河川維持管理（巡視・点検）における大規模視覚言語モデル（VLM）の適用性検討

旗手 聡[1]・相川 信幸[1]・坂本 遥峰[2]・菅田 大輔[2]・藤原 圭哉[1]

[1]日本工営株式会社 流域水管理事業本部 水工インフラマネジメント部（〒102-8539 東京都千代田区麹町5丁目4番地）

[2]日本工営株式会社 中央研究所先端研究センター（〒300-1259 茨城県つくば市稲荷原2304番地）

本稿では河川維持管理行為における河川点検および一次評価の効率化・標準化を目的としてVLMの適用可能性を検討した．現場で撮影される画像から変状や評価結果をVLMにより出力させ，結果の技術的説明性や例示された評価との正誤，回答出力時間について河川技術者の目線で評価した．また，VLMを含む自然言語モデルはプロンプト内容に大きく左右されるため，異なるプロンプト形式間の評価も実施した．

Key Words ： *Vision-Language Model,　河川維持管理　河川管理施設　河川点検　現場支援技術　Prompt engineering*

１．研究背景と目的

河川分野における維持管理業務は，河川構造物の健全性を把握し，補修・補強等の対策を実施することで，治水上の機能を確保するために行われる．

河川構造物及び河道の点検は，河川管理者や委託先の地元の調査会社等，河川の維持管理に精通していない受注者が実施することもある．河川区域を徒歩等で移動する点検者が目視により構造物等の変状を特定し，計測や触診・打音調査により実施される．

堤防等河川管理施設及び河道の点検・評価要領[1)]（以降，点検・評価要領）では，**図-1**のフローで河川構造物の点検および評価が実施される．点検のポイントや変状の評価方法が記載され，変状の評価結果の標準化に寄与している．一方で，実施者の違いによる点検結果・評価の揺らぎや変状の見落とし等の精度面に問題を抱え，AI技術を活用した河川構造物及び河道の点検の現場適用が検討されている．

近年の画像情報に基づいて判断するAI技術として，画像情報のみを扱う Convolutional neural network(CNN)ベースモデルと，画像とテキストの両方の情報を扱える Vision Language Model (VLM) が存在している．CNNベースモデルは，YOLOモデルに代表され，画像が属するクラスを分類するタスクに適するモデルである．一方で，VLMは，画像及びテキストを入力として，回答となるテキストを出力するモデルである．河川の点検では，点検者が目視した情報から必要な情報を読取り，記録することから，VLMの処理に近い．また，AI技術の性能を確保するためにCNNベースモデルは，検知事象に応じて教師データを作成する必要であるためにアノテーションの負担が大きい．一方で，VLMは画像と自然言語で構成される大規模なデータセットで事前学習されていることから，様々なタスクに対して追加学習をすることなく利用可能である．したがって，河川点検行為とVLMの親和性は高いと考えられる．

そこで本稿では，河川点検における一次評価に着目し，現場における巡視及び点検の効率化・点検結果の標準化を目的とした．本稿ではVLMを用いた河川点検画像の分類・文章生成の精度向上を目的として，プロンプトの違いが分類・文章生成の精度に及ぼす影響を検討した．

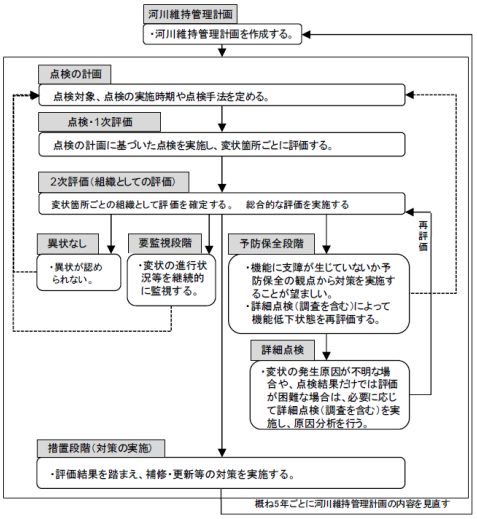

図-1 堤防等河川管理施設及び河道の点検・評価フロー[1)]

