

橋梁損傷検出における Transformer モデルを用いた画像処理の適用手法の研究

福岡 知隆¹・藤生 慎²

¹正会員 金沢大学特任助教 融合研究域融合科学系 (〒920-1192 金沢市角間町)
E-mail:tfukuoka@se.kanazawa-u.ac.jp (Corresponding Author)

²正会員 金沢大学准教授 融合研究域融合科学系 (〒920-1192 石川県金沢市角間町)
E-mail:fujii@se.kanazawa-u.ac.jp

日本では5年に一度橋梁の近接目視による定期点検が行われている。しかしこの点検は多大な労力を要し、技術者不足や予算の問題で今後の継続が困難であることが指摘されている。近年ではこの問題に対し、これまで人手で行ってきた作業を自動化する様々な取り組みが研究されている。

点検業務の省力化の手法の1つとして、画像処理技術を用いて橋梁画像から損傷を自動検出する手法が研究されている。近年はTransformerモデルを画像処理に適用した手法が提案されている。しかし、Transformerモデルを用いた手法において、損傷検出におけるモデル適用手法の議論は十分になされていない。本研究では橋梁の損傷検出における剥離と鉄筋露出を検出対象とし、入力画像の大きさを異なる画像データセットを用いて3つのモデルを訓練し、検出結果を比較し、評価をした。

Key Words: bridge maintenance, damage detection, image size, transformer model

1. はじめに

日本には橋長2m以上の道路橋が約73万橋あり、その多くが高度経済成長期に建設されている。そのため、今後高齢橋の割合は加速度的に高くなることを見込まれており、維持管理・更新を計画的かつ適切に進めていく必要がある¹⁾。現在の橋梁維持管理には、定期的な点検により損傷が軽微なうちに対策を講じる予防保全的な処置が求められている。2014年には、国が定める統一的な基準により、5年に一度の頻度で近接目視による全数監視が道路管理者に義務づけられた²⁾。しかし、人手による近接目視点検手法では作業コストなどの問題から将来にわたる継続的な維持管理は困難とされており、点検の効率化、低コスト化のため、様々な研究がなされている。

その1つが画像処理技術を用いた損傷の自動検出である。橋梁点検において点検者が確認する損傷は26種類あり³⁾、現在それらの損傷の多くは主に点検者の目視で確認されている。画像処理を用いることでそれらの損傷は検出が可能であると考えられる。深層学習を用いた検出手法も提案されており、ひび割れを対象とした手法⁴⁾⁵⁾、腐食、剥離・鉄筋露出、漏水などの自動検出手法が提案されている⁶⁾⁷⁾。深層学習による損傷検出では、大きく分けて2種類の検出手法が用いられている。検出対象を矩形で囲った結果を出力する物体検出手法と、検出対象

を画素単位で検出するセマンティックセグメンテーション手法である。近年では自然言語処理分野で主流となっているTransformerモデル⁸⁾を画像処理に適用した手法⁹⁾が提案されている。自然言語処理におけるTransformerモデルを用いた手法では、入力文における単語列を入力として、文全体における単語の位置情報から文脈を考慮した処理を行っている。一方で、画像処理に適用した場合は、入力画像を分割した画像の並びを単語に見立て、それらの分割画像の位置情報を考慮した処理を行っている。

一般的に深層学習を用いた手法がその性能を発揮するためには、モデルの訓練に膨大な訓練データが必要である。既存研究では、少数の訓練データに対して画像の回転や反転などの水増し処理を行うことで訓練データの総数を増やす手法がよく用いられている。しかし、モデルがその性能を発揮するためには訓練データの数量だけでなく、画像データの品質やクラスの割合も重要である。深層学習のモデルが想定する入力画像の大きさと実際に撮影された画像の大きさが異なる場合、画像の拡大、縮小、切り取りなどの前処理が必要である。しかし、それらの前処理は画像内の特徴が変化する、入力画像に含まれる領域が元の入力画像より狭くなるなど、深層学習モデルが検出処理を行う際に参照する情報量に変化があるため、検出精度への影響を考慮してそれらの前処理を実施しなければならない。しかし、Transformer モデルを用

いた橋梁損傷検出手法においては、それらの議論は十分になされていない。

本研究では Transformer モデルを用いた画像処理による、剥離、鉄筋露出を対象とした橋梁損傷検出手法において、モデルの訓練に用いた画像の大きさと検出時の入力画像の大きさの影響を評価する。検出結果はセマンティックセグメンテーション手法による画素単位での損傷検出画像とし、異なる大きさの画像データセットごとに検出モデルを訓練し、評価結果を比較する。

