

# 複数種類の高解像度衛星画像を用いた Mask R-CNN による建物抽出・被害分類モデル

内藤 昌平<sup>1</sup>・土屋 美恵<sup>2</sup>・友澤 弘充<sup>3</sup>・田口 仁<sup>4</sup>

<sup>1</sup>正会員 防災科学技術研究所 マルチハザードリスク評価研究部門 契約研究員  
(〒305-0006 つくば市天王台 3-1)

E-mail: naito@bosai.go.jp (Corresponding Author)

<sup>2</sup>非会員 みずほリサーチ&テクノロジーズ株式会社 (〒101-8443 東京都千代田区神田錦町 2-3)

E-mail: misato.tsuchiya@mizuho-rt.co.jp

<sup>3</sup>非会員 みずほリサーチ&テクノロジーズ株式会社 (同上)

E-mail: hiromitsu.tomozawa@mizuho-rt.co.jp

<sup>4</sup>非会員 防災科学技術研究所 防災情報研究部門 副部門長 (〒305-0006 つくば市天王台 3-1)

E-mail: tagchan@bosai.go.jp

地震や台風の発災直後に撮影された高解像度光学衛星画像を用いて、インスタンスセグメンテーション手法である Mask R-CNN により画像内の建物を自動抽出し、建物被害を無被害、損傷、倒壊、およびブルーシート被覆有無に自動分類するモデルを開発した。結果、建物抽出精度 (IoU) については約 35%、ピクセル毎の建物被害分類精度 (F 値) については約 52% となり、建物抽出と被害分類を同時に実施可能なモデルとして一定の性能をもつことを確認した。また、3 種の高解像度衛星画像を使用してモデルを構築したところ、IoU で約 39%、F 値については無被害が約 92%、損傷が約 69%、倒壊が約 56%、被覆有が約 85% であり、モデルの汎用性が一定程度あり、被害の早期把握に利用可能であることを確認した。

**Key Words:** Building detection, Damage estimation, Optical satellite, Deep learning, Mask R-CNN

## 1. はじめに

筆者らは自然災害発生直後に被害状況を迅速に把握し、災害対応支援に資する情報提供を行うことを目的とした研究開発を行っている。被害状況早期把握手段の一つとして光学衛星による画像取得が挙げられる。これは平常時においては観測が周期的に行われること、災害時には国際災害チャータ<sup>1)</sup>やセンチネル・アジア<sup>2)</sup>等の国際協定に基づき多数の衛星による緊急観測が可能なこと、航空機等と比較して広範囲の観測が可能であること等、利点が多い。また、画像取得から解析までを一元化し、基盤的情報流通ネットワーク<sup>3)</sup>を通じて情報プロダクトとして提供するワンストップシステム<sup>4)</sup>が開発されることにより、従来以上に迅速なデータ取得が可能となり、衛星データの利活用がより促進されることが期待されている。

本研究では、建物 1 棟毎の被害状況把握を目的とし、災害発生直後に取得された地上分解能が 50cm/pixel またはそれ以上の高解像度光学衛星を複数種類収集し、建物

抽出および被害分類を行うモデルを開発した。なお、既往研究<sup>5)</sup>においても画素単位でクラス分類を行うセマンティックセグメンテーション手法である U-Net を用いて高解像度光学衛星による建物被害抽出モデルを開発したが、本研究では画像中の物体検出とクラス分類を同時に行うインスタンスセグメンテーション手法である Mask R-CNN を使用し、建物の形状そのものの予測が可能になるように改良を行った。また、既往研究<sup>5)</sup>では熊本地震直後に撮影された WorldView-3 画像を使用した<sup>6)</sup>が、本研究では複数の災害において取得された、撮影時期や解像度の異なる高解像度光学衛星画像を用いてモデル構築を行っており、従来よりもモデルの汎用性を向上させることを目的としている。

## 2. 建物被害判読データ作成について

### (1) 使用した衛星画像

本研究において使用した高解像度光学衛星画像は、2016 年熊本地震本震直後の 5 月 5 日に撮影された

WorldView-3 画像の中から抽出した益城町, 西原村, 南阿蘇村の各地域合計約 120km<sup>2</sup>の領域 (図-1), 2018 年大阪府北部地震直後の 8 月 4 日に撮影された Pleiades 画像の中から抽出した高槻市付近における約 25km<sup>2</sup>の領域 (図-2), 2019 年山形県沖地震直後の 6 月 26 日に撮影された GeoEye-1 画像の中から抽出した鶴岡市小岩川地区付近の約 25km<sup>2</sup>の領域 (図-3 左), 2019 年台風 15 号直後の 9 月 25 日に撮影された GeoEye-1 画像の中から抽出した館山市米良地区付近の約 25km<sup>2</sup>の領域 (図-3 右) の各画像を使用した (表-1). なお, WorldView-3, GeoEye-1 画像は日本スペースイメーシング株式会社によりオルソ処理された画像を使用した. リサンプリングは 16bit, 内挿法は Cubic Convolution である. また Pleiades 画像は株式会社パスコによりオルソ処理された画像を使用した. いずれも RGB+近赤外の 4 バンドであるが, 本研究で使用する MaskR-CNN のモデルネットワークにおける入力 は 3ch であるため, RGB の 3 バンドのみを使用した. なお, 6 章で後述する異種画像を用いた汎用性の検証においては, 学習用データのサイズを合わせる為, 各画像の地上分解能を 50cm に統一している.

表-1 使用した衛星画像

対象災害	地域	種別	地上分解能
2016 熊本地震	益城町	WorldView-3	30cm
	西原村		
	南阿蘇村		
2018 大阪府北部地震	高槻市	Pleiades	50cm
2019 山形県沖地震	鶴岡市	GeoEye-1	40cm
2019 台風 15 号	館山市	GeoEye-1	40cm

## (2) 建物被害判読データ

次に, 国土地理院の基盤地図情報における建築物ポリゴンを基本とし, 一部は NTT インフラネット株式会社の GEOSPACE 電子地図を使用し, 建物形状に合わせて若干の修正を行った建物ポリゴンデータを作成した. このデータを GIS ソフトにより衛星画像と重ね合わせ, 目視判読により被害区分を無被害, 損傷, 倒壊の 3 段階に区分するとともに, ブルーシートによる屋根被覆の有無 (以下, 被覆有り・被覆無しと略) を記録し, 計 5 クラスの建物被害判読データ計 87,625 個を作成した (表-2). 建物被害判読データの作成範囲を図-1~図-3 において青枠で示す. なお, 目視判読による区分の基準は既往研究<sup>9)</sup>と同様であり, 建物の一部に被害や補修が確認できる場合を損傷, 建物全体が大きく変形もしくは破壊している場合を倒壊としている. 例として, 建物被害判読デー

タの内, 熊本地震における益城町における拡大図を図-4 に示す.

