FY2025/2025年度

Multimodal Visual Intelligence Team
Takayuki Okatani
マルチモーダル視覚知能チーム　岡谷 貴之

RIKEN　AIP Center for Advanced Intelligence Project

成果：自然災害リスク評価を行う画像理解AI

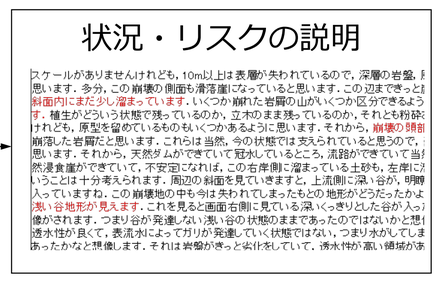
- 目標
  - 専門家が不在の現場をAIがサポート
  - 「専門家＋AI」でさらに的確な判断へ
  - 他のAIとの対話を通じた意思決定支援

災害現場の画像を理解する対話型AI

対話入力
(人や他のAIの関心)

AI

状況・リスクの説明



当チームのミッションと今年度の成果

ミッション

- 視覚を主な入力モダリティとするAIのフロンティアを前進させる
- マルチモーダルAIを実社会の多様な課題解決に応用する

主な研究成果（2025年発表）

(本ポスターの内容)

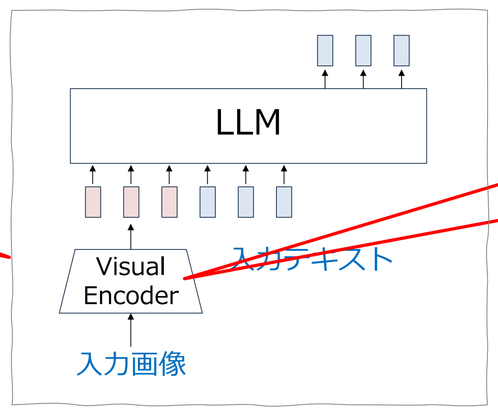
- 上空画像から斜面崩壊リスク評価を行うAI [Areerob+, Computer-Aided Civil and Infrastructure Engineerin誌]
- 密な予測タスクのための画像特徴の自己教師学習 [Liu+, International Conference on Machine Learning]
- 未知クラス物体検出の定式化の再考 [Hosoya+, International Journal of Computer Vision誌]
- Gaussian SplattingとSLAMの統合 [Bai+, International Journal of Computer Vision誌] 他

