

ローカル LLM を用いた画像分類点検支援システムの開発と評価

長岡工業高等専門学校
長岡工業高等専門学校

○柄沢 修宇
正会員 井林 康

1. はじめに

平成 26 年から義務化された橋梁等の道路施設点検は、令和 6 年度から 3 巡目に入り、道路管理者ごとに様々な仕様で膨大な点検・診断のデータが蓄積されており、令和 4 年からは全国道路施設点検データベース¹⁾として有料で公開されている。また、我が国では、少子高齢化による土木技術者数の不足、維持管理が必要な施設等の増加等で各種業務における効率化や省力化が求められている。この課題を解決するため点検等の維持管理業務で構造物の損傷診断に AI の画像認識技術を用いるなど、土木分野においても AI 技術を利用する事例が増加している。

一方で、Chat GPT や Gemini をはじめとした大規模言語モデル（Large Language Model, 以下 LLM）を用いた自然言語処理技術の発展は目覚ましい。LLM とは膨大なテキストデータと高度なディープラーニング技術を用いて構築された自然言語の理解や生成を可能とした AI 技術である。そのため、メール自動生成や文章の要約生成、チャットボットなど様々な応用が進んでいる。土木分野においても自然言語処理技術への期待は大きく、近年実用に向けた検討が増加している。しかし、土木分野における自然言語処理技術の適用事例はまだ少ない。

本研究で注目したのは、LLM の中でもローカルで動かすローカル LLM である。ローカル LLM とは、クラウドサーバーではなく、ユーザー自身の PC やサーバー上で動作する LLM のことであり、最大の特徴は、データ処理がローカル環境内で完結し、外部に情報が送信されないため、高いプライバシー保護を実現できる点にある。さらに、専門的な知識や社内独自のノウハウを学習させることで、特定分野に特化した高度な回答が可能となる。これらの特徴により、社内の機密情報を安全に扱いながら、利用環境や業務内容に適した LLM を活用できるという利点がある。デメリットとして、モデルサイズによっては高性能な GPU や大容量のメモリが必要となり、既存の PC 環境では動作しない、あるいは十分な性能を発揮できない場合がある。

本研究では全国道路施設点検データベースから取得した点検調書、損傷写真や点検要領等をローカル

LLM と組み合わせた新たな点検支援の形を検討し、維持管理業務の効率化を目指す。

2. LMStudio による検討

2.1 LMStudio

LMStudio²⁾は、ローカル環境で LLM を簡易的に実行できるプラットフォームである。Mac や Windows で動作し、GUI ベースで操作できるため非エンジニアでも手軽に試すことができる。多くのモデルをダウンロードして利用することができる。ここでは LMStudio ではモデルのパラメータ数における挙動の比較を行った。本研究で扱う PC の性能はメモリが 96GB、チップが Apple M3 Ultra の Mac Studio である。

2.2 モデルの比較

モデルは google/gemma-3-4b と google/gemma-3-27b を利用した。Gemma は Google が開発している生成 AI である Gemini の技術を活用した、軽量のオープンソースのローカル LLM である。Gemma3 はその中でもテキストに加えて画像を処理することもできる。モデル名に含まれる”b”は billion（10 億）を意味しており、40 億個のパラメータを持つモデルを 4b と表記する。一般に、パラメータ数が多いモデルほど高精度な推論や複雑な文脈理解・文章生成が可能になる一方で、必要とされるメモリ使用量や計算負荷が増大し、実行環境や応答時間に制約が生じる。そのため、用途や運用環境に応じて適切なモデル規模を選択することが重要である。

2.3 結果

1 つ目の質問として、橋梁の損傷画像を添付し、損傷部分および損傷箇所の特性が可能かを検証した。その結果、4b モデルと 27b モデルの回答には、文章量および内容の詳細さに大きな差が見られた。回答を表 1、添付した損傷画像を図 1 に示す。4b モデルでは損傷の存在自体は指摘できているものの、説明は比較的簡潔であり、詳細な状況把握には限界があると考えられる。一方、27b モデルでは、損傷が発生している部位の構造的な位置関係や、周辺部材との関係性についても言及されており、画像から読み