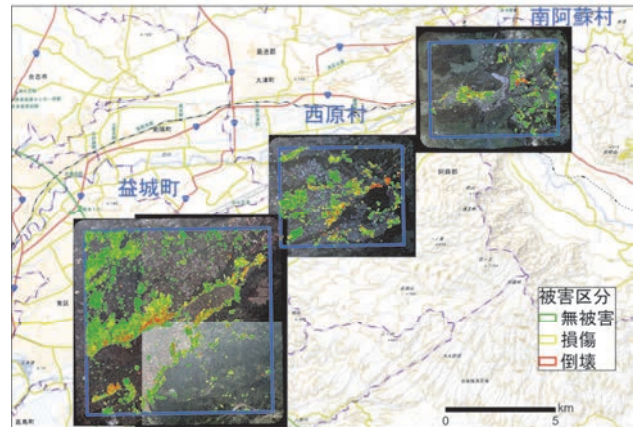


図-1 衛星画像と建物被害判読データの範囲 (熊本地震)

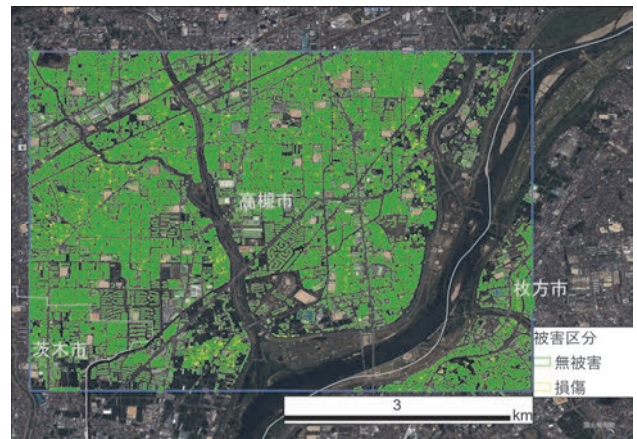


図-2 衛星画像と建物被害判読データの範囲 (大阪府北部地震)

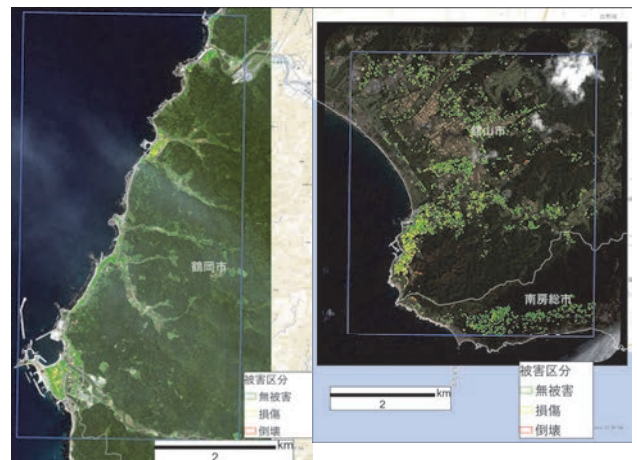


図-3 衛星画像と建物被害判読データの範囲 (左: 山形県沖地震, 右: 台風 15 号)

表-2 建物被害判読データの数

クラス	1	2	3	4	5	合計
	無被害	損傷		倒壊		
被覆	無	無	有	無	有	
益城町	15,052	1,452	3,721	1,365	114	21,704
西原村	2,817	541	679	255	89	4,381
南阿蘇村	1,145	234	182	104	10	1,675
高槻市	52,566	2	967	0	0	53,535
鶴岡市	1,348	30	175	1	0	1,554
館山市	3,939	35	795	7	0	4,776
合計	76,867	2,294	6,519	1,732	213	87,625

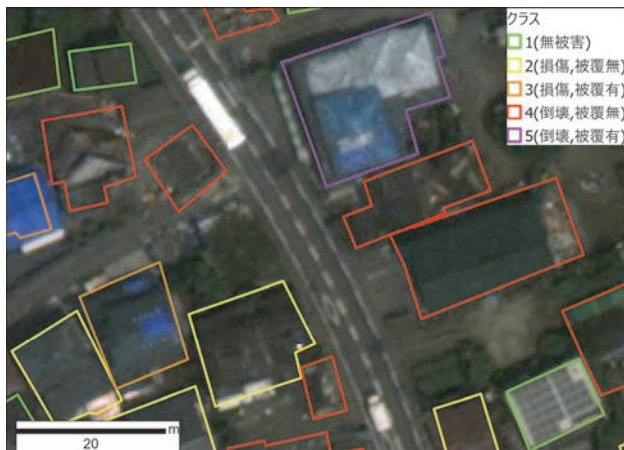


図-4 建物被害判読データの例（拡大図）

### (3) マスク画像

続いて、建物被害判読データを地図上に表示し、表-3に示すように分類ラベル毎に色分けしたマスク画像（図-5）を作成した。この画像は、深層学習モデル構築の際に、正解データとして使用した。なお、後述する被害分類においては、これらのクラスをまとめて、無被害（クラス1）、損傷（クラス2+3）、倒壊（クラス4+5）の3段階、あるいは被覆無し（クラス1+2+4）、被覆有り（クラス3+5）の2段階に区分している。

表-3 分類ラベルおよび配色

クラス	被害	被覆	色
0	建物以外		黒(黒)
1	無被害	無し	緑(緑)
2	損傷	無し	黄(黄)
3		有り	橙(橙)
4	倒壊	無し	赤(赤)
5		有り	紫(紫)

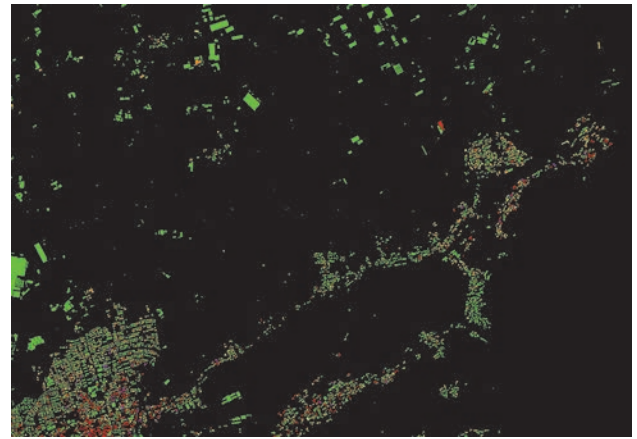
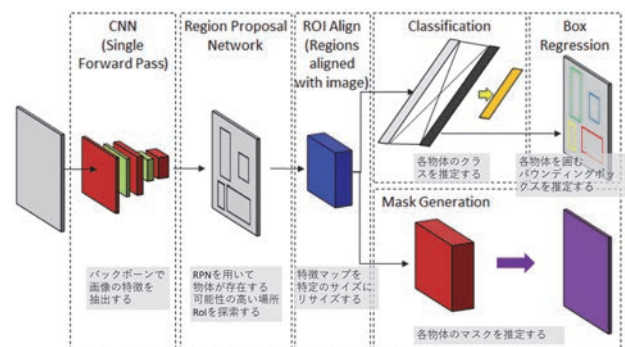


