

インフラ維持管理を支援する 画像理解 AI の開発と課題

岡谷 貴之 | 理化学研究所革新知能統合研究センター チームディレクター 東北大学情報科学研究科 教授

1 はじめに

我が国のインフラ施設は老朽化が進んでおり、その点検や維持管理の効率化は喫緊の課題となっている。このような背景のもと、AIの活用に対する期待は日増しに高まっている。著者らは理化学研究所 革新知能統合研究センターにおいて、インフラ管理ロボット技術チーム（2024年7月より「マルチモーダル視覚知能チーム」に改称）を組織し、2016年よりAIおよびロボット技術のインフラ分野への応用研究に取り組んできた。

本チームの目標は、最先端のAI・ロボット技術がインフラ維持管理にどのように貢献しうるかを明らかにし、その実現に必要な要素技術を見極めて、それらの研究開発を進めることである。具体的には、点検作業を支援し、場合によっては人の業務を代替することのできるシステムの構築を目指してきた。

研究の初期段階では、「画像を主たる入力とし、多様な分析や推論を行うAI」と「UAVを中心としたロボット技術」という二本柱でプロジェクトを開始した。ロボット技術については、東北大学情報科学研究科・田所研究室の協力を得て、橋梁床版の裏側など狭隘な空間にも自在に侵入できる球殻ドローンの開発・改良を進めた。しかしその後、ドローンの自律飛行技術が米国を中心に急速に進展している状況を鑑み、AIの研究開発に人的・予算的資源を重点的に投入する方針に切り替えた。

そこで中核的な課題として位置付けたのが、「画像を理解するAI」の実現である。これは、あるシーンの画像を入力として受け取り、その場の状況や物理現象を多角的に把握し、その内容を自然な言語で説明したり、必要に応じてユーザーと

対話したりできるAIを指す。

以下では、これまでに進めてきた取り組みの概要を述べ、最後に今後解決すべき課題について論じる。

2 画像理解 AI

われわれが目指してきた「画像理解 AI」——すなわち、単に画像を解析するだけでなく、その内容を深く理解し、言語によってユーザーと情報を共有できるAIシステム——は、2010年代当時、インフラ維持管理を対象としたものはおろか、他のいかなる分野においても存在しておらず、研究開発はほぼゼロからの出発であった。しかし、チームメンバーの努力により、2018年から2022年にかけて、いくつかの重要な成果を得ることができた。

その一つが、画像の内容を理解し、ユーザーの質問に答えることのできるAIの開発である [Nguyen et al. 2018]。このAIは、Transformer と呼ばれるニューラルネットワークのみで構成されるアーキテクチャ（図-1）を採用しており、この選択は現在のAI技術の発展にも直結するものであった。本システムは静止画を入力として、そのシーンを内部的に抽象化し、ユーザーの問い

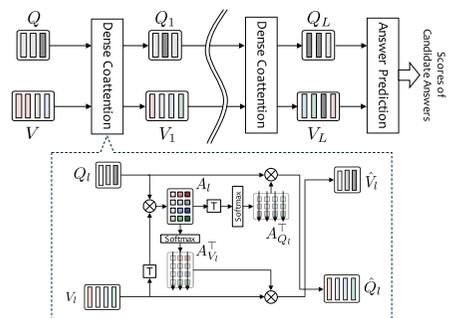


図-1 Transformer の注意機構を最初に画像理解に応用した DCN (Dense-Coattention Network)

かけに対して的確な回答を返す能力を有していた。

さらに、このAIは単発の質問応答にとどまらず、複数回の対話（ダイアログ）に対応し、文脈を維持したやりとりが可能となるよう拡張が加えられた [Nguyen et al. 2020]。また、視覚認識能力の強化や処理速度の最適化を図る研究も進められ、いくつかの成果が報告された (Nguyen et al. 2022)。これらの一連の研究は、現在の最先端AI技術の源流の一つと位置付けられ、その発展の礎となっている。

その後、大規模言語モデル (LLM) が登場し、急速な進化を遂げるとともに、視覚情報処理の能力も統合され、画像理解の精度や柔軟性は飛躍的に向上した。このような情勢の変化を受け、われわれは汎用的な画像理解AIの基盤技術の開発を一段落とし、その技術を活用して現実世界の課題解決に資する応用研究へと軸足を移すこととした。その具体例の一つに、次節で述べる橋梁点検への応用を試みた研究がある。

3 橋梁点検への応用

橋梁点検は、損傷箇所や部材の種類を点検者が目視で確認し、手作業で写真を撮影して報告書を作成するのが一般的である。報告書の作成には、熟練した技術者の知見が不可欠であり、一連の作業には多くの時間と人手を要する。さらに、経験の浅い担当者による点検では、見落としや判断のばらつきが生じやすいという問題もある。

われわれは、こうした課題の解決を目指し、橋梁点検を支援するAIを開発し、評価を行った (Kunlamai et al. 2023)。ポイントは、行政機関や点検業者によって蓄積されてきた大量の点検報

告書データを活用することを可能にしたことである。なおこの取り組みは、東京大学の全邦釘教授らとの共同研究として進めたものである。

まず、既存の報告書から約28万件にのぼる〈写真、説明文〉のペアを収集し、これを用いて画像と言語の関連性をモデルに学習させる事前学習を行った。この段階では、橋梁特有の部材や損傷パターンと、それに対応する言語表現との対応関係をモデルに習得させることを目的としており、新たなデータが少ない場合でも高精度な学習を可能とした。

次に、実際の点検現場で用いられている約90万件の〈写真、質問、回答〉ペアを用いてモデルを微調整した。ここでは、質問内容を「損傷の有無」「部材の種類」「損傷の種類」の三形式に整理し、精度評価を可能な枠組みを構築した。