Table1 Comparison of responses from the 4b and 27b models

Gemma3 4b	画像から判断するに橋梁の桁には以下の損傷があります。 ・ 桁の腐食：桁の表面に広範囲に腐食（錆）が発生しています。特に、桁の上面と側面に著しい腐食が見られます。 ・ 腐食による損傷：腐食によって、桁が変形し、一部が欠損しています。 ・ 損傷の場所：桁の上面と側面に、腐食と欠損が集中しています。 これらの損傷は、橋梁の安全性を脅かす可能性があり、早急な点検と補修が必要です。
Gemma3 27b	この画像から確認できる橋梁桁の損傷部分と場所は以下の通りです。 ・ 塗装の剥がれと錆び：桁全体にわたって赤色の塗装が大きく剥がれており、その下に広範囲な錆が発生しています。特に、鋼材の表面が露出している箇所は錆びが目立ちます。 ・ 腐食による鋼材の変形：桁の下部、特にリベット周辺や接合部に沿って、鋼材が変形しているように見えます。これは腐食によって鋼材の強度が低下し、形状が変化した可能性があります。 ・ リベット周辺の損傷：リベット周辺に錆や腐食が著しく、リベット自体も劣化している可能性があります。 ・ 鋼材の欠損：桁の一部に、小さな穴や欠損が見られます。これは腐食が進行し、鋼材が失われた状態です。 ・ 水漏れ跡：塗装の剥がれや錆びは、長期間にわたる水漏れによって腐食が進んだ結果である可能性があります。 これらの損傷は、橋梁の構造的な安全性を脅かす可能性があるため、早急な調査と補修が必要です。特に、リベット周辺の損傷や鋼材の欠損は、橋梁全体の強度に大きな影響を与える可能性があります。



Figure 1 Damage image

取れる情報を多角的に解釈できていることが確認できた。

2 つ目の質問では、橋梁の材種が鋼橋であるかコンクリート橋であるかを画像から判別できるかを検証した。鋼橋およびコンクリート橋の画像をそれぞれ 5 枚ずつ用いた結果、4b モデルでは正答率が 7 割であったのに対し、27b モデルではすべての画像を正しく判別し、正答率は 10 割となった。

以上の結果より、モデルのパラメータ数が増加することで、画像から抽出できる特徴量が増え、より詳細な回答がされる傾向が確認された。特に 27b モデルは、橋梁の材種判別や損傷状況の説明において高い表現力を示しており、ローカルで動作する LLM でも有効性が示唆される。

3. FileMaker による実装の検討

3.1 FileMaker

FileMaker はプログラミング知識が少なくても、ドラッグ＆ドロップなどの直感的な操作で、自社専用の業務用アプリやデータベースを開発・運用でき

るローコードプラットフォームであり、本研究室でも橋梁のタブレット点検システムの開発等に利用されている。令和 7 年 7 月にリリースされた FileMaker2025³⁾では AI を活用した新たな機能が追加され、そのうち LLM と RAG（検索拡張生成）⁴⁾の仕組みを活用し、ドキュメントに基づいた回答を生成する機能を使用した。

3.2 RAG

RAG とは検索（Retrieval）と生成（Generation）を組み合わせた AI の仕組みである。通常の LLM は、あらかじめ学習された知識の範囲しか回答できない。そのため、最新の情報への対応や専門的な内容について知識が少ない場合がある。こうした課題を解決するために RAG でのアプローチを行った。仕組みとしては、ユーザーの質問をベクトル化し、事前にベクトル化された文書（PDF から抽出したテキスト等）と照らし合わせて、意味的に近いものを検索する。検索で得られた文書（コンテキスト）をもとに、LLM が自然な形で回答を生成する。RAG のイメージを図 2 に示す。メリットとしては常にデータを参照できるため、最新情報に更新でき情報が古くなる心配がないことや、社内の専門知識やノウハウを学習させることで、専門的な内容にも対応可能になることである。また、根拠のある情報のみを基に回答を生成するため、企業にとって最も重要である信頼性の向上も期待できる。