2. 関連研究

セマンティックセグメンテーションでは画像内の物体を認識し、あらかじめ訓練した検出対象である物体の領域を画素単位で認識する^{10), 11), 12)}。セマンティックセグメンテーションによる損傷検出結果は画素単位であるため、撮影距離やレンズの焦点距離などの撮影時の情報を組み合わせることで、位置だけでなく、幅や長さを取得することが可能であり、損傷の程度を判断するための情報も取得可能である。

Transformer モデルを用いた様々なセマンティックセグメンテーション手法の提案もされている^{13), 14), 15)}。これらの研究では、Transformer モデルを用いた手法が、これまでよく使われていた畳み込み学習を用いたモデルよりも高い精度での分類が行われていることを示している。また、損傷検出における対象を限定したモデルも提案されている。Liu らはひび割れを対象としたセマンティックセグメンテーションモデル CrackFormer¹⁶⁾を提案し、既存のモデルと比較評価を行い、CrackFormer モデルの検出性能が既存モデルより高いことを示した。しかし、これらの既存研究では、モデルの訓練時の画像データの大きさについて議論はされておらず、また検出時の画像の大きさの影響は評価されていない。

3. 評価手法

本研究では Transformer モデルを用いた剥離、鉄筋露出の検出モデルを画像の大きさが異なるデータセット毎に訓練し、それらのモデルによる検出結果を比較し、画像の大きさの影響を評価する。また、既存手法との比較として、一部のデータセットでの畳み込み学習を用いたモデルによる検出結果との比較も行う。

(1) 深層学習モデル

本研究では Transformer モデルを用いたセマンティックセグメンテーションモデルとして、様々な対象のセグメ

ンテーションにおいて、高い分類性能を示した SegFormer モデル¹⁷⁾を用いる。また、比較モデルとして、畳み込み学習を用いた SegNet¹¹⁾モデルを用いる。SegFormer モデルではエンコーダ部のサイズが大きいほど分類性能が高いことが示されており、本研究では最も分類性能が高い SegFormer B-5 モデルを用いる。モデルの想定入力画像サイズは縦 512 ピクセル、横 512 ピクセルとする。SegNet の入力サイズは縦 224 ピクセル、横 224 ピクセルとする。また、SegFormer モデルは 1000 クラスに分類した大規模画像データセット ImageNet を用いて事前に訓練されたモデルを用いて剥離、鉄筋露出の検出モデルを訓練する。SegNet モデルにおいても同様に ImageNet を用いて訓練された VGG16 モデルを訓練済みモデルとして、剥離、鉄筋露出の検出モデルとして訓練する。

(2) データセット

本研究では剥離・鉄筋露出画像として K 県の橋梁点検調査に記載されている画像 179 枚を用いる。この画像データセットは一人のアノテータによって、剥離箇所を黒色、鉄筋露出箇所を赤色でアノテーションしている。画像の大きさ、縦横比は統一されておらず、最大長辺が 950 ピクセル、最小短辺が 270 ピクセルの画像データセットである。ランダムに抽出した 119 枚を訓練用画像、60 枚を評価用画像として分割する。訓練データには水増し処理を行い、画像データ数を増やす。水増し処理として、以下の処理をランダムな順番で実施する。また各処理は 66% の確率で実施されないものとした。

1. 水平反転

画像を水平軸に反転。

2. 垂直反転

画像を垂直軸に反転。

3. 移動

画像を水平方向に 0~20 ピクセル、垂直方向に 0~20 ピクセルの範囲でランダムに移動。

4. 拡大・縮小

画像を水平方向に 0.8~1.2 倍、垂直方向に 0.8~1.2 倍の範囲でランダムに拡大・縮小。

5. 回転

画像を 0~360° の範囲内でランダムに回転。

画像の大きさによる影響を評価するため、3 つの訓練用データセットを作成する。元の画像データセットに何も処理をしないデータセット A、元の画像データセットの各画像を幅 224 ピクセルのメッシュに分割したデータセット B、元の画像データセットの各画像を幅 448 ピクセルのメッシュに分割したデータセット C である。データセット A のデータがモデルの入力時に大きさと縦横比

