(41) 深層学習を用いた鋼骨組構造における出来形検出に関する基礎的検討

井筒　竜宇[1]・矢吹　信喜[2]・福田　知弘[3]

[1]学生会員　大阪大学大学院工学研究科　環境・エネルギー工学専攻　博士前期課程
E-mail: izutsu@it.see.eng.osaka-u.ac.jp

[2]フェロー会員　大阪大学教授　大学院工学研究科　環境・エネルギー工学専攻
（〒565-0871 大阪府吹田市山田丘 2-1）
E-mail: yabuki@see.eng.osaka-u.ac.jp

[3]正会員　大阪大学准教授　大学院工学研究科　環境・エネルギー工学専攻
E-mail: fukuda@see.eng.osaka-u.ac.jp

建設現場における出来高管理は，通常，写真を撮影し，図面詳細図や BIM/CIM（Building/Construction Information Modeling）モデルと照らし合わせて行う．この作業は時間がかかり人的ミスも発生しやすいことから，より正確で効率的に進捗状況を把握する手法が求められている．そこで本研究では，深層学習を活用し，施工途中の鋼骨組構造における梁や柱などの各構造部材をカメラで撮影した画像から検出することで，効率的に施工現場の進捗を把握することができるシステムの構築を目的とする．具体的には，既存の Object Detection Convolutional Neural Network (CNN)と Segmentation CNN をファインチューニングすることで，撮影した画像から施工途中の構造物を検出可能な CNN を構築する．その後，構築した二つの CNN モデルを統合することによって，画像から各構造部材を把握する事が出来るシステムを構築する．開発したシステムの精度検証と考察を実施する．

Key Words: *convolutional neural network, deep learning, as-built detection, steel frame, segmentation*

1. はじめに

近年，建設現場における出来高管理作業は，現場で撮影した写真や図面詳細図に加えて BIM モデルを用いることで，視覚的に進捗状況を把握する事が出来る．しかしながら，この作業は時間がかかるだけでなく，人的ミスが発生しやすい作業である．そのため，人の作業により進捗状況の把握を行うのではなく，システム上で進捗状況をより正確で効率的に把握する事が求められている．

一方で，深層学習を活用した物体検出の技術は日々発展してきている．デジタル画像のデータセットである Caltech 101 に始まり，画像データベースである ImageNet の登場により，深層学習を活用した物体の認識精度は大きく発展した．2010 年から開催されている大規模物体認識コンペティションである ILSVRC（ImageNet Large Scale Visual Recognition Challenge）において，2015 年以降から深層学習によって人を上回る検出精度を達成しており，研究者達は深層学習を活用した物体検出だけに留まらず，検出結果を活用して発展的なシステムを構築する傾向にある．

以上のような背景のもと本研究では，既存の Object Detection CNN や Segmentation CNN に対してファインチューニングを行うことで，実際の施工における鋼骨組構造物の柱や梁などの構造部材ごとに検出可能な CNN を構築し，二つの CNN モデルを統合することによって鋼骨組構造に対して物体検出とセグメンテーションを行うことが出来るシステムを構築する．はじめに，実際の施工現場は足場や防音シートなどで覆われている事が多いため鋼骨組構造体の模型を作成し，模型を撮影することにより学習用画像データセットを準備する．次に，深層学習を活用することによって構造部材を検出する事が出来るように，YOLOv3[1]（You Only Look Once）や U-Net[2]のような既存の深層学習モデルの学習の重みを作成した模型画像を用いてファインチューニングする事で変更する．その後，構築した二つの CNN モデルを統合することで，カメラを用いて撮影した画像から指定した対象構造物を検出可能なシステムの構築を行う．実際の施工現場に対する検出精度から開発したシステムの検証と考察を行う．