図-5 建物被害判読マスク画像の例（益城町）

## 3. モデルについて

### (1) Mask R-CNN

本研究では、代表的なインスタンスセグメンテーション手法であるMask R-CNN<sup>6)</sup>を使用している。Mask R-CNNではバックボーンでの画像全体の特徴をCNN（Convolutional Neural Network）により抽出後、検出した物体が存在する可能性が高い長方形の領域（RoI, Region of Interest）を畳み込みニューラルネットワークにより探索するRPN（Region Proposal Network）、RoIの位置を維持しながら特徴マップを特定サイズにリサイズし、各物体のクラスやバウンディングボックスの推定を行うRoI Alignの順に処理を行った後、各物体のクラスを推定してからバウンディングボックスを推定するBox Regressionと各物体のマスクを推定するMask Generationを並行して処理する構造を有している（図-6）。


図-6 Mask R-CNNの構成（文献<sup>7)</sup>に加筆）

また、Mask R-CNNにおいては多数のハイパーパラメータの設定が必要となることから、多くの組み合わせを試行し、効率的な学習を行うための工夫として、以下に示すバイズ最適化および枝刈りの2つの技術を適用した。



## (2) ベイズ最適化によるパラメータサーチ

深層学習におけるハイパーパラメータの組み合わせを選択する手法として、パラメータを一定間隔で変化させ、全ての組み合わせを試行するグリッドサーチ (Grid Search)、パラメータをランダムで決定し、試行を繰り返すランダムサーチ (Random Search)、試行すべきパラメータを統計的に推定しつつ試行を繰り返すベイズ最適化 (Bayesian Optimization) の3種類が有名であるが、本研究では、パラメータサーチの効率化およびモデルの性能向上を目的として統計的推定に基づくベイズ最適化を採用した (図-7)。

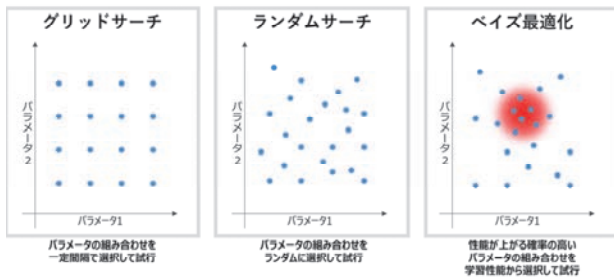


図-7 パラメータサーチの方法

ベイズ最適化の手法としては複数あるが、近年広く使用されているハイパーパラメータ最適化フレームワーク Optuna を使用し、TPE<sup>8)</sup> (Tree-structured Parzen Estimator) というアルゴリズムによりベイズ最適化を行った。TPE では以下①～③の処理が繰り返し実行される<sup>9)</sup>。

- ① パラメータ (input) を  $x$  に設定したときのスコア (output) を  $y$  とする。また、得られた  $y$  の値の集合において上位  $\gamma\%$  のスコアを  $y^*$  と定義する。このとき、試行結果を  $y^*$  以上の集合  $G$  とその他の集合  $L$  に分割する。 $G$  に属する確率密度関数を  $g(x)$ 、 $L$  に属する確率密度関数を  $l(x)$  とする (式(1))。

$$p(x|y) = \begin{cases} l(x) & \text{if } y < y^* \\ g(x) & \text{if } y \geq y^* \end{cases} \quad (1)$$

- ② 集合  $G$ 、 $L$  の各確率密度関数についてパラメータセットごとに、Parzen Estimator (カーネル密度推定法) による推定を行う。
- ③ 式(2)において定義されている Expected Improvement (EI) が最大となるパラメータセットを試行する。

$$EI_y(x) = \frac{g(x)}{l(x)} \quad (2)$$

## (3) 枝刈りによる学習効率向上

枝刈りとは、性能が高くなると予想される組み合わせについて、学習処理を途中で打ち切る手法である。本研究では、多数のハイパーパラメータの組み合わせについて、指定したエポック数だけ1つずつ学習処理を行い、学習後のモデルの損失値が最も低いモデルを採用している。ここで、学習処理の途中経過において損失値が下がらず、モデルの性能が高くなると推定される場合については学習処理を打ち切ることにより学習効率を向上させた (図-8)。

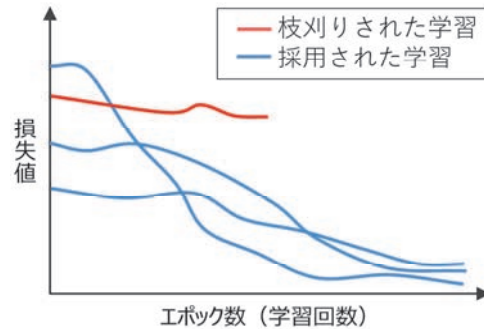


図-8 枝刈りのイメージ

## 4. データ取得およびモデル構築方法

### (1) 学習・検証・評価用データ取得

本研究では、準備として、各衛星画像から建物被害判読データの作成範囲 (図-1～図-3 の青枠部分) を切り出した後、各バンドにおける輝度値の平均  $\mu$ 、標準偏差  $\sigma$  を元に、以下の  $\beta$  の範囲が 0～1 になるように正規化した ( $R_{norm}$  は正規化のための係数) (式(3))。

$$\beta = \mu \pm R_{norm} \cdot \sigma \quad (3)$$

次に、この画像全域を用いて、各画像の左上から右下にかけて一定のピクセル幅 (patch\_size) のウインドウを、上下左右一定のピクセル幅 (overlap\_size) だけ重なるように走査してパッチ画像を切り出した。また、マスク画像についても同じサイズのパッチ画像を抽出し、正解ラベルの学習・検証・評価用に使用した。ただし、衛星画像内には建物が存在しない領域の割合が多い為、走査時に建物が全く含まれないパッチ画像については一定割合 (non\_building\_rate) のみランダムに取得し、学習用データとして使用した。

学習用+検証用データとしては衛星画像上部の指定した領域割合 (training\_data\_rate) から取得したパッチ画像を使用し、そのうちランダムに取得した一定割合

(validation\_data\_rate) を検証用データ, 残りを学習用データとして使用した. また, 評価用データとしては衛星画像下部の残り (1- training\_data\_rate) の領域割合からパッチ画像を取得した. このような手順により取得した学習用, 検証用, 評価用データの各領域の一例を衛星画像上に投影した図を図-9 に示す.