表-1 分割データセットの画像数

| | データセット B | データセット C |
|------------|----------|----------|
| 分割後画像数 | 1718 | 478 |
| アノテーション画像数 | 369 | 464 |

が調整されるのに対して，データセット B, C は大きさだけが調整される．訓練データセットの画像の剥離，鉄筋露出領域は画像内の一部であるため，分割した場合はそれらの領域が含まれない画像が生成される場合があるが，本研究ではそれらの画像も訓練データとして利用する（表-1）．各データセットに対する水増し処理は分割後に実施し，各データセットの画像数を 17810 枚に水増した．

(3) 検出手法

各モデルは入力画像に対して，訓練データと同じ大きさに変換する前処理を実施する．データセット A から訓練した検出モデル A による検出時は前処理を行わない．データセット B, データセット C から訓練した検出モデル B, C による検出時は入力画像をそれぞれの訓練データの大きさ 224 ピクセル，448 ピクセルのメッシュに分割する．分割後の画像を入力データとして受け取り，それぞれの画像に対して検出結果を出力し，結合しなおすことで元の入力画像の検出結果を生成する（図-1）．

モデル B, C では画像を分割することにより，入力画像の範囲が元画像よりも狭くなる．本研究ではこれを補うため，分割画像の始点をずらした複数の分割画像セットの結果を重ね合わせる手法を用いる．分割の始点をずらさない場合，分割画像のピクセルサイズの半分だけ x 軸方向に分割始点をずらした場合，分割画像のピクセルサイズの半分だけ y 軸方向に分割始点をずらした場合，分割画像のピクセルサイズの半分だけ x 軸，y 軸方向に分割始点をずらした場合の 4 つの検出結果を重ね合わせた結果を最終的な検出結果とする（図-2, 図-3）．

(3) 評価指標

本研究では評価尺度として精度(Precision)，再現率(Recall)を用いる．それぞれ式(1)，(2)に基づき，ピクセル単位の検出の成否に基づいて算出する． N_{much} はモデルにより検出対象損傷として検出したピクセルのうち，正しく検出対象損傷を検出できていたピクセル数， N_{detect} はモデルが検出対象損傷として検出したピクセル数を示す． N_{damage} は検出対象損傷のピクセル数である．

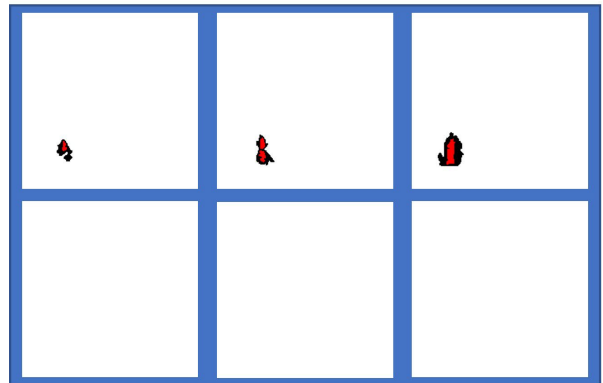
本稿における精度は検出結果で検出対象損傷と判断したピクセルのうち，どの程度正しく検出対象損傷ピクセルだけを検出対象損傷として検出しているかを評価する．