表-1 本稿で検証したプロンプト形式とその概要

プロンプト名称	プロンプト概要
Zero-shot	例を与えずただ命令する
Few-shot	例を提示し、タスクを実行するための方法を学習させる 例　次の例に倣って状況説明をしてください。(画像例およびその解釈の添付)
CoT	思考の過程例に倣って回答させる 例　この画像の状況説明をする際は、画像内の確認物体とその関係を示してから状況説明をしてください。
Zero-shot-CoT	思考の過程を記述させる 例　この画像の状況説明をしてください。最終的な説明に至るステップを記述し、答えてください。
Few-shot-CoT	いくつかの段階的な思考過程の例を与える 例　以降に示す例に倣って画像の状況説明をしてください。画像内の確認物体とその関係を示してから状況説明をしてください。(画像例およびその解釈の添付)
GCoT	1　写真の背景、キャプションを説明させる 2　質問の対象を特定させ、その位置情報を説明させる 3　質問に推論が必要な場合は段階的な推論ステップを展開する 4　質問の答えを示す

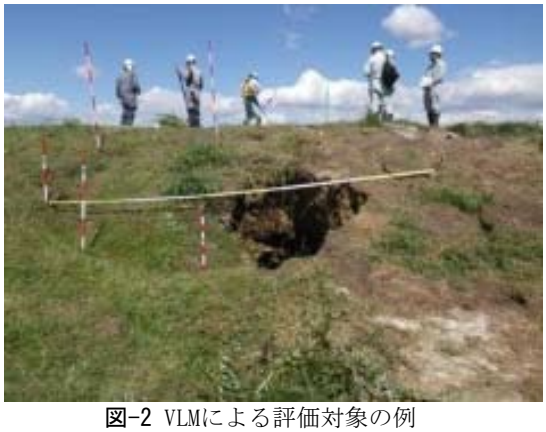

図-2 VLMによる評価対象の例
点検・評価要領「III. 事例写真集」から抜粋[1)]

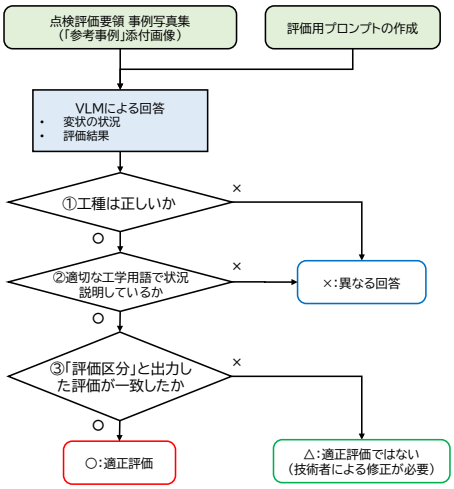

図-3 VLM出力結果に対する評価フロー

２．研究手法

(1) プロンプトの違いによる精度検証に向けたプロンプト形式の設定

VLMはプロンプト（命令文・質問文等）の違いによって出力内容が大きく変化するため，出力結果の精度向上が期待されるプロンプトエンジニアリングの手法を活用し，合計6種類のプロンプトを作成し，VLMの精度を検証した．

プロンプト形式の例は**表-1**に示す．VLMから出力させる内容としては大きく分けて「変状の状況」「評価区分」「評価理由」の3つを設定している．

(2) VLMへの入力画像データ

点検・評価要領「III. 事例写真集」（以下，「事例集」）の「参考事例」に掲載されている河川管理施設の変状画像87枚を入力画像データ(例：**図-2**)とした．VLM出力結果の評価を行う際には事例集内に記載されている「施設区分」「変状種別」「評価区分」「状況と評価区分」の内容とVLMの出力結果との比較・評価を実施した．

(3) 使用したVLMモデル

VLMは，画像と自然言語で構成される事前学習のデータセット内容によって，Visual Question Answering（VQA）性能が変わる．本稿のVLMモデルは，河川点検現場への適応を見据えて，VLMのVQAベンチマークスコアが比較的に高く，商用利用が可能なモデルとして，gpt-4o（OpenAI）を使用した．

VLMに入力する内容は，(1)で作成したプロンプトおよび(2)から出力した画像を合わせて入力した．

(4) VLM出力精度の評価

VLMを活用した技術者支援を行うには「技術的な説明性」が担保されている必要がある．本稿でのVLMの精度評価として，**図-3**の評価フローを用いた．フロー内①～③の評価内容を以下に示す．

① 工種が正しいか
VLMから出力された「変状の状況」内に記載された内容と事例集の「施設区分」が一致しているか確認した．

② 適切な工学用語で状況説明しているか
出力された「評価理由」を事例集の「状況と評価理由」の記載内容と比較して，工種各部の名称（天端，基礎工等）が適切に使用されているか，技術者が評価した．

③ 「評価区分」と出力した評価が一致したか
出力された「変状に対する評価結果」において，事例集「評価区分」（「a:異常なし」を除くb～dの3段階）の記載内容と一致するか，確認した．