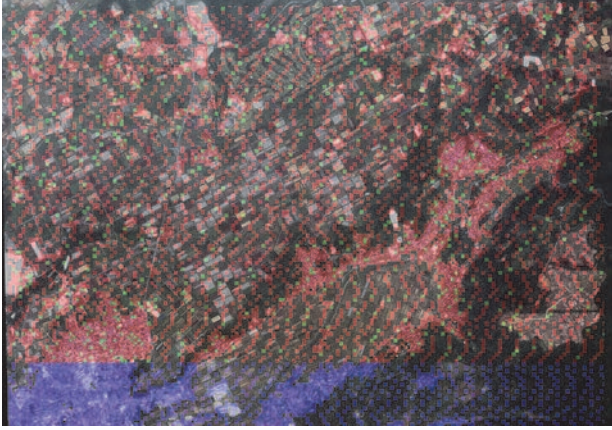


図-9 学習・検証・評価用パッチ画像のイメージ  
(赤: 学習用, 緑: 検証用, 青: 評価用)

### (3) モデル構築

本研究ではGitHub上で公開されているオープンソースライセンスの Mask R-CNN プログラム<sup>10)</sup>をダウンロードして使用した. 開発環境としては Python 言語, および深層学習フレームワークの Tensorflow を使用した. 本研究で使用した機械学習関連ソフトウェアのバージョンについて表-4 に示す. また, 学習・評価を行った計算機環境について表-5 に示す.

Mask R-CNN の構造は複雑であり, 一から性能を高めるためにはデータ量や学習時間に多大なコストを要するため, モデルネットワークのヘッダー部分に ResNet50 の重みファイル<sup>11)</sup>を使用し, 転移学習を行った.

表-4 使用した機械学習関連ソフトウェアのバージョン

Python	3.8.12
datatoolkit	11.3.1
cdnn	8.2.1 cuda11.3_0
optuna	2.10.0
tensorflow	tensorflow-gpu 2.5.0

表-5 計算機環境

マシン名	ELSA VELUGA-D 5000D
OS	Windows 10 Pro
CPU	AMD Ryzen9 5900X (12 コア/24 スレッド)
メモリ	64GB
GPU	NVIDIA RTX 3090 (24GB)

損失関数としては多クラス交差エントロピー (式(4)) を使用した. 学習用データと検証用データを比較し, 正解ラベル  $t_k$  と予測結果  $y_k$  との誤差である loss 値 ( $E$ ) が最小になるように最適化を行った.

$$E = - \sum_k t_k \log y_k \quad (4)$$

このような学習を繰り返し行い, Mask R-CNN のハイパーパラメータ約 40 種類の中から, 表-6 に示す各パラメータを変化させ, ベイズ最適化によるパラメータサーチを行った. 各出力結果から, 200epoch の学習試行における学習用・検証用データの loss 値が最小となる重みファイルを選択し, その中から評価用データのピクセル毎の予測精度が比較的良くなるパラメータを採用した.

表-6 ベイズ最適化によるパラメータサーチの対象

パラメータ名	設定範囲
DETECTION_MIN_CONFIDENCE	0.6~0.8
DETECTION_NMS_THRESHOLD	0.01~0.2
LEARNING_RATE	$1.0 \times 10^{-5} \sim 1.0 \times 10^{-4}$
RPN_ANCHOR_SCALES	(16,32,64,128), (32,64,128,256), (16,32,64,128,256)
RPN_NMS_THRESHOLD	0.7~0.8
RPN_TRAIN_ANCHORS_PER_IMAGE	128, 64, 256

採用した試行結果における, epoch 毎の loss 値の推移の例を図-10 に示す. このように 200epoch 程度の学習で検証用データの loss 値が概ね収束していることを確認している.

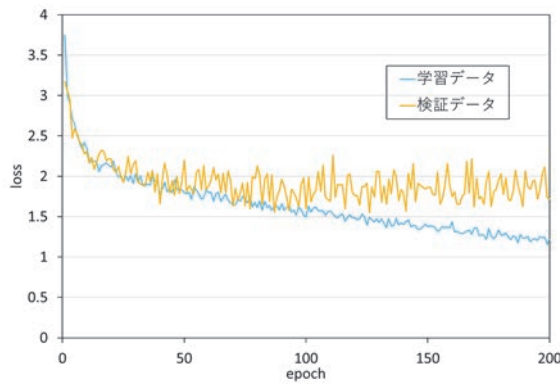


図-10 epoch 毎にみた loss 値の推移

#### (4) 被害予測結果のマッピング

各パッチ画像における Mask R-CNN による被害予測結果は図-11 のようになり、表-3 に示したクラス毎に色分けされた推定領域および各物体を表すバウンディングボックスがモデルから出力される。例えば、図中の紫・赤色は倒壊、橙色は損傷、緑色は無被害と、それぞれ推定された領域を表す。

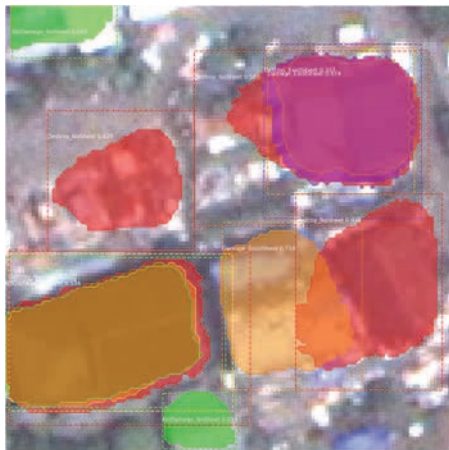


図-11 Mask R-CNN による被害予測結果の例

衛星画像全域における被害予測結果作成にあたり、各パッチ画像におけるモデルの出力結果を、元の衛星画像全域において左上から右下に向かってオーバーラップなしでつなぎ合わせ、GeoTIFF 形式のファイルとして保存し、マッピングを行った(図-12)。なお、マッピングの際、画素に複数の推定領域が重なっている場合には被害が大きいクラスを優先した。また、モデルから出力された建物被害推定領域については被害区分ラベルを付与したシェープファイル形式のデータとしても保存可能になっており、これらの予測結果を GIS ソフト等で地図上に表示し、分析することを可能にした。以下では、このモデルにおける被害予測結果について説明するが、それ

ぞれの予測結果については最適化されたモデルを用いた 1 回の試行結果である。



図-12 マッピングされた被害予測結果の例

### 5. 既往モデルとの予測精度比較

#### (1) 本研究におけるモデルの学習条件

本研究において開発した Mask R-CNN を用いた建物被害抽出モデルの精度を評価するため、既往研究<sup>5)</sup>において開発した U-net を用いたモデルを元に、入力チャンネル数を 4ch (RGB+近赤外) から 3ch (RGB) に修正したモデルとの比較を行った。衛星画像および教師データとしては既往研究<sup>5)</sup>と同一になるように、熊本地震直後に益城町、西原村、南阿蘇村を含む範囲において撮影された WorldView-3 画像 5 枚、および図-1 に示した各地域における建物被害判読データを使用した。また、データ取得方法等、学習条件は可能な限り両手法で同様になるように設定したが、パラメータサーチ方法と転移学習の有無が異なっている(表-7)。

データ取得、および学習に係る条件は表-8 のように設定した。なお、パラメータサーチの試行時間はハイパーパラメータの組み合わせの数と 1 モデル当たりにかかる計算時間の観点から 3 日間に設定した。