a: 入力画像



b: 分割後の入力画像



c: 分割画像の検出結果



d: 最終出力結果

図-1 画像分割モデルによる入出力例

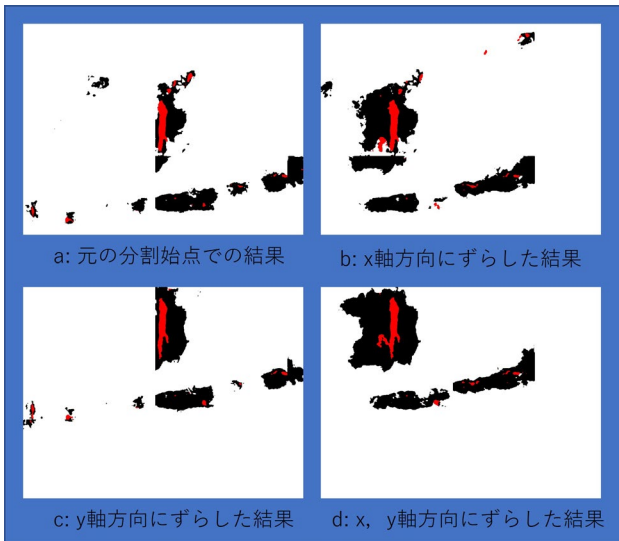


図-2 分割始点ごとの検出結果画像



図-3 分割始点ごとの検出結果を重ね合わせた画像

$$Precision = \frac{N_{much}}{N_{detect}} \quad (1)$$

$$Recall = \frac{N_{much}}{N_{damage}} \quad (2)$$

すなわち、精度が高いほどモデルが検出した結果は正しく検出対象損傷のみを検出できているといえる。再現率は正解である検出対象損傷ピクセルのうち、どの程度モデルが検出できたかを示す。再現率が高いほどモデルの結果は対象損傷検出の取りこぼしが少ないといえる。

4. 評価結果

SegFormer モデルによる検出結果はデータセット A, B, C の 3 つのデータセットで訓練したモデルの結果とする。比較モデルである SegNet モデルはデータセット B のみを

表-2 剥離検出結果の比較

| | SF_A | SF_B | SF_C | SN_B |
|-----|--------|--------|--------|--------|
| 精度 | 0.808 | 0.771 | 0.769 | 0.793 |
| 再現率 | 0.741 | 0.658 | 0.674 | 0.441 |

表-3 鉄筋露出検出結果の比較

| | SF_A | SF_B | SF_C | SN_B |
|-----|--------|--------|--------|--------|
| 精度 | 0.708 | 0.747 | 0.765 | 0.693 |
| 再現率 | 0.685 | 0.486 | 0.563 | 0.420 |

評価する。表-2 に剥離の検出結果の精度、再現率、表-3 に鉄筋露出の検出結果の精度、再現率を示す。

SegFormer モデルを用いた検出モデル A を SF_A 、検出モデル B を SF_B 、検出モデル C を SF_C とする。SegNet モデルを用いた検出モデル B を SN_B とする。表-2 の結果から、 SF_A が最も精度、再現率が高いことが示された。その他の SegFormer モデルについても、SegNet モデルと比較して若干精度は低いだが、再現率が大きく向上しており、Transformer モデルは剥離検出に有効なモデルであると考えられる。 SF_B と SF_C の結果を比較すると、分割サイズを大きくした方が若干検出性能が向上している。また、 SF_A と SF_B の結果を比較した場合、 SF_A の方が検出性能が高い。SegFormer モデルの結果を比較した場合、剥離検出の場合、入力画像の範囲が広いほど検出性能が向上しており、縦横比の調整による影響は小さいと考えられる。表-3 の結果から、SegFormer モデルは SegNet モデルと比較して検出性能が向上しており、Transformer モデルは鉄筋露出の検出に有効なモデルであると考えられる。一方で、 SF_A 、 SF_B 、 SF_C の再現率が剥離と同様に入力画像の範囲が広いほど向上しているが、精度が最も高いのは SF_C となっている。これは鉄筋露出の検出においては入力画像の範囲が広いだけでなく、縦横比が調整されず、鉄筋の形状情報が変化しない方が望ましいためと考えられる。

これらの結果から剥離、鉄筋露出検出において、Transformer モデルは入力画像の大きさが大きいほど検出性能が向上する傾向にあると考えられる。一方で、鉄筋露出では精度が低下していることに鑑みるに、検出対象によっては、過剰な検出を抑えるためにデータセットの縦横比の変化は避けた方が望ましいものと考えられる。また、入力画像の大きさを検出モデルの想定入力サイズより小さくしたメッシュ画像とした場合、精度に -5.7%~3.9% の増減があり、再現率に 19.9% の低下が起こることが示された。

5. おわりに

近年の橋梁点検では近接目視点検手法の代替手法が求

められており、画像処理を用いた損傷検出手法は多く研究されている。本研究は最新の深層学習ベースの検出モデルであるTransformerモデルを用いた損傷検出手法において、画像の大きさに着目した適応手法について評価を行った。モデルの訓練、評価時の入力画像の大きさを変化させた結果を比較し、検出性能の増減を確認した。