①②を満たさない回答を「×：異なる回答」とし，①②を満たす回答を「説明性を満たす回答」とし，「説明性を満たす回答」については③の評価を行い，

◎：適正評価，×：過小評価，○：過大評価の区分を整理した．予測モデルの出力結果の評価では混同行列を用いた正解率や適合率，再現率等が用いられるが，本稿の出力結果では評価結果が3つ以上となるため，評価指標（**図-4**）を新たに定義する．

◎：適正評価は出力結果の評価と要領の評価が同じ状態を示している．×：過小評価は出力結果の評価が点検・要領の評価よりも軽微な変状として評価している．○：過大評価は，出力結果が要領の評価よりも重篤な変状として評価している状態にある．

◎：適正評価の数が多い件数となったプロンプトはVLMの出力結果が点検・評価要領に沿った適切な評価であることを示している．また◎：適正評価の数が少ないプロンプトであったとしても，○：過大評価の件数が多いプロンプトは変状を実際の評価よりも安全側に評価しているため，技術者のフォローが必要となるが良くない評価とは考えにくい．

ただし，×：過小評価に関しては実際の評価よりも低い評価をVLMが行っているために重篤な変状を見逃していると判断されるため，変状評価には適さないプロンプトと考えられる．点検・評価要領における評価として，c評価は予防保全段階として対策実施または詳細調査による機能低下状態の再評価，d評価では補修又は更新等の対策が必要と整理されている．そのため，これら2つの評価以下に見逃さずに維持管理を行うかがポイントの1つとなる．

		AIの評価		
		b	c	d
要領の評価	b	◎	○	○
	c	×	◎	○
	D	×	×	◎

◎:適正評価 →状況判断・評価ともに適切に実行
×:過小評価 →重篤な変状を見逃す可能性が有る
○:過大評価 →変状を安全側に捉えている

図-4 本稿における評価指標および混同行列

３．研究成果

(1) 出力結果の技術的説明性の評価

入力用に準備した画像およびプロンプトは87個あり，各プロンプトに対する回答結果を技術者による確認を行った．結果を**表-2**に示す．

説明性を満たす回答として最も良く回答したのは「CoT」であった．ただし，割合としては約40%で半数以上は技術的には説明不足の回答となっている．「×：異なる回答」に分類した例を**図-5**に掲載する．左図では樋門樋管を護岸と回答するような工種の取り違え，右図では色味が似通っている状態の鋼矢板をコンクリートと表現，構造物の構成材料を取り違えがあり，全体の構造が判別し難い近距離で撮影した画像や色味が判別し難い画像が多かった．

表-2 各プロンプトの技術的説明性の評価結果(件)

プロンプト形式	説明性を満たす回答（×:異なる回答 ではない）	×:異なる回答
Zero-shot	33 (37.9%)	54
Few-shot	25 (28.7%)	62
CoT	35 (40.2%)	52
Zero-shot-CoT	31 (35.6%)	56
Few-shot-CoT	19 (21.8%)	68
GCoT	16 (18.4%)	71

図-5 「×：異なる回答」に分類した画像例
（左図：Few-shot-CoT，右図：Zero-shot）

(2) 出力結果と点検・評価要領の評価区分との比較

前節で「説明性を満たす回答」と判断された回答結果の内，点検・評価要領 事例集 「評価区分」に記載されている評価と出力結果における評価区分との正誤を確認した．結果は**表-3**，×：過小評価となった回答例を**図-6**に示す．◎：適正評価が最も多く，×：過小評価が最も少ないプロンプトはCoTとなっている．「説明性を満たす回答」に対する割合に関しても，CoTが最も良い評価となっており，◎：適正評価が多く，×：過小評価も少ない結果となった．○：過大評価が最も多くなったプロンプトはZero-shotである．技術者のフォローが必要となるが，変状を安全側に評価可能となっている．

表-3 評価区分に関する正答の確認結果（件）

プロンプト形式	◎:適正評価	×:過小評価	○:過大評価	「説明性を満たす回答」
Zero-shot	15 (45.5%)	2 (6.1%)	16 (48.5%)	33
Few-shot	10 (40.0%)	11 (44.0%)	4 (16.0%)	25
CoT	19 (54.3%)	1 (2.9%)	15 (42.9%)	35
Zero-shot-CoT	13 (41.9%)	6 (19.4%)	12 (38.7%)	31
Few-shot-CoT	11 (57.9%)	3 (15.8%)	5 (26.3%)	19
GCoT	8 (50.0%)	2 (12.5%)	6 (37.5%)	16

