(48) Deep Learningによる橋梁撮影画像からの損傷状況説明文の自動生成

山根　達郎[1]・全　邦釘[2]・渡部　達也[3]

[1]学生会員　東京大学大学院新領域創成科学研究科　博士後期課程（〒277-8561 千葉県柏市柏の葉 5-1-5）
E-mail: yamane.tatsuro.20@dois.k.u-tokyo.ac.jp

[2]正会員　東京大学大学院特任准教授　工学系研究科総合研究機構（〒113-8656 東京都文京区本郷 7-3-1）
E-mail: chun@i-con.t.u-tokyo.ac.jp

[3]非会員　愛媛大学大学院理工学研究科　博士前期課程（〒790-8577 愛媛県松山市文京町 3）
E-mail: watanabe.tatsuya.15@cee.ehime-u.ac.jp

橋梁点検における撮影画像には，特定の損傷以外にも部材名称などの様々な情報が含まれている．本研究では，多様な損傷や部材が写っている画像を基に，損傷状況を説明した文章を生成できる Deep Learning モデルの構築を行った．さらに，Deep Learning モデルが文章を生成する際に，入力画像のどの部分に着目しているのかの可視化を行った．構築された Deep Learning モデルは高い精度で損傷状況を説明する文章を生成できることが示された．また，損傷状況を説明した文章の生成時に，損傷の発生箇所や各部材に着目して単語を出力していることが示された．

Key Words: *deep learning, image captioning, image analysis, maintenance, damage detection*

1. 序論

近年，橋梁などのインフラ構造物の高齢化が問題となっており，効率的な維持管理手法の確立が求められている．そこで，近年幅広い分野で高い性能を発揮している Deep Learning に代表される機械学習手法をインフラの維持管理に活用し，損傷の自動検出などを目的とした効率化を図る研究が国内外で進められている．Zhang らによる道路表面ひび割れの検出に関する研究[1]，Xue らによるトンネル表面のひび割れや漏水の検出に関する研究[2]，著者等による Deep learning や LightGBM, Random Forest などを用いてコンクリート表面のひび割れを画素単位で検出する研究[3)-5)]，構造物の振動から損傷検出を行う研究[6)]など，様々な研究が進めらている．

一方，情報工学の分野では，撮影画像を元に画像に写っている物体の状況を説明する文章を生成する Image Captioning に関する研究が，Deep Learning によって近年大きく進展している[7]．画像と文章を結びつけて処理することで，AI によって実行可能なタスクが大きく増えると考えられ，様々な分野への応用が期待されている．

維持管理分野においても，単純な画像分類 AI で扱える範囲には限界があり，Image Captioning の枠組みで扱うことで高度な判断が可能になると期待できる．例えば橋梁の撮影画像には，損傷の種類，損傷がどの部材に発生しているかなど，健全度の判断において非常に重要な情報が多数含まれている．また，部材情報や損傷種類などは相互に関係している．このような様々な情報は，単純な画像分類 AI では的確に拾い上げることは難しい．逆に，Image Captioning のような相互関連性を考慮できる手法であれば，「床版において，コンクリート補強材にうきが発生している」のように，部材情報と紐付いた損傷の状況など，関係性を含めた複雑な損傷評価を行えるようになる．

加えて，インフラの維持管理においては，維持管理方針の策定など様々な段階で人間による意思決定が行われる．そこで，Image Captioning のように，人間にとって理解しやすい自然言語を扱うことのできる AI の活用は，維持管理業務における人間の補助というユースケースを考えた場合にも有用である． よって本研究では，橋梁撮影画像を基に損傷状況の説明文を生成する Deep Learning モデルの構築を行った．

ただし，維持管理に関する意思決定の補助として AI を活用することを視野に入れる場合，Deep Learning による評価結果は説明性に乏しく解釈が困難であることから，