本研究ではテキスト入力と複数の PDF ファイルから抽出したテキストを活用して、ユーザーの質問に回答するチャットボットを作成した。操作画面を図 3 に示す。RAG 情報のボタンで追加したデータが閲覧でき、データを一括削除または個別で削除できるボタンも各々作成した。本研究ではこれを使い、チャットボットへの質問内容ごとでの回答の傾向と

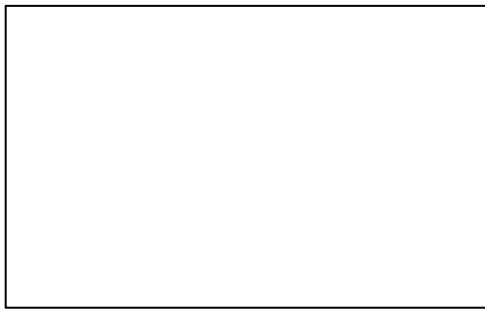


Figure 2 RAG image



Figure 3 Chatbot operation screen

埋め込みモデルや RAG データの量、内容、組み合わせによる違いについて比較を行った。

3.3 RAG データ及びモデル

本研究では、RAG データとして全国道路施設点検データベースの有料版で取得可能な PDF ファイル形式の道路橋点検調査および国土交通省が出している道路橋定期点検要領を使用した。また、モデルはテキスト埋め込みモデルに BAAI/bge-m3、テキスト生成モデルに google/gemma-3-12b-it を使用した。BAAI/bge-m3 は北京智源人工知能研究院が出したモデルで意味の近さ・文書の構造・キーワードの一致という異なる方法で同時に検索できる特徴がある。

3.4 結果

3.4.1 各質問による回答の比較

10 橋分の点検調査を RAG データとして登録したうえで得られた各質問に対する回答結果を表 2 に示す。

結果として、特定の橋梁を指定して諸元情報を質問した場合や、諸元情報を条件として橋梁名を回答させる質問に対しては、概ね適切な回答が得られた。また、各損傷に対する代表的な補修工法についても、点検調査の記載内容に基づいた補修工法が適切に提示されており、RAG による文書参照が有効に機能していることが確認された。一方で、健全度に関する質問については、いずれのケースにおいても正確

Table 2 Response results for each question

質問	回答
〇〇橋について	一部の諸元情報が記載
橋梁 ID で検索	指定した橋梁名が記載
管理者で検索	橋梁名が複数記載 不足あり
〇〇橋の健全度	健全度が記載 誤答あり
健全度 II の橋梁	複数記載 誤答、不足あり
〇〇橋の点検時の注意点	注意点を複数記載 誤答あり
腐食がある橋梁	誤答
損傷が一番激しい橋梁	誤答
各損傷の代表的な補修工法	各損傷の補修工法が記載

な回答が得られなかった。さらに、特定の損傷が存在する橋梁の抽出や、損傷程度が大きい橋梁を対象とした質問に対しても、期待される回答を十分に得ることはできなかった。これらは点検調査の様式上、表形式のため埋め込みの際に単語ごとが結び付けられづらいことが考えられ、特に健全度と損傷内容は点検調査を見たときに諸元情報と異なる表に記載されていたため誤答になってしまったと考える。

また、前述の質問において橋梁名を質問した際、本来は橋梁名で回答されるべきところが、橋梁コードで回答される事例が確認された。そこで、埋め込みモデルを paraphrase-multilingual-MiniLM-L12-v2 に変更し、同様の検証を行った。本モデルは 50 言語以上に対応する多言語埋め込みモデルであり、大規模なテキストデータを用いて事前学習されている。検証の結果、橋梁名を橋梁コードで回答する事例は確認されなかったものの、回答精度は BAAI/bge-m3 を用いた場合と比較して低かった。以上より、本研究では精度を優先し、埋め込みモデルには BAAI/bge-m3 を継続して使用することとした。

