特集≫≫　建設 DX と生成 AI

土砂災害現場の二次災害リスク評価における大規模言語モデルとマルチモーダル AI の活用

全　邦釘・岡谷貴之・島田　徹

本研究では，土砂災害現場における二次災害リスクを効率的かつ精確に評価するための大規模言語モデル（LLM）活用手法を提案する。専門家による現地評価は，アクセス困難，人材不足，即時対応の難しさなどの課題がある。これに対し，本研究では航空写真から情報を抽出する VQA（Visual Question Answering）と LLM を組み合わせた手法，および画像と言語を統合処理するマルチモーダル AI を開発・評価した。実験の結果，両手法とも専門家の知見を再現可能であり，特にマルチモーダル AI はより高い精度でリスク評価を行えることが示された。提案手法は遠隔かつリアルタイムでの評価を可能にし，将来的には建機搭載カメラとの連携により，土砂災害対応の効率化・高度化に貢献することが期待される。

キーワード：土砂災害，大規模言語モデル（LLM），Visual Question Answering（VQA），マルチモーダル AI，航空写真解析

1. はじめに

日本は地形的・気象的特性から土砂災害が頻発する国であり，年間約1,000件の土砂災害が発生している[1)]。これらの災害は主に土石流，地すべり，崖崩れの３つに分類され，人命や財産に深刻な被害をもたらす。土砂災害発生後，復旧作業を迅速に行うためには二次災害のリスク評価が不可欠である。

現在，土砂災害現場における二次災害リスクの評価は，主に専門家による現地調査に依存している。しかし，この従来手法には３つの重大な課題がある。第一に，災害発生直後の現場は危険かつアクセスが困難であり，専門家が迅速に現地調査を行うことができない場合が多い。第二に，広域災害では限られた数の専門家で多数の現場を評価する必要があり，人的リソースの制約から即時対応が難しい。第三に，土砂災害リスク評価には地質学や土木工学などの専門知識が必要であり，十分な数の専門家を確保することが困難である。

近年，人工知能（AI）技術の発展により，画像認識や自然言語処理の分野で目覚ましい進歩が見られる。特に大規模言語モデル（Large Language Models, LLM）[2)] は膨大なテキストデータから学習し，人間に近い言語理解・生成能力を獲得している。また，Visual Question Answering（VQA）[3)] 技術は画像を分析し，その内容に関する質問に回答する能力を持つ。これらの技術を土砂災害リスク評価に応用することで，従来手法の課題を克服できる可能性がある。

本研究では，土砂災害現場の二次災害リスクを適切に評価するための AI モデル開発を目的とする。具体的には，航空写真から得られる視覚情報と専門知識を組み合わせて二次災害の危険性を評価するシステムの構築を目指す。このようなシステムが実現すれば，現場へのアクセス困難性を克服し，限られた専門家リソースの効率的活用が可能となる。また，将来的には建機に搭載されたカメラからのリアルタイム映像を用いた評価も視野に入れており，災害対応の迅速化と高度化に貢献することが期待される。

本研究では，二つの異なるアプローチを検討する。一つ目は，VQA と LLM を組み合わせたハイブリッドモデルであり，画像から情報を抽出し，それを LLM に入力して二次災害リスクを推論する。二つ目は，画像と言語の両方を同時に処理できるマルチモーダル AI（Multimodal Large Language Model, MLLM）を用いたアプローチである。これらの手法を用いて，土砂災害現場の画像から災害タイプ，原因，観察事項，将来リスクを包括的に評価するモデルの開発と評価を行う。

本論文の構成は以下の通りである。第２章では提案手法の詳細について説明し，第３章では実験設定と評価方法について述べる。第４章では実験結果の提示と考察を行い，第５章で結論と今後の課題について述べる。なお，本稿は文献（4）の概要の解説を示したも

のである。そのため結果や学習データ構築プロセスで英語を使って進めているが，基本的に日本語でも大きく精度が変わらず実現できると見込まれる。

2. 提案手法

本章では，土砂災害現場における二次災害リスク評価のための二つの異なるアプローチについて詳細に説明する。まず全体の研究フローを示し，続いて VQA-LLM ハイブリッドモデルとマルチモーダル LLM の二つの手法について述べる。