主な課題

- 専門家が持つマルチモーダル（非言語的）専門知をどのようにデジタル化し、AIに学習させるか？
- マルチモーダルAIに十分な視覚情報処理能力を与えるには？

課題Ⅰ
専門知学習：
データ・方法

「専門家の知識」をデジタル化・AIに学ばせるには？

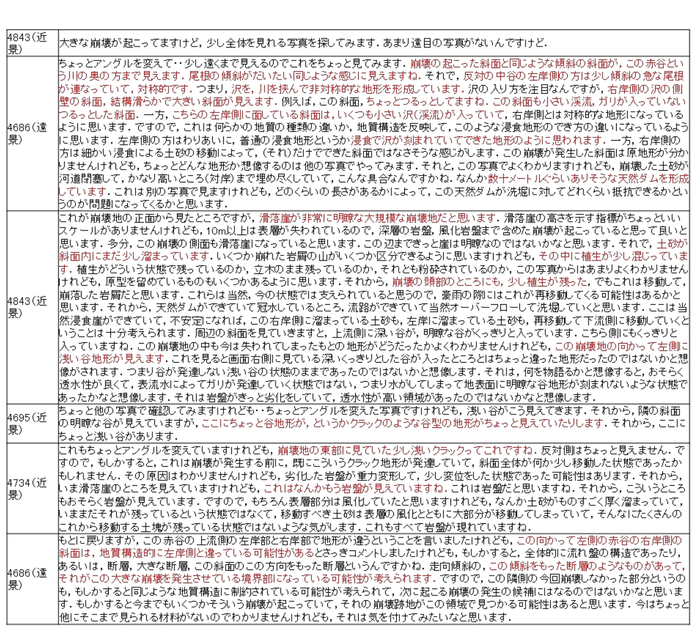


課題Ⅱ
視覚情報処理能力の不足

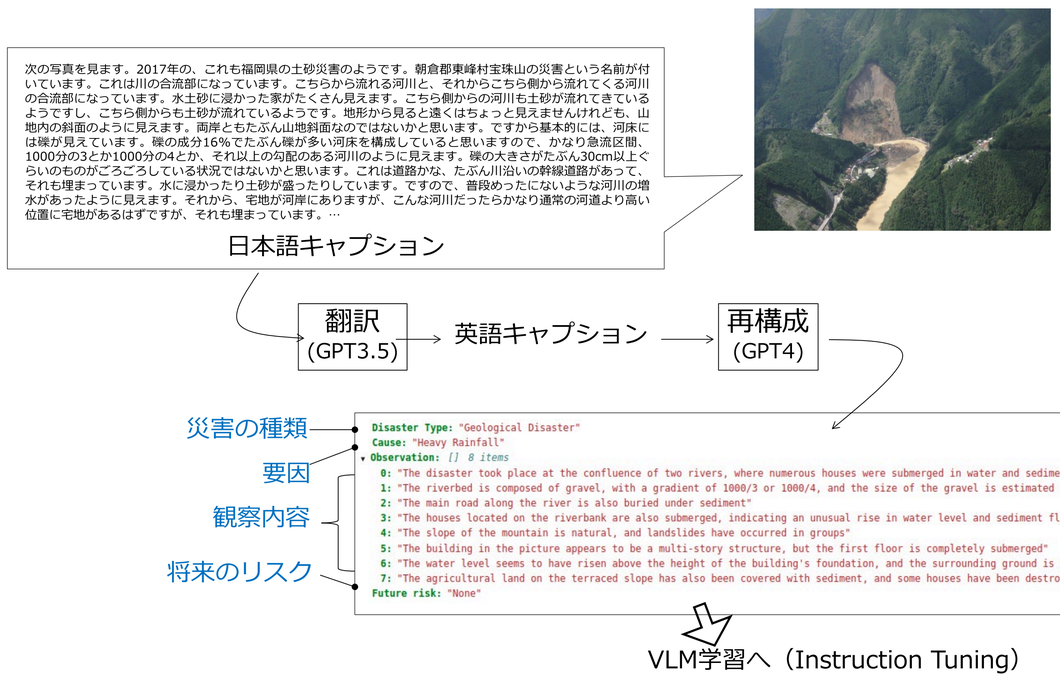
何をどのように取り出すか？

学習データの取得～構築

- 専門家に自由形式で説明してもらう → 音声認識でテキスト化

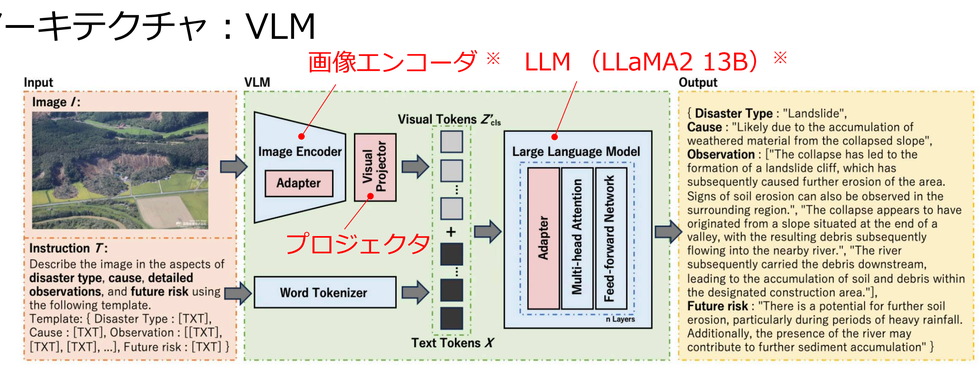


- LLMによる翻訳・要約・フォーマット



AIの構造

- アーキテクチャ：VLM



- 学習：教師ありファインチューン

学習データ

Dataset	Average Length of Tokens	Unique Words Count
Original Data	172.60	978
Augmented Data	169.47	1332

学習プロンプト

Instruction: The descriptions below explain the conditions at a sediment disaster site investigated from a disaster response helicopter. Please analyze the possible future risks associated with this disaster and include the detailed reasons for your conclusions. Describe within 150 words.

Input: {Disaster Type: [VQA_prediction], Cause: [VQA_prediction], Observation: [[VQA_prediction], [VQA_prediction], [VQA_prediction],...] }.