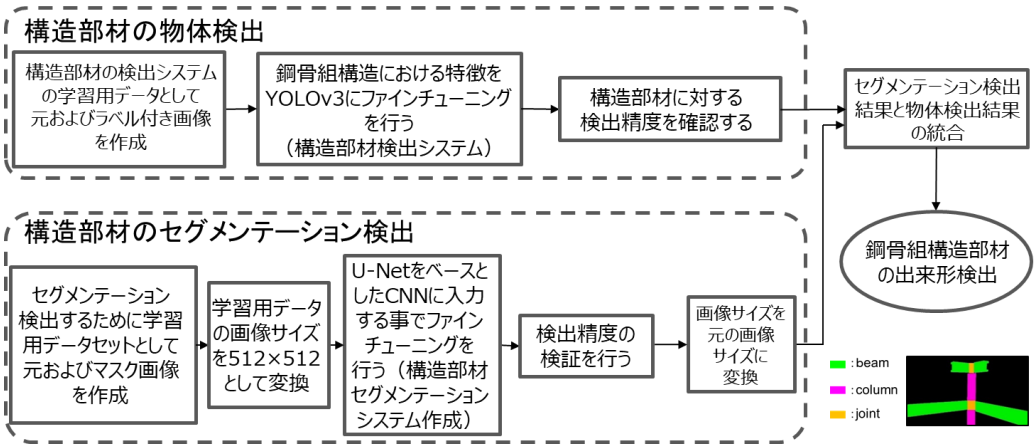


図-1　提案するシステムの開発手順

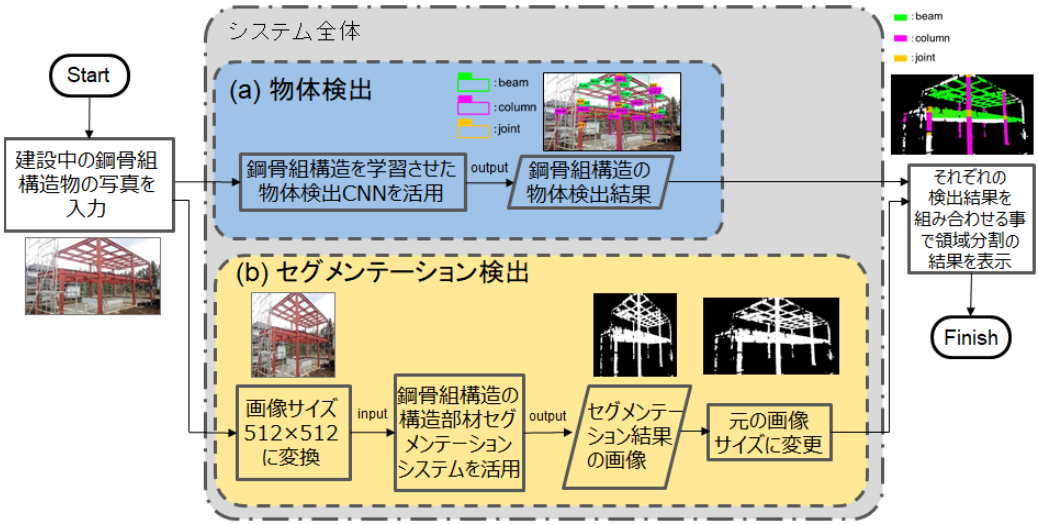


図-2　実際の鋼骨組構造に対する検出手順

2.　提案手法

本研究で提案するシステムの開発手順を図-1 に示す．提案するシステムは作成した鋼骨組模型を用いて学習用データセットを作成し，CNN にファインチューニングを行うことで鋼骨組構造の検出を行うことが出来るシステムである．図-2 に実際の鋼骨組構造物に対して検出を行う手順を記載する．図-2 (a)の部分はカメラで撮影した画像から柱や梁などの各構造部材を確認するシステムの手法を示している．実際の鋼骨組構造物の画像を入力し，それぞれの構造部材ごとに検出を行う．その後，深層学習モデル活用したセグメンテーション結果と検出結果を重ね合わせて二つの結果を結合させることによって領域分割を行う．図-2 (b)の部分は出来形検出システムの手法を示している．入力した実際の施工現場写真の画像サイズを 512×512 pixel に変形させた後，学習させた CNN を用いることで入力した画像から施工完了部分を検出し，マスク画像を作成する．作成したマスク画像を元の画像サイズに戻すことで，領域分割に利用する結果の一つとして使用する．なお，鋼骨組構造は，各種水路，水槽のカバー，橋脚，土留め，仮設備など土木分野で幅広く利用されている．

(1)　検出対象構造物に対する物体検出

開発したシステムは構造物を部材ごとに検出する事を目的としている．しかし，ImageNet のような既存の画像データベースを用いるだけでは，データベースに含まれていない建設構造物の検出を行うことは不可能である．そこで，実際の工事現場から学習用データセットに使用するための画像データを作成することを試みたが，一般的な工事現場では足場や仮囲いなどが存在するため，本研究の検出対象である柱や梁などを直接撮影することは困難である．解決策として，縮尺 1/30 の鋼骨組の模型を作成することによって模型の画像データを撮影し，学習用データセットとして用いる．このデータセットで再学習を行わせることによって，実際の施工現場における