(1) 研究フロー

本研究の全体フローを**図**―1 に示す。従来の手法では，災害現場の画像から直接特徴を抽出し，それを言語モデルに入力してリスク評価を行っていた。この手法はシンプルで効率的であるが，専門知識の考慮が不足している可能性がある。一方，本研究で開発するモデルでは，災害現場の画像をまず詳細な情報をテキストに変換し，その後テキストを言語モデルに入力する。専門知識を追加して言語モデルの精度を向上させることにより，より正確で信頼性の高い二次災害リスク評価が可能となる。

本研究では，以下の二つのアプローチを採用する：

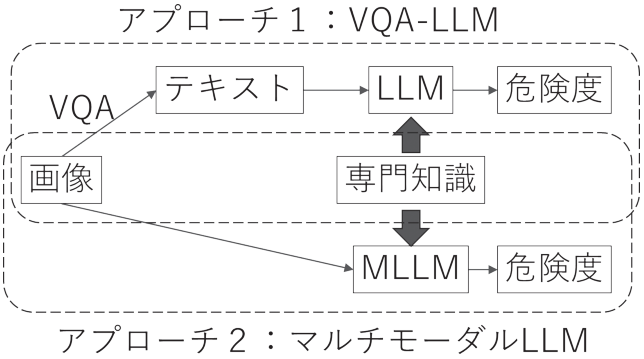
① VQA-LLM ハイブリッド：画像に対する質問応答を行う VQA モデルと，その出力を用いて推論を行う LLM を組み合わせたアプローチ

②マルチモーダル LLM：画像と言語を同時に処理できる統合モデルを用いたアプローチ

(2) VQA-LLM ハイブリッドモデル

(a) モデル概要

VQA-LLM ハイブリッドアプローチでは，土砂災害リスク評価という複雑なタスクを二段階のプロセスに分割する。これは，一度に難しい問題を解くのではなく，段階的に取り組むことで精度を高める戦略である。



図―1 本研究の全体フロー

第一段階（VQA による情報抽出）：Visual Question Answering（VQA）モデルが入力画像から直接認識可能な情報を抽出する。具体的には，災害現場の航空写真に対して「これはどのタイプの災害か？」「近くに川はあるか？」といった具体的な質問を投げかけ，それらの質問に対する回答を生成する。この段階では，純粋に画像から読み取れる客観的な情報のみを抽出することに集中する。

第二段階（LLM によるリスク予測）：第一段階で得られた情報を基に，Large Language Model（LLM）が総合的な理解とリスク予測を行う。LLM は，VQA から得られた「土砂崩れが発生している」「近くに川がある」「斜面に亀裂がある」といった情報をもとに，「将来的に河川閉塞のリスクがある」といった専門的判断を行う。この段階では，抽出された情報の関連性分析や，潜在的なリスクの推論といった高次の思考プロセスを実行する。

このアプローチの主な利点は以下の通りである。

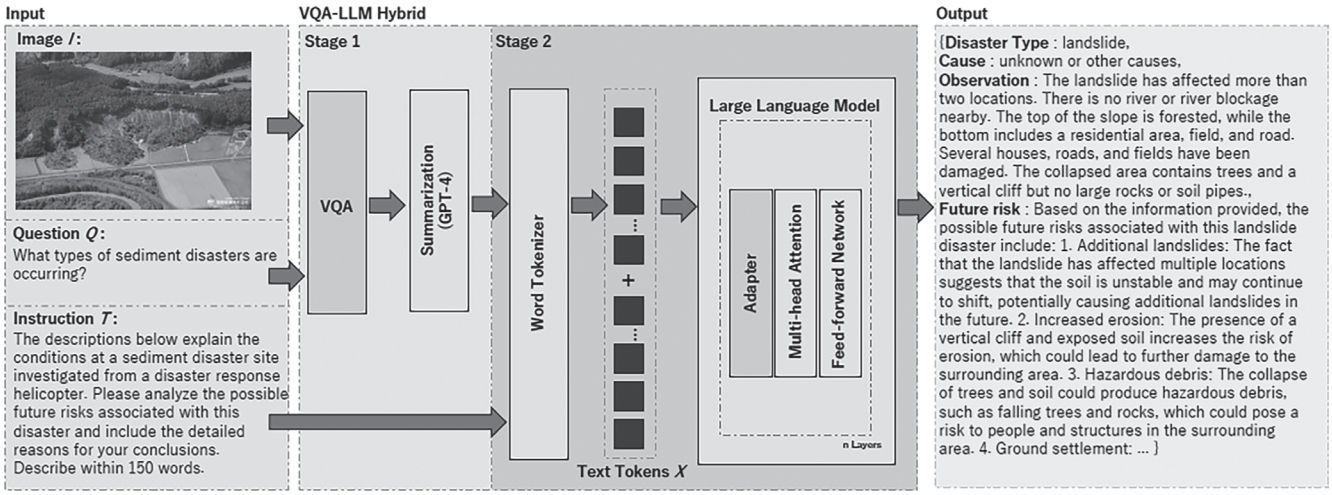
- 学習効率の向上：タスクを分割することで，各段階のモデルがそれぞれ特化した役割に集中できる。VQA モデルは画像認識に，LLM は推論と予測に特化することで，全体としての性能が向上する。
- 解釈可能性の確保：中間出力が「川がある」「斜面に亀裂がある」といった自然言語であるため，モデルがどのような情報に基づいて判断を下したのかを人間が理解しやすい。いわば，AI の「思考過程」を追跡できるようになる。