評価実験では、上述の事前学習を行わない従来モデルの正答率が約79%にとどまったのに対し、事前学習を経たモデルは85～88%にまで性能が向上した。さらに微調整を施した最終モデルでは、正答率が90%を超えた。とりわけ「損傷の有無」を問う比較的単純な質問においては、96～99%というほぼ完璧な精度を示した。一方で、「損傷の種類」に関する質問では、正答率が60～65%程度にとどまった。この要因としては、出現頻度の低い部材や、視覚的に類似した損傷の判別において誤りが生じやすいことが挙げられる。

本研究は、既存データを有効に活用することで、新たなデータ収集の負担を大幅に軽減しつつ、高性能な点検支援システムの実現可能性を示した点に、意義があると考えている。

4 斜面崩壊リスク評価への応用

近年、気候変動の影響により土砂災害の発生が増加しており、UAVなどを用いて上空から撮影した画像を活用し、被災状況を迅速かつ的確に分析することが求められている。しかし、地すべりの危険度を正確に評価するには、深い地質学的知見や専門家による現地調査が不可欠であり、時間的なコストや安全性の面で大きな制約がある。

そこでわれわれは、このような業務に資する、

	Question	Answer
	What is the damage in the image?	Corrosion
	What is the member in the image ?	Bearing
	Are there any corrosion in the bearing?	Yes
	Question	Answer
	What is the damage in the image?	Degradation of anticorrosive
	What is the member in the image ?	Bearing
	Are there any corrosion ?	Yes
	Does this image show the drainage pipe ?	No

図-2 橋梁の一部の画像から、当該部分の変上等に関するユーザーの質問に答える AI

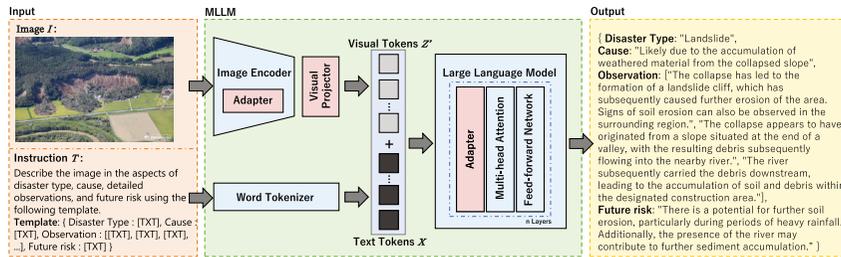


図-3 斜面崩壊現場を画像から分析し、リスク評価を実行する AI

自然言語による解説生成機能を備えた AI システムを開発し、その性能を検証した (Areerob et al. 2025)。

本研究の貢献は、大きく二点に整理される。

第一の貢献は、専門家の暗黙知を音声対話形式で収集し、その文字起こしデータを用いて、大規模言語モデル (LLM) によって「災害種別」「原因」「観察点」「将来リスク」の四つの要素に体系的に整理する手法を確立した点である。具体的には、30年以上の地すべり調査経験を有する専門家による自由な対話内容を、GPT-4によって要素別に解析し、さらに「言い換え」によるデータ拡張を行うことで、限られたデータ量でも効果的に AI を学習させられるようにした。

第二の貢献は、得られたデータをもとに、画像と言語を統合的に扱う二種類のマルチモーダル AI アーキテクチャを提案し、その比較検証を行った点にある。一つは「VQA LLM ハイブリッド」と呼ばれる二段階処理モデル、もう一つはマルチモーダル大規模言語モデル (MLLM) である。評価の結果、それぞれのモデルは特性に応じた強みを示し、目的やニーズに応じて最適な AI 設計を選択可能であることを示した。

今後は、地理的・地質的に多様な現場データを収集し、さらなる性能向上を図るとともに、その有効性を検証することで、災害対応の迅速化およびリスク管理への一層の貢献が期待される。

5 おわりに

近年の AI の研究は、大規模言語モデル (LLM) を中核に据えることで大きく進展した。とりわけ、大量のデータを用いた事前学習とモデル規模

の拡大によって性能が向上した点に特徴がある。米中のビッグテック企業を中心に研究開発が加速する中、一部では「汎用人工知能 (AGI) 実現へのカウントダウン」が語られ、AI が仕事を奪うという懸念も広がっている。

しかしながら、AI は著しい進歩を遂げつつも、依然として、できることとできないことの差は大きい。たとえば、コーディングや数学といった「言語で閉じた世界」では、人間をしのぐ水準に達しつつあるが、実世界の複雑な問題解決はなお人間の水準には (場合によって遠く) 及ばない。

画像理解 AI も同様であり、日常物体の認識などでは実用レベルに達している一方、橋梁点検や災害リスク評価といった高度に専門的な領域の問題に対しては、熟練技術者の代替が見通せる状況にはない。

今後の課題は二つある。一つは、専門家が現場で培う「暗黙知」を AI にどう学ばせるかという点である。AI は、文書化された知識は吸収できるが、直観や視覚的判断基準のように言語化しにくい知見は難しい。

もう一つは、画像の深い理解を可能にする仕組みの構築である。現在の AI は、物体の検出や名称の特定には優れるが、空間構造や時間的変化、因果関係を把握する「世界モデル」に基づく推論は苦手である。たとえば、目の前のコンクリート床版のひび割れを認識した後、それが橋全体の健全度にどのように影響を与えるかを推論するといったことは困難である。

こうした課題の克服が、今後の AI 研究の重要な方向性である。決定的な解法はまだないため、当面は、既存技術を組み合わせたハイブリッドな枠組みの構築など、現実的なアプローチを模索することになろう。