図-6 変状を×：過小評価とした例
（Few-shot，点検・評価要領 d評価，VLM c評価）

(3) 結果出力に必要とする平均時間

現場実装を想定した際に，出力結果の精度と共に即時に回答が出力されるかどうかも重要視されると考えられる．そのため，現場レベルでモデルの有効性を判断するには出力結果の正確性と共に回答出力時間が重要となると考える．

作成したプロンプトは87個あり，各プロンプトに対して必要とした回答の平均時間を**表-4**に示す．最も素早く回答したのはZero-shotの4.5秒となり，最も遅く回答したのはCoTの10.3秒となった．

CoTは他プロンプトと比較として，文章の正確性は良いプロンプトであったが，回答時間に関しては最も回答時間が短いプロンプト（Zero-shot）よりも2倍強の回答時間を要した．評価結果の過程やプロンプト内に評価結果例の提示を実施している．評価結果の過程の説明およびプロンプトで提示した例の理解等に多くの処理が実行されている可能性が示唆される．

表-4　各プロンプト形式の回答出力時間

プロンプト形式	プロンプト1つに対する回答出力の平均時間(秒)
Zero-shot	4.5
Few-shot	5.2
CoT	10.3
Zero-shot-CoT	6.2
Few-shot-CoT	7.0
GCoT	9.8

(4) 総合評価

総合評価の方法としては，評点が高くなるほど評価を高く設定し各項目に対しての重みは設定せず，最高は6点である．(1)出力結果の技術的説明性の評価は「×：異なる回答ではない」に分類された結果の割合，(2)出力結果と点検・評価要領の評価区分との比較は「◎：適正評価」または「○：過大評価」に分類された結果の割合，(3)結果出力に必要とする平均時間は回答時間の昇順（短い順）で評価を実施した．

総合評価の結果は**表-5**，◎：適正評価となった画像例を**図-7**に示す．評価点が最も高くなったのはZero-shotとなった．

表-5 総合評価結果

プロンプト形式	説明性	評価正答率	出力時間	総合評価	総合評価順位
Zero-shot	5点	5点	6点	16点	1
Few-shot	3点	1点	5点	9点	4
CoT	6点	6点	1点	13点	2
Zero-shot-CoT	4点	2点	4点	10点	3
Few-shot-CoT	2点	3点	3点	8点	5
GCoT	1点	4点	2点	7点	6

図-7　◎：適正評価となった例
（Zero-shot，点検・評価要領 d評価，VLM d評価）

4．今後の課題と展望

(1) 本稿における課題

表-2の通り，「×：異なる回答」と分類されたプロンプトは6割～8割となっており，正確な技術用語を使用した出力結果となっていないのが現状である．ファインチューニングやRAG技術の併用により技術用語を理解可能にしたモデルの改良を目指す．

また，本稿ではプロンプト形式間の比較検討を行ったが，河川技術者が経験的に習得してきた河川構造物の確認ポイントを基にプロンプトテンプレートの作成や入力画像の幾何補正や鮮明化等を適用するすることで出力結果の画一化と精度改善を図る．

(2) モデルの適用範囲の拡大

近年，ドローン等を使用したインフラの点検が広がり，河川維持管理においてもUAVから撮影される画像・動画を利活用した取り組みが進められている．

従来の河川維持管理行為との差別化点としては，人間が実施していた河川区域内の巡回や変状の確認・写真撮影等の記録行為を全てUAVで行う点にある．UAVは河道上を飛行するため，河道や低水護岸等の状況確認が可能となり，現場作業の省力化や高度化が可能となる．ただし，UAVの記録データの確認には労力が必要となるため，解決策としてVLMによるスクリーニングが効果的であると考える．

(3) 業界におけるVLM適用のメリット

業界が抱える課題として，生産性向上および若手の技術者人材の育成が求められている．ベテラン技術者が経験的に培ってきた技術の継承が困難となり技術の空洞化を招くことが危惧されている．

VLMは状況把握と解釈行為の大部分に適用可能であると考えられ，受発注者関係無く経験の浅い技術者を支援可能となる．また，同一モデルを使用することで評価結果の画一性を高めることが可能になると共に，出力結果の正誤を適切に判断する経験を積むことで若年技術者の技能向上に資すると考える．

参考文献

1) 国土交通省 水管理・国土保全局 河川環境課：堤防等河川管理施設及び河道の点検・評価要領，2023，https://www.mlit.go.jp/river/shishin_guideaaline/kasen/tenkenhyouka/index.html