図―2 に VQA-LLM ハイブリッドモデルの全体構造を示す。プロセスをさらに詳細に説明すると，まず災害現場の入力画像と「これはどのタイプの災害か？」「近くに川はあるか？」といった一連の質問が VQA モデルに与えられる。VQA モデルは各質問に対する回答を予測し，それらの回答を構造化された形式で LLM に渡す。LLM はこれらの情報と「この災害現場の将来的なリスクを評価せよ」などのタスク指示に基づいて，最終的なリスク評価レポートを生成する。このレポートには，災害タイプ，原因，観察事項，そして将来リスクの予測が含まれる。

(b) VQA コンポーネント

VQA コンポーネントは，画像から情報を直接抽出するための重要な役割を担っている。人間の専門家が災害現場の写真を見て最初に行う観察と同様の処理を行うのがこのコンポーネントである。

本研究では，土砂災害分野の専門家との綿密な協議



図―２　VQA-LLM アプローチ

を通じて，災害現場の評価に必要な情報を抽出するための 13 の質問を設定した。これらの質問は，専門家が現場写真から読み取るべき重要な情報を網羅するように設計されている。質問の種類は多岐にわたり，以下のようなものが含まれる：

①災害タイプの分類：「どのような種類の土砂災害が発生しているか？」（土石流，地滑り，崖崩れなど）
②原因の特定：「災害の原因は何か？」（降雨，地震，その他）
③数量的質問:「災害は何か所で発生しているか？」（1 か所，2 か所，3 か所以上）
④はい／いいえの質問：「斜面の近くに川はあるか？」「河道閉塞は発生しているか？」
⑤土地利用に関する質問：「斜面の上部／下部の土地はどのように利用されているか？」

この VQA モデルにより，例えば「斜面に亀裂はあるか？」という質問に対して，画像内の斜面部分に注目し，亀裂の有無を判断することができる。

モデルの学習には，合計 372 枚の土砂災害画像を使用した。このうち 68 枚は本研究の主実験にも使用される画像である。学習データの量を増やすために，画像の一部を切り取る，反転させる，回転させるなどのデータ拡張技術を適用し，訓練データと検証データをそれぞれ2,600 枚と 400 枚に増加させた。これにより，限られた画像データからでも高い汎化能力を獲得することを目指した。

（c）LLM コンポーネント

LLM コンポーネントは，VQA から得られた情報を基に総合的な理解とリスク予測を行う。このコンポーネントは，専門家が現場情報を分析してリスクを予測する思考プロセスをシミュレートするよう設計されている。

本研究では，オープンソースの大規模言語モデルである Llama2 を fine-tuning したモデルを使用した。VQA コンポーネントからの出力である災害タイプ（例：地滑り），原因（例：豪雨），観察事項（例：斜面に亀裂がある，近くに川がある）をもとに，将来発生する可能性のあるリスク（例：河道閉塞の危険性がある）を予測するようにモデルを調整した。

LLM の fine-tuning には，専門家の思考プロセスを学習させるための特別なデータセットが必要である。このため, 以下のステップでデータセットを作成した:

①専門家の回答を模倣したデータの作成：土砂災害の専門家が行うような将来リスク予測を模倣した合成データセットを作成した。
② GPT-3.5 を活用した教師データの生成：GPT-3.5 という言語モデルを用いて，土砂災害に関する質問（例：「この地滑り現場の将来リスクは？」）と回答（例：「豪雨が続けば二次的な崩壊が発生する可能性がある」）のペアを大量に生成した。
③教師データを用いた学習：生成された質問回答ペアを教師データとして用い，LLM に土砂災害のリスク予測能力を学習させた。