表-7 各モデルにおける学習条件の比較

項目	既往研究 <sup>5)</sup>	本研究
使用モデル	U-net	Mask R-CNN
パッチ画像	128pixel	128pixel
オーバーラップ	16pixel	16pixel
チャンネル数	3ch	3ch
転移学習	なし	あり
パラメータサーチ	グリッドサーチ	ベイズ最適化
エポック数	200	200

表-8 既往モデルとの比較におけるデータ取得・学習条件

設定項目	値	説明
$R_{\text{norm}}$	1.95	画像正規化の係数
patch_size	128	パッチ画像のサイズ(pixel)
overlap_size	16	パッチ画像のオーバーラップ(pixel)
non_building_rate	0.2	建物が写っていないパッチの取得割合
training_data_rate	0.8	学習用+検証用データの取得割合
validation_data_rate	0.1	検証用データの取得割合
trial_max_time	259,200	パラメータサーチの試行時間(秒)

## (2) ピクセル毎の被害予測精度比較

建物被害予測マスク画像(例えば図-5)を正解とし、既往研究<sup>9)</sup>におけるU-netモデル、および本研究におけるMask R-CNNモデルをそれぞれ用いてマッピングされた予測結果との比較を行った。精度評価の指標としては、正解とするピクセル領域と予測されたピクセル領域における積集合(AND)と和集合(OR)のピクセル比率であるIoU(式(5))、各ピクセル領域の積集合(AND)と正解データ(truth)とのピクセル比率である再現率(Recall)(式(6))、各ピクセル領域の積集合(AND)と予測データ(pred)とのピクセル比率である適合率(Precision)(式(7))、およびRecallとPrecisionとの調和平均であるF値(Fmeasure)(式(8))を用いた。

$$IoU_{\text{pred}}^{\text{truth}} = \frac{\text{pred} \cap \text{truth}}{\text{pred} \cup \text{truth}} \quad (5)$$

$$\text{Recall} = \frac{\text{pred} \cap \text{truth}}{\text{truth}} \quad (6)$$

$$\text{Precision} = \frac{\text{pred} \cap \text{truth}}{\text{pred}} \quad (7)$$

$$F\text{measure} = \frac{2}{\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}} \quad (8)$$

既往研究(U-net)モデルにおけるピクセル毎の予測精度評価指標(表-9)では学習用データにおける全建物のIoUが約71%、評価用データにおける全建物のIoUが約46%であるのに対し、本研究(Mask R-CNN)モデル(表-10)では学習用データにおける全建物のIoUが約42%、評価用データにおける全建物のIoUが約35%となり、U-netモデルの方がやや建物形

状の抽出精度がやや高い結果となった。結果を分かりやすくするため、図-13に評価用データにおけるU-net、Mask R-CNN各モデルのIoU(青色)の比較を示す。

表-9 既往研究<sup>9)</sup>(U-net)におけるピクセル毎予測精度(%)

		IoU	Recall	Precision	Fmeasure
学習用データ	無被害	74.9	84.6	86.8	85.7
	損傷	64.8	74.3	83.4	78.6
	倒壊	53.2	63.0	77.4	69.5
	被覆無	70.6	79.3	86.5	82.8
	被覆有	71.5	81.4	85.5	83.4
	全建物	71.3	81.1	85.6	83.3
評価用データ	無被害	55.6	67.8	75.5	71.4
	損傷	36.4	48.1	59.8	53.3
	倒壊	28.4	36.5	56.3	44.3
	被覆無	48.7	60.9	70.8	65.5
	被覆有	45.6	57.3	69.1	62.6
	全建物	46.1	57.9	69.4	63.1

表-10 本研究(Mask R-CNN)におけるピクセル毎予測精度(%)

		IoU	Recall	Precision	Fmeasure
学習用データ	無被害	46.3	69.2	58.4	63.3
	損傷	38.0	51.9	58.7	55.1
	倒壊	17.5	37.1	24.8	29.7
	被覆無	48.5	63.0	67.9	65.3
	被覆有	41.0	63.4	53.7	58.2
	全建物	42.1	63.3	55.7	59.3
評価用データ	無被害	40.8	62.2	54.2	57.9
	損傷	29.8	42.5	49.8	45.9
	倒壊	24.0	44.2	34.5	38.7
	被覆無	42.3	58.5	60.5	59.5
	被覆有	33.7	53.5	47.6	50.4
	全建物	34.9	54.3	49.5	51.7



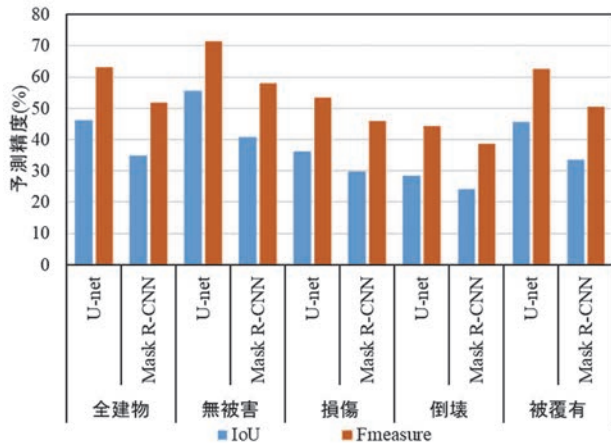


図-13 評価用データにおけるピクセル毎の予測精度比較

また、各被害区分のF値（茶色）を比較すると、U-netモデルでは無被害が約71%、損傷が約53%、倒壊が約44%、被覆有りが約63%であるのに対し、Mask R-CNNモデルでは無被害が約58%、損傷が約46%、倒壊が約39%、被覆有りが約50%であり、これについてもU-netモデルの精度の方がやや高くなった。また、いずれのモデルでも無被害や被覆有りに比べて、損傷や倒壊では精度がやや低くなった。このようにU-netモデルのほうがMask R-CNNモデルよりも予測精度が高くなったが、後者は個々の建物を一つの物体として認識しており、建物ポリゴンを別途用意しなくても建物形状の予測が可能である点で優れていると考えている。

加えて、被災度が上がるほどピクセル毎の予測精度が下がる結果となったが、これは主に、高い被災度に区分したラベルほど学習用データ数が少ないことに起因するものと考えられる。また、倒壊やブルーシート被覆等、被災の特徴を示す箇所は建物の一部であることが多いのに対し、建物被害判読データ（2章(2)参照）は建物全体を囲んだ領域を1つのクラスとみなしていることから、ピクセル毎の評価とは判断基準が異なり、このことが精度評価に影響を与えていると考えられる。視点を変えると、U-netやMask R-CNNを適用することにより、建物に生じた部分的な損傷箇所を抽出できている可能性があるということもできる。