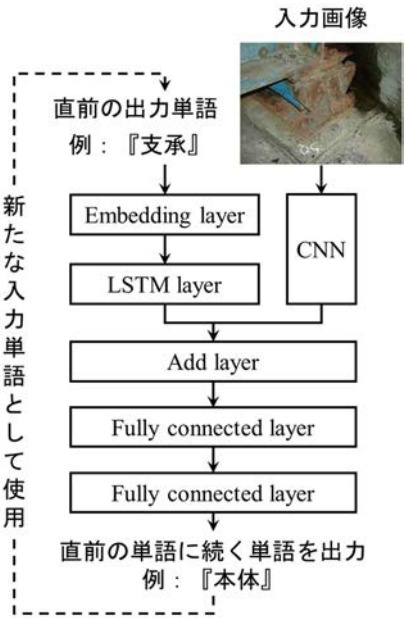

図-1 文章生成モデルの構造

判断の説明性が求められる現場では導入が進みづらいという課題が挙げられる．そこで本研究では，損傷状況の説明文を生成する Deep Learning モデルの構築のほかに，Deep Learning モデルが出力を行う際に，入力画像のどの部分を重要視して出力に至ったのか可視化する手法についても実装を行った．Deep Learning モデルの判断を可視化することで，説明性が求められる実際の現場への導入を円滑にするだけでなく，モデルの精度改善のための方針も立てやすくなると考えられる．

2. 文章生成

(1) 文章生成モデル

本研究では，文章生成モデルの構築にニューラルネットワークライブラリの一種である Keras を用いた．構築した文章生成モデルの概要を**図-1** に示す．本モデルは，主に画像を認識するための Convolutional Neural Network（CNN）と，時系列データである文章を取り扱うための Long short-term memory（LSTM）で構成されている．

本モデルでは，CNN に画像を入力するとともに，並行して LSTM を組み込んだネットワークにモデル自身が直前に出力した単語を入力する．これら 2 つのネットワークの出力は Add layer とよばれる 2 つの出力を受け取る層へと入力される．その後全結合層を介して最終的な出力が得られる．このようにして，モデルの出力として直前に出力された単語に続く単語が出力される．

モデルの CNN には，ILSVRC-2014 model with 16 weight layers（VGG16）[8]の学習済みモデルを使用した．また，本研究では，LSTM layer の前に Keras に用意されている

『支承本体において、腐食が発生している。』

図-2 学習データの一例

Embedding layer を設置している．Embedding layer は単語を分散表現とよばれるベクトルに変換するネットワークであり，変換された分散表現が LSTM layer に入力される．LSTM は過去の出力を行った際の状態を内部で保存するため，過去に出力された単語を考慮した上で続く単語を出力することが出来る．出力される単語は，学習データに存在する単語の中から確率が高い単語が選ばれる．

ここで，橋梁撮影画像のなかには，複数の部材や損傷が 1 枚の画像の中に混在しているものも多く存在する．そこで，本研究では，最も確率が高い単語のほかに，確率 の高い上位複数の単語を用いて文章を生成することで複数の文章の生成を行った．

(2) モデルの学習

Deep Learning によって Image Captioning を行う場合，モデルの学習データであるキャプションを大量に準備する必要がある．そのための学習用データセットとして，PASCAL[9] や MS-COCO (Microsoft Common Objects in Context)[10]などの学習用データセットが公開されているが，ただし，このようなデータセットを基に学習を行ったモデルを用いて，土木分野固有の画像を解析しても正しい結果は得られない．何故なら，上記のデータセットは特に土木工学分野固有の用語や概念が含まれたものになっていないためである．

そこで本研究では，実際の橋梁の損傷を撮影した画像を用いて学習データセットを構築し，そのうえで Image Captioning を行う Deep Learning モデルの学習を行う．本データセットには，愛媛県内の橋梁点検調書の作成に用いられた画像を用いた．画像の枚数は合計 10084 枚で，橋梁定期点検要領に記載されている部材区分および損傷の種類を参考に，すべての画像に部材名と損傷名を記述した文章を一文ずつ対応付けた．学習データの一例を**図-2** に示す．本研究では，構築したデータセットのうち 80% を training データとして用いて，残りの 20%を validation データとして用いて学習を行った．

(3) 文章生成結果

本研究では，学習用データセットに含まれない新たな

(a) (b) (c)

図-3 損傷状況の説明文の生成結果

表-1 各画像に対する生成された文章の一致率(%)