学習の技術的詳細としては，ベースモデルとして Llama-2-13B-chat-hf（130 億パラメータを持つチャット用に最適化されたモデル）を使用し，QLoRA という fine-tuning 技術を適用した。QLoRA は，モデル全体ではなく一部のパラメータのみを調整することで，計算資源を効率的に使いながら高い性能を達成できる手法である。これにより，限られたデータと計算資源でも専門家レベルのリスク予測能力を持つモデルの開発が可能となった。

(3) マルチモーダル LLM

(a) モデル概要

マルチモーダル LLM アプローチでは，単一のモデルで画像理解と言語理解の両方を行う。このアプローチでは，画像エンコーダが入力画像を処理し，その結果を LLM が処理できる形式に変換して，LLM に入力する。LLM は画像情報とテキスト指示を組み合わせて最終的な予測を生成する。

図―3 にマルチモーダル LLM の全体構造を示す。モデルは，画像エンコーダ，ビジュアルプロジェクタ，LLM の三つの主要コンポーネントから構成される。画像エンコーダと LLM は事前学習されたモデルを使用し，アダプタを介して統合される。ビジュアルプロジェクタは学習可能なパラメータを持つ。

(b) モデルアーキテクチャ

マルチモーダル LLM は，LaVIN モデルをベースに構築した。画像エンコーダとしては，CLIP（Contrastive Language-Image Pre-training）で事前学習された Vision Transformer（ViT-L/14）を採用した。このコンポーネントは，インターネット上の膨大な画像とテキストのペアで訓練されており，様々な画像の特徴を理解する能力を持っている。各画像は線形変換などを通じてビジュアルトークン（AI が理解できる数値表現）に変換される。

ビジュアルプロジェクタは，画像から得られた情報を言語モデルが理解できる形式に変換する「通訳」のような役割を果たす。具体的には，画像全体の特徴を表す特別なトークン（クラストークン）を，言語モデルが処理できる形式に変換する。また，効率的に学習させるために，「アダプタ」という小さな調整用モジュールを言語モデルと画像処理部分の両方に追加した。これはちょうど，大きな機械の一部だけを交換して新しい機能を追加するようなものである。全体を作り直す代わりに，必要な部分だけを調整することで効率的に学習ができる。

(c) モデル学習

マルチモーダル LLM の学習には，「visual instruction tuning」と呼ばれる微調整方法を採用した。これは人間が先生役となって，画像を見せながら「この画像について説明してください」といった指示を出し，AI に正しい回答方法を教えるようなプロセスである。

具体的な学習手順：

① AI に「土砂災害の種類，原因，観察事項，将来リスクを説明してください」という指示と災害画像を与える

② AI が回答を生成する過程で，次の単語を予測する能力を向上させる

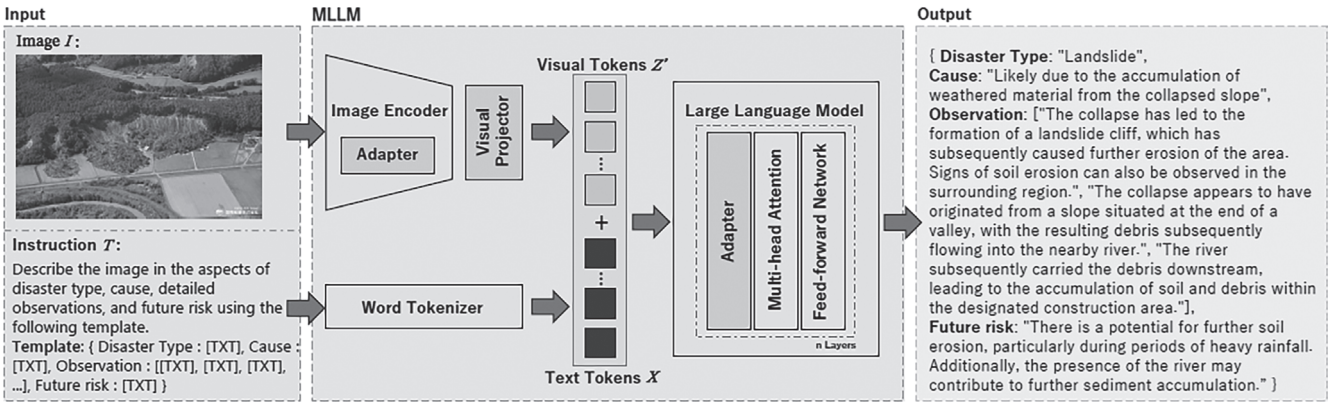
③学習を効率化するため，大部分のモデルパラメータはそのままにして，画像と言語をつなぐ部分（ビジュアルプロジェクタ）と調整モジュール（アダプタ）のみを更新する