Output:
Future risk: [GT future risk]

性能評価

- 専門家の記述と意味的な類似度を計算(GPT-4利用)
  - VQA-LLM:VQA単体モデル＋LLM, MLLM:上記モデル

GPT-4による自動評価(10点満点評価)

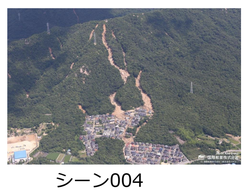
Model	Fine-tuning	GPT-4 similarity score				
		Disaster type	Cause	Observation	Future risk	Overall
Visual question answering-large language model (VQA-LLM)	✗	7.33	5.83	7.33	**8.00**	7.12
VQA-LLM	✓	7.33	5.33	7.33	7.67	6.92
Multimodal large language model (MLLM)	✗	4.33	4.50	3.67	4.67	4.29
MLLM	✓	**7.33**	**6.83**	**7.50**	7.50	**7.29**

専門家2名による評価(5点満点評価)

Model	Fine-tuning	Expert evaluation score				
		Disaster type	Cause	Observation	Future risk	Overall
Visual question answering-large language model (VQA-LLM)	✗	3.83	3.00	**4.00**	**4.17**	**3.75**
VQA-LLM	✓	3.83	3.00	3.50	3.17	3.38
Multimodal large language model (MLLM)	✗	1.50	3.00	2.67	2.17	2.33
MLLM	✓	**4.00**	**3.50**	3.83	3.67	**3.75**

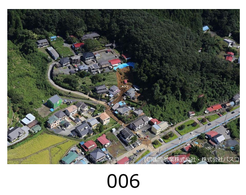
結果の例

項目別スコア(10点満点)



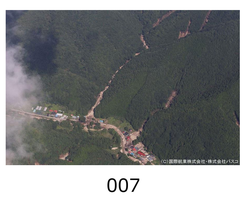
シーン004

8 Disaster Type: "Landslides"
6 Cause: "Human development that has altered the terrain's natural drainage and caused soil erosion from surrounding slopes."
Observation: [] 5 items
0: "The picture shows a mountainous region where numerous small landslides have occurred"
1: "The landslides appear to have originated from the ridge lines of the valley systems"
7 2: "The soil from the landslides has flowed into the low-lying areas and created temporary lakes"
3: "It is likely that the area experienced heavy rainfall at the time the picture was taken"
7 4: "The presence of trees suggests that the slopes have a relatively gentle incline"
Future risk: "The area is likely prone to similar landslides in the future, particularly during heavy rainfall."



006

4 Disaster Type: "Geological disaster (landslide)"
3 Cause: "Not specified"
Observation: [] 7 items
0: "The village of Hiroga in Sakurada town, Miyazaki prefecture was hit by a geological disaster on November 2019"
1: "The slope has collapsed and has reached the residential area, causing damage to houses and making them uninhabitable"
2: "The slope has also affected the road, and there are reports of one death and several injured persons"
3 3: "The aftermath shows that the slope has moved downstream, causing damage to houses and making them uninhabitable"
4: "The slope has a large incline and is composed of weathered material, suggesting that it is a old landslide slope"
5: "The slope has a distinct border and is covered with vegetation, making it difficult to observe from the air"
6: "There is a possibility that the slope is a remnant of an older landslide"
1 Future risk: "None"



007

8 Disaster Type: "Landslides, Debris flow"
9 Cause: "The area has experienced heavy rain, and the slope is steep with exposed sedimentary rocks, which has likely caused the land slides and debris flow."
Observation: [] 5 items
0: "The picture depicts a mountainous slope in a mountainous region of Hiroshima Prefecture, Japan, with clear evidence of landslid es and debris flow."
1: "The debris flow has reached the downstream area and is still flowing."
8 2: "The slope is steep with exposed sedimentary rocks, and there are signs of cracks, which suggests that the area has likely exper ienced weathering and fracturing."
3: "The vegetation on the slope is sparse, and the surface is covered in loose material, indicating that the soil has likely slid d own the slope."
4: "The color of the soil in the debris flow is different from the surrounding soil, suggesting that the sediment contained by the cracks has been transported by the debris flow."
9 Future risk: "The area is likely to experience similar landslides and debris flows in the future, especially during heavy rainfall, as the slope is steep with exposed sedimentary rocks and cracks that have likely caused the landslides."