鋼骨組の柱・梁・接合部などの構造部材を検出することが可能か検証を行った．学習用画像データセットとして2000枚の鋼骨組模型画像を用意し，その内400枚をテスト用画像として無作為に選択して使用している．

はじめに，作成した鋼骨組模型の写真データを学習用データセットとして使用するために，LabelImgを用いてアノテーション作業を行った．試験的に検出を行うクラス数は3として設定し，柱（column）・梁（beam）・接合部（joint）の三種類を建設構造物から検出する対象として撮影画像からアノテーションデータを作成している．また，XMLファイルによって画像内におけるそれぞれの部材を示す矩形の位置が示されている．

(2) 鋼骨組建方模型を用いたセグメンテーション検出

検出対象としている建設構造部材のセグメンテーションによる検出を行うために学習用データセットを作成した．矩形検出を行う際とは違い，検出対象物である柱・梁・接合部の三種類を白色で表示し，その他の背景を黒色で表示するようにマスク画像を作成することで学習用データセットとして使用している．それぞれの学習用データセットは入力する際に画像サイズが大きすぎると学習がうまくいかないため，画像サイズを512×512に統一している．この大きさで学習を行うことによって，サーバ内で画像に表示されている検出対象物の特徴量を把握することが容易になっている．

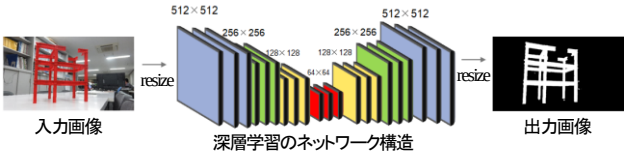
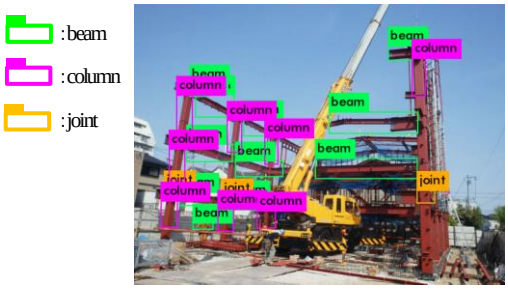


図-3 U-Netをベースとした CNNの構造と学習内容の一例

(a) 入力した施工現場画像



(b) 入力画像に対する検出結果

図-4 構築した物体検出システムの検出結果

3. システム開発

物体検出用に作成した学習用データセットを適切な形式にすることでYOLO v3の再学習を行い，鋼骨組模型に対して物体検出を行った．YOLO v3は優れた物体認識CNNであり画像内における位置情報を扱うタスクを与えることによって，様々な物体のクラスと位置を扱うことが可能になっている．その後，セグメンテーション用に作成した学習用データセットを用いて，U-Netの学習の重みを変更するためにファインチューニングを行った．U-Netをベースとして，検出対象物である三種類の部材に対してセグメンテーションで検出を行う事ができるCNNを構築した（図-3）．U-Netはセマンティックセグメンテーションのモデルの一つであり，元々は医療画像のセグメンテーションのために提案されたモデルである．作成した学習用データセットの前処理を行う事で検出対象物同士の境界を正確に表示するために，検出対象物と背景の境界部分のロスを多くするなど検出するための工夫がされている．一方で，同様にセマンティックセグメンテーションのモデルの一つである SegNet[3]は，風景画像のセグメンテーションを高速，省メモリで行うといった深層学習モデルで，Encoder-Decoder構造を採用しているため処理を高速で行うことが出来る．しかしながら，今回のシステムでは検出対象物を撮影した画像を用いて，施工完了部分だけを精度よく検出する必要がある．そのため，高速で処理を行うリアルタイム性は重要視していないことからU-Netを採用した．

4. 検証実験

再学習させることで鋼骨組模型の柱、梁および接合部を検出可能になったYOLOv3に実際の建設施工現場の写真を入力することで構造部材を検出することが可能か検証を行った．検出精度の検証実験として，CNNの学習用データセットに形状が類似していない画像を用意して検出を行った（図-4）．図-4 (a)は用意した画像は重機のような遮蔽物が存在しており，学習用データセットに含まれていないパターンの画像となっている．