学習データとしては，68 枚の土砂災害画像と専門家による詳細な説明文を使用した。データ量を増やすために，元の説明文を言い換えた文章も追加して，合計 136 サンプルの学習データを作成した。このような工夫により，限られたデータからでも専門家のような判断能力を持つ AI モデルの開発を目指した。

(4) データ収集と前処理

本研究で使用するデータについて説明する。データセットは日本全国の土砂災害現場を撮影した航空写真とそれに対応する専門家の注釈から構成される。

データ収集では，土砂災害調査の専門企業に所属する 8 名の専門家（全員が 30 年以上の経験と技術資格を持つ）が参加した。各画像について，2 名の専門家が画像に描写されている事象を口頭で説明し，その会話を録音した。専門家には，画像から読み取れる情報のみに基づいて注釈を行うよう指示した。録音された



図―3 マルチモーダル LLM アプローチ

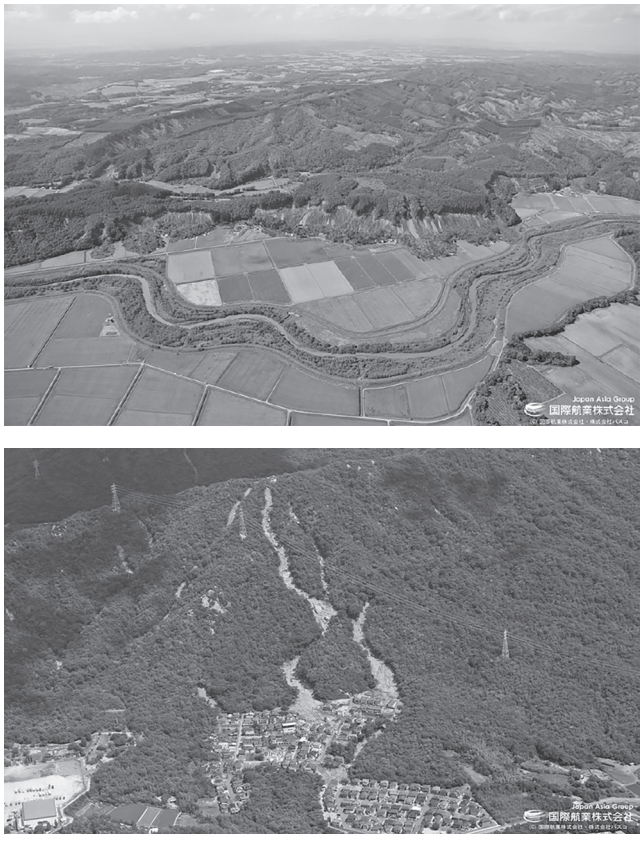
口頭説明は音声認識システムを用いてテキストに変換された。これらのテキストはGPT-3.5を用いて標準化され，英語に翻訳された。続いて，GPT-4を用いて以下の4つの要素を含む統一フォーマットに変換された:

①災害タイプ（Disaster Type）：土砂災害の種類（地滑り，崩壊など）
②原因（Cause）：災害の推定原因（降雨，地震など）
③観察事項（Observation）：画像から観察される複数の特徴や状態
④将来リスク（Future risk）：将来的に発生する可能性のあるリスク

本章では，土砂災害現場における二次災害リスク評価のための二つのアプローチ，VQA-LLMハイブリッドモデルとマルチモーダルLLMについて詳細に説明した。次章では，これらのモデルの評価方法と実験結果について述べる。

3. 実験結果

テストデータセットから選んだ2つの画像（**写真**―1）に対して各モデルがどのような評価をしたか事例を示す。なお定量的な評価は文献(4)で行っており，適宜参考にされたい。



写真―1　解析画像

(1) 事例1：複数の崩壊が発生している現場

写真―1上の画像は複数の崩壊が発生している現場を示している。この画像に対して，VQA-LLMハイブリッドモデルは，災害タイプを「Landslide」と正しく識別したが，原因を「不明」としており，この部分はVQAモデルが質問に回答できなかったことを反映している。観察事項は比較的詳細で，正解データと多くの共通点があった。将来リスクの予測は詳細かつ妥当なものだった。