図-14にU-net（上）、Mask R-CNN（中）の各モデルにおける評価用画像のマッピング結果、および正解ラベル（下）の例を示す。各図を比較すると被害区分等の予測結果が若干異なっており、Mask R-CNNモデルでは道路の一部を無被害建物と判断する等、若干の誤分類が確認できる。しかし、いずれのモデルを使用した場合においても全体的な被害程度や被害集中箇所等、被害の概要把握という観点では十分利用可能な程度に、建物位置や被害箇所を予測できていることが分かる。



図-14 U-net（上）、Mask R-CNN（中）各モデルによる予測結果、および正解クラスラベル（下）との比較



### (3) 建物毎の被害予測精度比較

本研究では、ピクセル毎の被害予測結果は建物内における詳細な被害の特徴を示しているという前提に基づき、建物全体の被害をより高精度に予測する工夫として、建物毎の予測ピクセルの比率に基づいた被害予測を行った。具体的には2章で作成した建物被害判読データを使用し、各ポリゴン内の予測クラスのピクセル比率を元に、**図-15**のフローチャートに示されたルールに従って建物被害の予測を行った。なお、ルール作成にあたり、災害対応への活用の観点から、特に倒壊建物の見逃しを小さくすることを重視した。

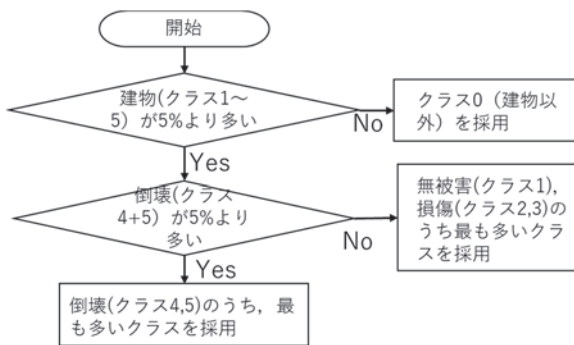


図-15 建物被害予測のフローチャート

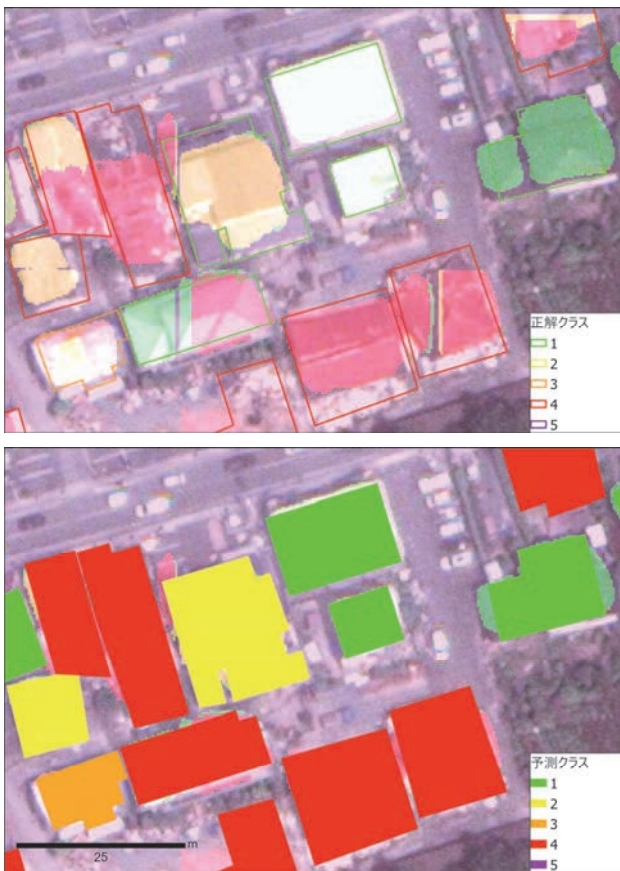


図-16 Mask R-CNNによる予測結果(上)と建物毎の予測結果(下)

このルールに従った建物被害区分の例を**図-16**に示す。上図において線で囲まれた部分は建物ポリゴン、線の色は正解クラスラベルであり、塗りつぶされた領域はMask R-CNNによる予測領域である。また、下図において塗りつぶされた領域の色は**図-15**のルールに従って予測された建物毎の予測結果をそれぞれ示している。

このようなルールに基づく建物毎の被害予測結果の精度検証を目的として、建物被害判読データを正解とし、全画像における予測結果の精度を混同行列により評価した。本研究に使用した学習用データ、評価用データの被害区分毎の内訳を**表-11**に、U-netモデルにおける評価用データの精度評価結果を**表-12**に、Mask R-CNNモデルにおける評価用データの精度評価結果を**表-13**にそれぞれ示す。なお、**表-12**、**表-13**における建物毎の被害予測における再現率、適合率は、ピクセル数ではなく、正解クラスと予測クラスのラベルが一致するかどうかを基準としている。

表-11 学習用データ、評価用データの被害区分別内訳

	無被害	損傷	倒壊	合計
学習用	14,468	4,896	1,126	20,490
検証用	1,190	366	76	1,632
評価用	2,697	1,296	733	4,726
合計	18,355	6,558	1,935	26,848
	被覆無し		被覆有り	合計
学習用	16,933		3,557	20,490
検証用	1,378		254	1,632
評価用	3,957		769	4,726
合計	22,268		4,580	26,848

表-12 既往研究<sup>5)</sup>モデル(U-net)における精度評価結果

被害区分		予測					再現率
		無被害	損傷	倒壊	他	合計	
正解	無被害	2,232	239	60	166	2,697	82.8
	損傷	313	787	149	47	1,296	60.7
	倒壊	109	114	449	61	733	61.3
	合計	2,654	1,140	658	274	4,726	-
適合率		84.1	69.0	68.2	-	-	-
F値		83.4	64.6	64.6	-	正答率	73.4
被覆		予測					再現率
		無し	有り	他		合計	
正解	無し	3,579	126	252		3,957	90.4
	有り	94	653	22		769	84.9
	合計	3,673	779	274		4,726	-
適合率		97.4	83.8	-		-	-
F値		93.8	84.4	-		正答率	89.5

表-13 本研究モデル(Mask R-CNN)における精度評価結果

被害区分		予測					再現率
		無被害	損傷	倒壊	他	合計	
正解	無被害	1,903	265	277	252	2,697	70.6
	損傷	168	690	340	98	1,296	53.2
	倒壊	23	44	581	85	733	79.3
	合計	2,094	999	1,198	435	4,726	-
適合率		90.9	69.1	48.5	-	-	-
F 値		79.4	60.1	60.2	-	正答率	67.2
被覆		予測					再現率
		無し	有り	他		合計	
正解	無し	3,420	147	390		3,957	86.4
	有り	99	625	45		769	81.3
	合計	3,519	772	435		4,726	-
適合率		97.2	81.0	-		-	-
F 値		91.5	81.1	-		正答率	85.6