結果として，実際の建設現場における画像から検出された結果は精度が高いものの対象全てを確認することが

出来ていない状況にある（図-4 (b)）．この原因として，写真の左部分は検出対象である構造部材が重ならずにそれぞれを確認できる状態で撮影されているため検出精度が高くなっているが，写真の右部分に存在する検出対象物の柱部分が複数本重なることによって境界線が曖昧になっていることが確認できる．検出対象物の構造が複雑に写ってしまうことで発生した問題であると考えられる．この問題については，今後実際の施工現場の画像データを学習用データセットに加え，構築したシステムに多様性を持たせることによって改善することが可能であると考えている．

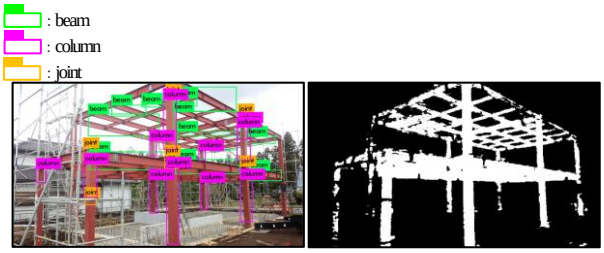
同様にファインチューニングによって鋼骨組模型の画像データを学習させた U-Net に対して実際の建設施工現場の写真を入力することで構造部材を検出することが可能か検証を行った（図-5 (a), (b)）．IOU（Intersection over Union ）によってシステムの精度検証を行った．ここで，IoU とはセグメンテーション結果により，検出対象物であると予想した領域の内，モデルが正しく検出対象物である事を認識できた領域がどの程度かを示したものになっている．実際の鋼骨組に対しての精度は約 8 割程度であり一定以上の精度が得られているものの，それぞれの部材ごとの精度を確認すると，柱および梁と比較して接合部の検出精度が悪いことが確認できる．この結果は，接合部の構造が他二つの検出対象と比較して複雑である事と足場など構造が複雑になっている部分を接合部として誤検出してしまっている事が原因と考えられる.セグメンテーション結果と物体検出結果を組み合わせることで領域分割を行った結果を示す（図-5 (c)）．鋼骨組模型と実際の施工現場に対する構築した深層学習モデルの検出精度を表-1 に記載する．

今回の学習用データセットは鋼骨組模型を用いて行ったため，鋼骨組模型に対する検出精度が高くなることは予想が出来たが，実際の施工現場においても一定以上の検出精度を確保する事が出来ている．この検出精度は建設模型だけの学習用データセットに対して実際の施工現場の画像データを数割ほど加える事によって，多様性を持たせることで向上させることが出来ると考えられる．

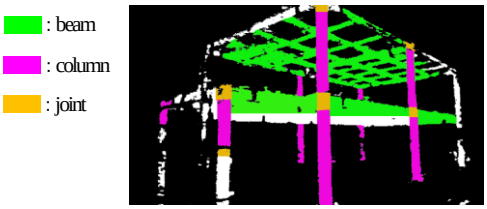
表-1　領域分割によって得られた精度

カテゴリ	建設部材	Intersection over Union（%）
鋼骨組模型	Column	95.7
	Beam	95.5
	Joint	84.1
実際の建設施工現場	Column	80.3
	Beam	78.4
	Joint	50.4

（a）入力画像



（b）入力画像から得られた物体検出および
セグメンテーション結果



（c）得られた検出結果から領域分割を行った結果

図-5　実際の鋼構造体に対して検出を行った結果

5. まとめ

本研究では，鋼骨組構造の部材である柱・梁・接合部を対象として検出を行った.今後の展開として，様々な種類の検出対象物の画像データを加える事によって，構造部材に対する学習用データセットを作成し，精度の向上および検出対象物の種類の増加を行う．また，セグメンテーションを行う CNN の精度を向上させるために特徴量の把握がより精確に行えるシステム構築を行う．

参考文献

1) Redmon, J., Farhadi, A.: Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767. 2018.
2) Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*. Springer, Cham. pp. 234-241, 2015.
3) Badrinarayanan, V., Kendall, A., Cipolla, R.: Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE* transactions on pattern analysis and machine intelligence, 39(12), pp.2481-2495, 2017.