マルチモーダルLLMは，災害タイプを「Landslide and creation of a natural dam」と識別し，原因も「地質･地質構造･風化度の違いによる」と推測しており，専門的な見解を示した。観察事項と将来リスクの両方で妥当な分析を行っており，特に「豪雨時に植生に覆われた土砂が下流へ移動する可能性」という具体的なリスクを指摘した。

(2) 事例2：地滑り現場

写真―1下の画像は斜面の地滑り現場を示している。この画像に対して，VQA-LLMハイブリッドモデルは，災害タイプを誤って「debris flow」と識別した。原因は再び「不明」とされたが，観察事項には「2か所以上で崩壊が発生している」「斜面の上部は森林で覆われている」など，部分的に理解可能な情報が含まれていた。将来リスク予測では，「さらなる土砂流出の可能性」など妥当な指摘があった。

マルチモーダルLLMは，災害タイプを正しく「Landslide」と識別し，原因を「人為的活動と自然侵食」と推測した。観察事項では「細長い山岳地域内の急斜面」「露出した岩石」などの特徴を正確に捉えており，将来リスクも「さらなる斜面崩壊と侵食の可能性」と妥当な予測を行った。

これらの定性的評価結果は，定量的評価と一貫しており，特にマルチモーダルLLMが専門家の判断に近い分析を行えることを示している。また，VQA-LLMハイブリッドも一定の精度を発揮できることが確認された。

4. 結論

本研究は，土砂災害現場の二次災害リスク評価を航空写真から自動化することを目指した。開発したモデルは，現場アクセスの困難性，時間的制約，専門家不足という課題に対応し，遠隔評価を可能にする。本研究では，土砂災害現場における二次災害リスク評価のための二つのアプローチ，VQA-LLMハイブリッド

モデルとマルチモーダル LLM を開発・評価した。

VQA-LLM ハイブリッドの最大の利点は解釈可能性にある。中間出力が自然言語で表現されるため，モデルの判断根拠を人間が理解しやすい。また，専門家との協議で設計された質問セットにより，必要な情報を確実に抽出できる。一方，マルチモーダル LLM は画像理解と言語生成を単一のモデルで行う統合アプローチである。事前に定義された質問に制限されず，より柔軟な情報抽出が可能だが，専門分野に特化した学習データに強く依存する。本研究の範囲内では VQA-LLM より性能がよかったが，一方で判断根拠の理解はやや難しくなる。

本研究の成果は，土砂災害対応における AI 活用の可能性を示すとともに，視覚情報と言語処理の組み合わせが専門分野の意思決定支援に有効であることを実証している。AI モデルは専門家を置き換えるものではなく，専門家判断を支援し効率的な災害対応を可能にするツールとして位置づけられるべきである。

今後の課題としては，多様なデータの拡充，マルチソースデータの統合，リアルタイム評価システムの構築，人間との協働インターフェース設計などが挙げられる。さらに将来的には，建機搭載カメラからのリアルタイム映像評価や時系列データ分析への拡張が考えられる。これらに取り組むことで，AI 技術を活用した災害対応の高度化が進み，より安全な社会の実現に貢献することが期待される。

謝　辞

本稿は，JST【ムーンショット型研究開発事業】グラント番号【JPMJMS2032】の支援を受けたものである。

J C M A

《参 考 文 献》

1）国土交通省砂防部．（2025）．令和 6 年の土砂災害（https://www.mlit.go.jp/mizukokudo/sabo/content/001877375.pdf）（2025 年 4 月 7 日アクセス）
2）乾健太郎．（2023）．ChatGPT の出現は自然言語処理の専門家に何を問いかけているか．自然言語処理，30（2），273-274.
3）Yu, Z., Yu, J., Cui, Y., Tao, D., & Tian, Q.（2019）. Deep modular co-attention networks for visual question answering. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*（pp. 6281-6290）.
4）Areerob, K., Shimada, T., Chun, P., Okatani, T., et al.（2025）. Multimodal artificial intelligence approaches using large language models for expert-level landslide image analysis. *Computer-Aided Civil and Infrastructure Engineering*（Accepted）.

［筆者紹介］

全　邦釘（ちょん　ぱんじょ）
東京大学大学院
工学系研究科　社会基盤学専攻
特任准教授

岡谷　貴之（おかたに　たかゆき）
理化学研究所
革新知能統合研究センター
チームリーダー，
東北大学　大学院情報科学研究科
システム情報科学専攻
教授

島田　徹（しまだ　とおる）
国際航業㈱
国土保全部
フェロー・事業企画担当部長（兼務）