全文完全一致	1文以上完全一致	部材一致	損傷一致	全て不一致
33.0	82.0	86.0	91.0	5.0

画像200枚を用いて文章の生成を行った．出力した結果の一例を**図-3**に示す．**図-3**(a)は，主桁の防食機能が劣化し，部分的に腐食が発生している画像である．出力された文章では，主桁の腐食と防食機能の劣化について正しく文章が生成されていることがわかる．また，**図-3**(b)は，竪壁にひび割れをはじめとした複数の損傷が発生している画像であり，各損傷について説明した文章が正しく生成されていることが分かる．**図-3**(c)は，床版にうきや剥離が生じている画像であり，こちらも正しく文章が生成されていることが分かる．

(4) 生成された文章の評価

本節では，前節で文章生成を行った200枚の画像から得られた出力結果をもとに精度の評価を行う．文章生成や自動翻訳の精度評価においては，生成された文章と人間による文章との類似度を計算するBLEU[11]が多く用いられる．しかし，本研究のように画像内に様々な部材や損傷があるようなデータの場合，人間が準備する文章にも様々なパターンが考えられるため，類似度の評価は必ずしも生成された文章の正確な評価につながらない恐れがある．そこで，本研究では以下のように生成された文章と画像を比較して精度の評価を行った．

各画像について1文出力したときの評価基準は，「損傷とその損傷が発生している部材が画像と一致している（完全一致）」，「損傷が発生している部材が一致している（部材一致）」，「発生している損傷が一致している（損傷一致）」，「部材・損傷ともに一致していない（全て不一致）」の4項目に分けて評価を行った．複数の文章を出力した場合の評価基準は，「出力した全ての文章で，損傷とその損傷が発生している部材が一致している（全文完全一致）」，「損傷とその損傷が発生している部材が一致している文章が1つ以上ある（1文以上完全一致）」，「損傷が発生している部材が一致している文章が1つ以上ある（部材一致）」，「発生している損傷が一致している文章が1つ以上ある（損傷一致）」，「部材・損傷ともに全ての文章で一致していない（全て不一致）」の5項目に分けて評価を行った．評価は生成された文章と入力画像を著者が1枚ずつ確認することでどの項目に該当するかの分類を行った．

出力結果に対する評価を**表-1**に示す．**表-1**より，完全に一致した文章が1つ以上出力された割合は82.0%であることが分かる．この程度の精度が確保できていれば，例えばドローンなどを用いて橋梁全体を撮影することで損傷状態の大まかなスクレイピング等も可能になり，点検時により注視すべき箇所の目星などをつけることもできることから，人間が点検する際の補助の目的としては十分使えるレベルに達していると考えられる．さらなる精度の向上には，曖昧な写真や珍しい構図，形状の画像をさらに増やして学習を行う必要があると考えられる．

3. 着目領域の可視化

機械学習の手法としてDeep Learningを用いる場合，生成したモデルの判断はブラックボックスと見なされることから，判断の説明性が求められる現場では忌避されやすい．また，出力結果の一例を示した際に述べたように，判断が難しいと思われる写真については学習データを増やしたりするなどの対応が求められるが，闇雲に学習データを増やしても精度の向上は難しいと考えられる．

そこで，Deep Learningモデルが出力を行う際に，入力画像のどの部分を重要視して出力に至ったのかをヒートマップとして可視化する方法であるGradient-weighted Class Activation Mapping（Grad-CAM）[12]と呼ばれる手法を文章生成時に適用した．本研究では，単語を逐次的に出力することで文章の生成を行っているため，単語の出力ごとに着目領域のヒートマップを計算して出力すること

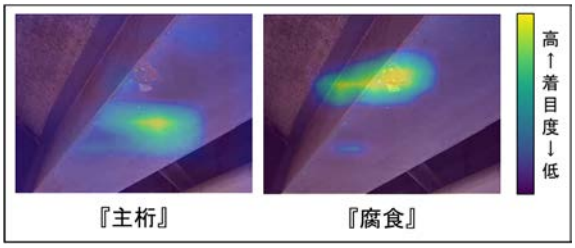

図-4 単語の出力時における着目領域の可視化結果

によって，モデルがどこに着目しながら各単語を出力しているのかを明らかにする．出力したヒートマップの例として，**図-3** (a)の文章生成時の部材名出力時および損傷名出力時の着目領域を**図-4** に示す．

図-4 から，『主桁』という単語を出力する際には，主桁を中心に高い反応を示していることがわかる．また，『腐食』という単語が出力されたときには，腐食が生じている部分に強い反応が生じていることが分かる．このように，Deep Learning モデルが部材や損傷の発生箇所に着目して単語を出力していることが確認できた．