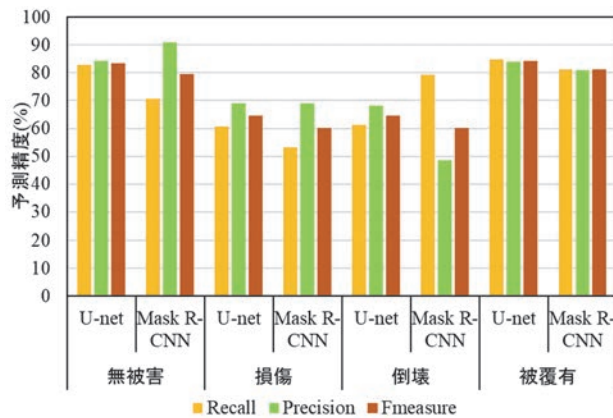


図-17 各モデルによる予測精度比較

図-17 に U-net, Mask R-CNN 各モデルの評価用データにおける予測精度の比較を示す。F 値（茶色）について比較すると、U-net モデルでは無被害が約 83%, 損傷が約 65%, 倒壊が約 65%, 被覆有りが約 84% であり、Mask R-CNN モデルでは無被害が約 80%, 損傷が約 60%, 倒壊が約 60%, 被覆有りが約 81% となった。各区分の F 値や正答率から見るとやや U-net モデルの方が高精度であるが、いずれのモデルも各被害区分の F 値が 60% 以上であり、高い予測精度を持っていることを確認した。また、Mask R-CNN モデルでは特に倒壊の再現率 (Recall) が高くなっており、見逃しが少ないモデルになっていることが分かる。

## 6. 異種画像を用いた汎用性の検証

本研究の成果を今後発生する災害において活用するためには、モデルの汎用性を向上させることが重要である。本研究では、災害発生直後に入手可能である複数の衛星画像に対応できるよう、WorldView-3, GeoEye-1, Pleiades という解像度が異なる 3 種類の高解像度光学衛星画像及び建物被害判読データを活用し、各画像からそれぞれ学習用・検証用・評価用データを取得後、各画像が混在した学習用・検証用データによりモデル構築を行い、各評価用データにおける予測精度を検証し、モデルの特徴について考察した。

解析にあたり、各画像の地上分解能は 1 ピクセルあたり WorldView-3 で 30cm, GeoEye-1 で 40cm, Pleiades で 50cm とそれぞれ異なっているため、事前に各画像の地上分解能を 50cm に合わせた。

学習データの取得手順は 4 章と同様であるが、設定条件がやや異なる (表-14)。具体的には、衛星種別が増えたことにより輝度値のばらつきが大きくなったため  $R_{norm}$  を 2.1 に増加させた、また評価用データをより広域から取得する為、`training_data_rate` を 0.7 にした。

表-14 汎用性検証におけるデータ取得・学習条件

設定項目	値	説明
$R_{norm}$	2.1	画像正規化の係数
<code>patch_size</code>	128	パッチ画像のサイズ (pixel)
<code>overlap_size</code>	16	パッチ画像のオーバーラップ (pixel)
<code>non_building_rate</code>	0.2	建物が写っていないパッチの取得割合
<code>training_data_rate</code>	0.7	学習用+検証用データの取得割合
<code>validation_data_rate</code>	0.1	検証用データの取得割合
<code>trial_max_time</code>	259,200	パラメータサーチの試行時間 (秒)

学習用・検証用・評価用データ数の被害区分別内訳を画像毎に集計したものを表-15 に示す。なお、学習時は 4 章と同様、ベイズ最適化によるハイパーパラメータサーチを行い、`loss` 値が最小となった試行のモデルを採用した。

表-15 学習用データ, 評価用データ数の被害区分別内訳

		無被害	損傷	倒壊	合計
学習用	WorldView-3	13,264	4,304	941	18,509
	Pleiades	39,065	733	0	39,798
	GeoEye-1	3,348	560	2	3,910
	合計	55,677	5,597	943	62,217
検証用	WorldView-3	998	353	76	1,427
	Pleiades	3,137	51	0	3,188
	GeoEye-1	232	48	1	281
	合計	4,367	452	77	4,896
評価用	WorldView-3	4,819	2,138	961	7,918
	Pleiades	9,659	178	0	9,837
	GeoEye-1	1,650	422	5	2,077
	合計	16,128	2,738	966	19,832
		被覆無し	被覆有り	合計	
学習用	WorldView-3	15,392	3,117	18,509	
	Pleiades	39,065	733	39,798	
	GeoEye-1	3,384	526	3,910	
	合計	57,841	4,376	62,217	
検証用	WorldView-3	1,162	265	1,427	
	Pleiades	3,137	51	3,188	
	GeoEye-1	235	46	281	
	合計	4,534	362	4,896	
評価用	WorldView-3	6,526	1,392	7,918	
	Pleiades	9,659	178	9,837	
	GeoEye-1	1,684	393	2,077	
	合計	17,869	1,963	19,832	

各画像種別におけるピクセル毎の被害予測精度を図-18に, また詳細な結果を表-16に示す. 建物形状の予測指標であるIoU(青色)について, 全建物の場合において比較すると, WorldView-3(図中のWV3)画像が約34%, Pleiades画像が約51%, GeoEye-1画像が約21%であり, Pleiades画像が最も建物形状の抽出精度が高かった. また, 被害区分毎の予測精度をF値(茶色)により比較すると, WorldView-3画像では無被害が約55%, 損傷が約47%, 倒壊が約31%, 被覆有が約60%であった. Pleiades画像では無被害が約68%, 損傷が約55%, 被覆有が約55%であった. GeoEye-1画像では無被害が約33%, 損傷が約58%, 倒壊が約11%, 被覆有が約61%であった. また, 全画像の集計値では無被害が約58%, 損傷が約49%, 倒壊が約31%, 被覆有が約60%であった(表-16). これらの結果から, ピクセル単位の予測精度は無被害や被覆有については6割程度である一方, データ数の少ない倒壊では3割程度とやや低くなった.

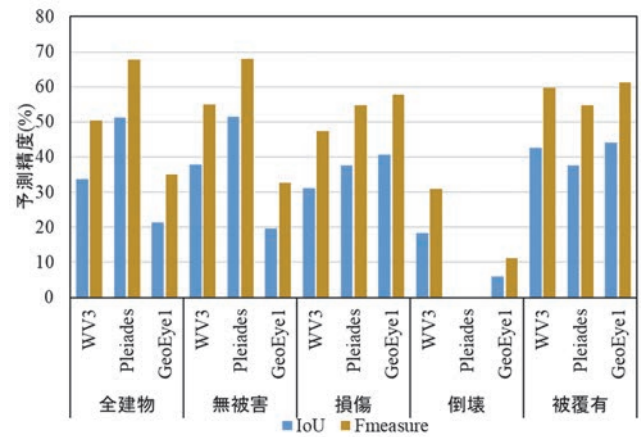


図-18 衛星画像種別におけるピクセル毎の予測精度比較

表-16 評価用データのピクセル毎予測精度(%)