本研究の評価においては、分割画像のサイズが検出モデルの想定入力画像サイズより大きいデータセットを用いた評価実験を行い、適切な画像の分割サイズの評価する。

参考文献

- 1) 国土交通省白書 2021 , <https://www.mlit.go.jp/hakusyo/mlit/r02/hakusho/r03/pdf/kokudo.pdf> [2022年9月30日]
- 2) 国土交通省道路局, 道路橋定期点検要領, https://www.mlit.go.jp/road/si-saku/yobohozen/tenken/yobo4_1.pdf [2022年9月30日]
- 3) 全邦釘, 井後 敦史: Random Forest によるコンクリート表面ひび割れの検出, 土木学会論文集 F3, Vol.71, No. 2, I_1-I_8, 2015.
- 4) Yokoyama, S., Matsumoto, T.: Development of an automatic detector of cracks in concrete using machine learning. The 3rd International Conference on Sustainable Civil Engineering Structures and Construction Materials - Sustainable Structures for Future Generations, 171, pp.1250-1255, 2016.
- 5) Cha, Y.-J., Choi, W., Büyüköztürk, O.: Deep learning-based crack damage detection using convolutional neural networks. Computer-Aided Civil and Infrastructure Engineering, 32(5), pp.361-378, 2017.
- 6) Dong H., Gang T.: Damage detection of quayside crane structure based on improved faster R-CNN. International Journal of New Developments in Engineering and Society, 3(2), pp.284-301, 2019.
- 7) Zhang, C., Chang, C., Jamshidi, M.: Concrete bridge surface damage detection using a single-stage detector. Computer-Aided Civil and Infrastructure Engineering, 35(4), pp.389-409, 2020.
- 8) Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez N. A., Kaiser Ł. Polosukhin I.: Attention is All You Need. Proceedings of the 31st International Conference on Neural Information Processing Systems, pp.6000-6010, 2017.
- 9) Dosovitskiy, A., Beyer L., Kolesnikov A., Weissenborn D., Zhai X., Unterthiner T., Dehghani M., Minderer M., Heigold G., Gelly S., Uszkoreit J., Houshy N.: An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale, Proceedings of International Conference on Learning Representations, 2021.
- 10) J. Long, E. Shelhamer, T. Darrell: Fully convolutional networks for semantic segmentation. CVPR, 2015.
- 11) V. Badrinarayanan, A. Kendall, R. Cipolla, SegNet: Deep Convolutional Encoder-Decoder Architecture for Image Segmentation, ArXiv:1511.00561v3, 2016.
- 12) Li, S., Zhao, X., Zhou, G.: Automatic pixel-level multiple damage detection of concrete structure using fully convolutional network, Computer-Aided Civil and Infrastructure Engineering, 34(7), pp.616-634, 2019
- 13) Y. Jin, D. Han, H. Ko.: TrSeg: Transformer for semantic segmentation, Pattern Recognition Letters, Vol. 148, pp.29-35, 2021.
- 14) Xu Z., Zhang W., Zhang T., Yang Z., Li J.: Efficient Transformer for Remote Sensing Image Segmentation, Remote Sensing, 13(18), 2021.
- 15) Li Z., Xu P., Xing J., Yang C.: SDFormer: A Novel Transformer Neural Network for Structural Damage Identification by Segmenting the Strain Field Map, Sensors, 22(6), 2022.
- 16) Liu H., Miao X., Mertz C., Xu C. Kong H.: CrackFormer: Transformer Network for Fine-Grained Crack Detection, IEEE/CVF International Conference on Computer Vision (ICCV), pp.3763-3772, 2021.
- 17) Xie E., Wang W., Yu Z., Anandkumar A., Alvarez M J., Luo P.: SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers, arXiv preprint arXiv:2105.15203, 2021.

(2022.9.30 受付)

A BASIC STUDY OF APPLYING THE TRANSFORMER MODEL-BASED BRIDGE DAMAGE DETECTION METHOD

Tomotaka FUKUOKA, Makoto FUJIU

In Japan, periodic inspections of bridges by the close visual inspection method are conducted once every five years. This bridge inspection needs much cost. It causes a lack of engineers and budget. Some local governments couldn't complete the bridge's aggressive preventive maintenance in Japan. To solve those problems, studies of automation have been made to reduce the inspection task which depends on human power.

Recently, deep learning-based damage detection methods have been studied by many researchers to reduce the cost of the bridge's aggressive preventive maintenance. We focus on the state-of-the-art technology for semantic segmentation method which uses the Transformer model. This kind of method has high accuracy to detect the target from an input image. On the other hand, there is not enough discussion about how to use this new method. In this study, we set the task of detecting the peeling and the rebar exposure on the surface image of a bridge. We evaluate the effectiveness of the size of the image for training the model by comparing three models trained with different image sizes.