4. 結論

本研究では，CNN と LSTM を組み合わせた Deep Learning モデルを用いることで，橋梁撮影画像から損傷状況を説明する文章を生成する方法を提案した．

構築した Deep Learning モデルを用いて 1 枚の画像から複数の文章を生成した結果，生成した全ての文章で部材および損傷種別が完全に正解しているものは 33.0%，部材および損傷種別が完全に正解している文章が 1 つ以上生成されているものは 82.0%に達した．

また，Grad-CAM を活用することによって，損傷状況説明文の生成時の着目箇所を可視化することで，Deep Learning モデルが部材や損傷の発生箇所に着目して単語を出力していることが示された．

本研究の成果である橋梁撮影画像からの損傷状況説明文の自動生成を発展させることによって，さらなる判断が可能になると考えられる．例えば，「主桁に腐食が発生している」および「床版にひび割れが発生している」といった情報を基に，「床版ひび割れからの漏水は，主桁に腐食を発生させる原因になる」といった既存の知識と連携させることで，損傷の原因を推定することができるようになる．損傷原因を推定することができれば対策方法の立案も当然視野に入るため，将来的にはドローンなどを活用して撮影した画像を基に，損傷状況や損傷原因を判断し，自動的に補修・補強の方針の立案ができるようになるなど，既存の維持管理体制を非常に省力化できるようになると考えられ，現在研究を進めている．

謝辞：本研究は「日本鉄鋼連盟鋼構造研究・教育助成事業」の助成を受けた研究の一部です．記して謝意を表します．

参考文献

1) Zhang, L., Yang, F., Zhang, Y. D. and Zhu, Y. J.: Road crack detection using deep convolutional neural network, *IEEE international conference on image processing*, pp.3708–3712, 2016.
2) Xue, Y. and Li, Y.: A Fast Detection Method via Region-Based Fully Convolutional Neural Networks for Shield Tunnel Lining Defects, *Computer-Aided Civil and Infrastructure Engineering*, Vol.33, Issue8, pp.638–654, 2018.
3) 全邦釘，嶋本ゆり，大窪和明，三輪知寛，大賀水田生，ディープラーニングおよびRandom Forestによるコンクリートのひび割れ自動検出手法，土木学会論文集 F3，73 巻，2 号，pp.I_297-I_307，2017.
4) 山根達郎，全邦釘：Deep learning による Semantic Segmentation を用いたコンクリート表面ひび割れの検出，構造工学論文集 65A，pp.130–138，2019.
5) Chun, P., Izumi, S. and Yamane, T.: Automatic detection method of cracks from concrete surface imagery using two-step Light Gradient Boosting Machine, *Computer-Aided Civil and Infrastructure Engineering*, in press, 2020.
6) Chun, P., Yamane, T., Izumi, S. and Kuramoto, N.:Development of a Machine Learning-based Damage Identification Method using Multi-point Simultaneous Acceleration, Measurement Results, *Sensors*, Vol.20, Issue.10, 2780, 2020.
7) Vinyals, O., Toshev, A., Bengio, S. and Erhan, D.: Show and Tell: A Neural Image Caption Generator, *The IEEE Conference on Computer Vision and Pattern Recognition*, pp.3156–3164, 2015.
8) Simonyan, K. and Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition, *International Conference on Learning Representations*, 2015.
9) Everingham, M., Gool, L. V., Williams, C. K. I., Winn, J. and Zisserman, A.: The PASCAL Visual Object Classes (VOC) Challenge, *International Journal of Computer Vision*, Vol.88, pp.303–338, 2010.
10) Lin, T., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P. and Zitnick, C. L.: Microsoft COCO: Common Objects in Context, *European Conference on Computer Vision*, pp.740–755, 2014.
11) Papineni, K., Roukos, S., Ward, T. and Zhu, W.: BLEU: a Method for Automatic Evaluation of Machine Translation, *Proc. of the 40th Annual Meeting on Association for Computational Linguistics*, pp.311–318, 2002.
12) Selvaraju, R. R., Cogswell, M., Das, A, Vedantam, R, Parikh, D. and Batra, D.: Grad-CAM: Visual Explanations From Deep Networks via Gradient-Based Localization, *The IEEE International Conference on Computer Vision*, pp.618–626, 2017.