3.4.2 RAG データの違いによる回答の比較

RAG に追加するデータ量の違いによる回答への影響を検証するため、点検調査を 10 橋分および 100 橋分の場合で比較を行った。まず、回答時間については、初期稼働時に約 1 分を要したものの、それ以降は質問内容に依存し、数秒から数十秒程度であった。この傾向は 10 橋分の場合と 100 橋分の場合でほとんど差が見られなかった。回答内容については、「腐食の代表的な補修工法は？」という質問に対し、100 橋分を用いた場合には 10 橋分の場合と比較して回答文量が増加する傾向が確認されたが、一方で、回答内容にほとんど差が見られない場合もあり、回答の違いは質問内容に依存することが示唆された。これは、追加した点検調査の大半が健全度 I であり、RAG データ全体としての情報量が大きく変化しなかったためであると考えられる。

次に、道路橋定期点検要領を RAG データとして登録した場合について検証を行った。定期点検要領のみを登録した場合、資料内容を適切に抽出できる事例が多く確認された。回答は概ね資料中の文章に基づいて生成されており、回答内容の信頼性は高いと考えられる。これは、点検調書と比較して表形式の記載が少なく、文章主体で構成されているため、テキスト情報を適切に抽出できたためであると考えられる。

さらに、定期点検要領に加えて点検調書 10 橋分を登録し、同様の検証を行った。質問内容を「〇〇橋の点検時に注意する点は？」とした結果、点検調書のみを RAG データとした場合と比較して、回答文量は増加した。一方で、橋梁固有の詳細な情報はあまり含まれず、定期点検要領の内容に依存した一般的な回答となる傾向が確認された。

最後に、補修工法に関する内容を詳細に検討するため、健全度Ⅱの橋梁を対象とした点検調書 10 橋分を RAG データとして登録し、検証を行った。その結果、健全度Ⅰの点検調書が含まれている場合と比較して、回答に含まれる情報量が増加し、より適切な内容が得られた。これは、健全度Ⅱ以降の点検調書では、点検結果の補足欄に損傷に対する代表的な補修工法が多く記載されているため、各損傷に対応する補修工法の提示数が増加したことによるものと考えられる。

4. まとめ

本研究で作成したチャットボットは、補修工法の事例参照や、点検前における事前確認、さらには若手技術者の教育といった場面において有効に活用できる可能性があると考えられる。特に、FileMaker 上でローカル環境として構築されている点は、橋梁の点検調書などの機密性の高いデータを外部に送信することなく利用できるという点で、大きな利点で

ある。一方で、点検調書に多く含まれる表形式データの理解が十分でないことや、健全度判定のような構造的かつ定量的な情報の取り扱いに課題が残ることが明らかとなった。これらの結果を踏まえると、あくまで技術者の判断を補助する支援ツールとして位置付ける必要がある。

今後の改善点としては、点検調書の表構造を考慮した前処理を行い、項目ごとに情報を再構成する手法の検討が挙げられる。また、損傷写真などの画像情報を入力することができれば、文章情報だけでは捉えきれない損傷の特徴を反映した回答が可能になると考えられる。さらに、橋梁点検分野に特化した埋め込みモデルの比較・検証や、ローカル LLM に対するファインチューニングを行うことで、専門性の高い質問に対する回答精度の向上が期待される。これらの改善を段階的に進めることで、より実務に即した橋梁点検支援システムへと発展させることが可能である。これらの AI 技術を活用した維持管理業務が少しずつ確立されていくことで、今後の建設業界の維持管理業務の効率化および省力化に寄与すると考えられる。

文 献

- 1) 日本みち研究所 “全国道路施設点検データベース”
<https://road-structures-db.mlit.go.jp>
- 2) LM Studio – Local AI on your computer
<https://lmstudio.ai>
- 3) Claris FileMaker 2025 の提供を開始
<https://www.claris.com/ja/blog/2025/announcing-the-2025-release-of-claris-filemaker>
- 4) Claris FileMaker 2025 - RAG (Retrieval Augmented Generation : 検索拡張生成) 機能でカスタム App のデータを AI で参照
<https://ywc.com/filemaker/?p=55944>