at once, *Advances in Neural Information Processing Systems*, Vol. 36, pp. 19769–19782 (2023).