ABCI 3.0開発加速利用 （2025年度）成果概要（公開用）

課題名：
大規模言語モデルのマルチモーダル適応に関する研究

実施時期：2025年度
所属機関名：東北大学
代表者氏名：斉藤いつみ

成果概要：
画像を入力してベクター形式のダイアグラムを生成するタスクにおいて，複数のフォーマットを対象として包括的に評価可能なベンチマークを構築し，それぞれのフォーマットに対して既存の視覚言語モデルがどの程度正しく変換できるかの評価を実施した．これらのベンチマークの構築にあたっては，GithubやWikimedia等のWeb上から収集を行った．また，収集したデータを用いたfinetuningを行うことで，一部のフォーマットについては大幅に性能を向上できることを確認した．

成果のポイント：

画像をベクターフォーマットに変換するタスクにおいて，Graphviz, Mermaid, Plantuml, draw.io, TikZ, SVGなど複数のフォーマットでベンチマークを作成し，各モデルにおいて評価を実施した．また，Qwen3VL 8B Instructモデルをfine-tuningした結果も比較した．
その結果，Graphviz, Mermaid, PlantUML, draw.ioについては大幅に性能が向上し，proprietaryモデルにも迫る性能を示した．このことから，これらのフォーマットについては公開データが十分に存在しないことから，データ自体を整備し学習させることの効果が高かったと考えられる．一方，TikZやSVGについては学習を行っても性能の向上があまり見られなかった．こちらについては，公開モデルがすでにWeb上のデータで学習を行っている可能性や，SVGのコードの煩雑さが影響していると考えられる．これらの点についてはより詳しく調査を行う予定である．

	Graphviz		Mermaid		Plantuml		draw.io		TikZ		SVG	
	ImgSim	CSR	ImgSim	CSR	ImgSim	CSR	ImgSim	CSR	ImgSim	CSR	ImgSim	CSR
Proprietary models												
Claude 4.5 sonnet	81.0	0.985	**86.5**	**0.970**	**80.3**	**0.944**	**84.4**	**0.968**	67.8	0.858	**49.7**	0.991
GPT-5	79.2	0.988	67.6	0.781	43.3	0.512	77.6	0.924	**74.9**	0.963	48.1	0.994
Open-source models												
Qwen3VL 4B	65.7	0.870	69.2	0.807	46.8	0.589	0.0	0.000	56.8	0.861	43.6	0.921
Qwen3VL 8B	68.2	0.867	75.3	0.873	50.3	0.615	0.0	0.011	58.6	0.877	40.0	0.864
InternVL3.5 1B	2.8	0.070	41.3	0.608	2.2	0.060	0.0	0.000	11.7	0.499	20.8	0.959
InternVL3.5 2B	7.0	0.110	44.1	0.608	13.1	0.170	0.0	0.000	22.7	0.697	21.0	**1.000**
InternVL3.5 8B	41.2	0.578	63.5	0.776	38.2	0.479	12.2	0.294	49.5	0.939	30.5	0.931
InternVL3.5 14B	56.5	0.779	63.9	0.766	36.5	0.456	30.8	0.593	50.1	0.945	40.1	0.962
InternVL3.5 38B	53.4	0.709	67.8	0.796	50.3	0.628	64.5	0.914	57.0	**0.982**	40.1	0.969
FT models												
Qwen3VL 8B FT	**88.8**	**0.994**	79.0	0.868	76.4	0.864	68.8	0.805	63.1	0.870	43.2	**1.000**

成果についてより詳細な情報を提供しているWebページ、発表論文などの情報：