		IoU	Recall	Precision	Fmeasure
WorldView-3	無被害	37.9	69.5	45.4	55.0
	損傷	31.1	45.6	49.4	47.4
	倒壊	18.2	34.8	27.7	30.8
	被覆無	32.3	59.1	41.7	48.9
	被覆有	42.5	59.4	59.9	59.7
	全建物	33.7	59.1	44.0	50.4
Pleiades	無被害	51.4	75.1	61.9	67.9
	損傷	37.7	44.2	71.8	54.7
	倒壊	-	-	-	-
	被覆無	51.4	75.1	61.9	67.9
	被覆有	37.7	44.2	71.8	54.7
	全建物	51.2	74.6	62.0	67.7
GeoEye-1	無被害	19.5	69.0	21.4	32.7
	損傷	40.5	51.3	65.9	57.7
	倒壊	5.9	6.0	90.0	11.2
	被覆無	19.3	66.5	21.4	32.4
	被覆有	44.1	57.2	65.9	61.3
	全建物	21.3	64.8	24.1	35.1
全画像	無被害	41.0	72.7	48.5	58.2
	損傷	32.6	46.3	52.4	49.2
	倒壊	18.1	34.3	27.8	30.7
	被覆無	38.3	67.9	46.8	55.4
	被覆有	42.4	57.7	61.5	59.6
	全建物	38.6	66.9	47.8	55.7



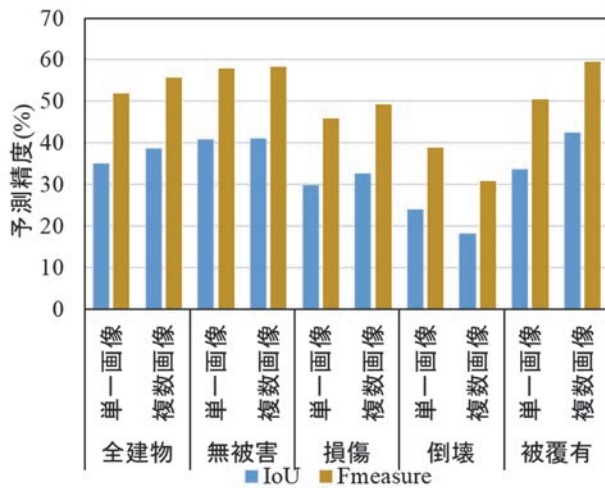


図-19 単一・複数画像におけるピクセル毎の予測精度比較

図-19に熊本地震のWorldView-3画像のみを用いて構築したMaskR-CNNモデルにおける被害判別精度（単一画像，5章の表-10参照）と複数の異種画像を使用したモデルにおける判別精度の全画像集計値（複数画像）との比較を示す。これを見ると，全建物のIoU（青色）は単一画像の約35%から複数画像では39%に向上し，被害区分毎では特に損傷のF値（茶色）が約46%から約49%に向上，被覆有りのF値が約50%から約60%に向上するなど，異種画像の使用により全体的な精度向上に寄与していることが確認できる。ただし，倒壊については精度が低下しており，これはこれらのクラスの教師データが熊本地震に集中しており，PleiadesやGeoEye-1画像において十分に学習用データが構築できなかったことに起因していると考えられる。

続いて，建物単位の被害予測結果について述べる。画像種別毎に分けた再現率（黄色），適合率（緑色），F値（茶色）の比較を図-20に，混同行列を表-17～表-20にそれぞれ示す。WorldView-3画像においては無被害のF値が約84%，損傷のF値が約67%，倒壊のF値が約56%，被覆有りのF値が約84%となった。Pleiades画像においては無被害のF値が約96%，損傷のF値が約76%，被覆有りのF値が約76%となった。GeoEye-1画像においては無被害のF値が約90%，損傷のF値が約79%，倒壊のF値が約33%，被覆有りのF値が約81%となった。全体の集計値は，無被害のF値が約92%，損傷のF値が約69%，倒壊のF値が約56%，被覆有りのF値が約85%となった。

結果，学習用データの数が不足しているGeoEye-1の倒壊のみF値が十分な精度を示していないが，それ以外はF値で約56～96%の範囲内となり，被害状況の早期把握に必要な情報として利用可能な水準にあることを確認した。

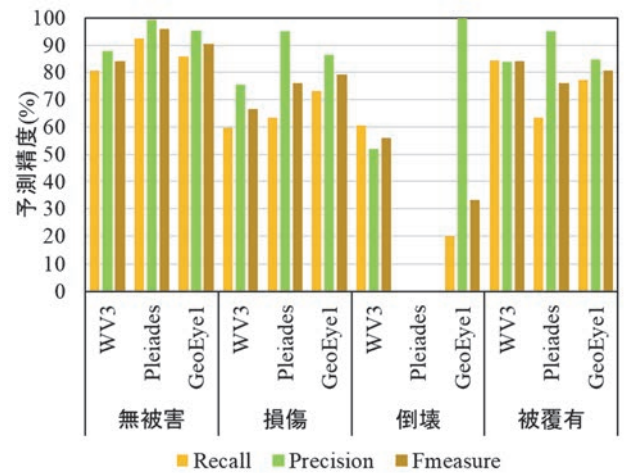


図-20 衛星画像種別における建物毎の予測精度比較

表-17 WorldView-3 画像における予測結果（評価用データ）

被害区分		予測					再現率
		無被害	損傷	倒壊	他	合計	
正解	無被害	3,882	310	259	368	4,819	80.6
	損傷	436	1,274	280	148	2,138	59.6
	倒壊	104	101	582	174	961	60.6
	合計	4,422	1,685	1,121	690	7,918	-
適合率		87.8	75.6	51.9	-	-	-
F値		84.0	66.6	55.9	-	正答率	72.5
被覆		予測					再現率
		無し	有り	他		合計	
正解	無し	5,673	226		627	6,526	86.9
	有り	154	1,175		63	1,392	84.4
	合計	5,827	1,401		690	7,918	-
適合率		97.4	83.9		-	-	-
F値		91.8	84.1		-	正答率	86.5

表-18 Pleiades 画像における予測結果（評価用データ）

被害区分		予測					再現率
		無被害	損傷	倒壊	他	合計	
正解	無被害	8,938	6	0	715	9,659	92.5
	損傷	55	113	0	10	178	63.5
	倒壊	0	0	0	0	0	-
	合計	8,993	119	0	725	9,837	-
適合率		99.4	95.0	-	-	-	-
F値		95.8	76.1	-	-	正答率	92.0
被覆		予測					再現率
		無し	有り	他		合計	
正解	無し	8,938	6	715		9,659	92.5
	有り	55	113	10		178	63.5
	合計	8,993	119	725		9,837	-
適合率		99.4	95.0	-		-	-
F値		95.8	76.1	-		正答率	92.0

表-19 GeoEye-1 画像における予測結果（評価用データ）

被害区分		予測					再現率
		無被害	損傷	倒壊	他	合計	
正解	無被害	1,416	49	0	185	1,650	85.8
	損傷	67	309	0	46	422	73.2
	倒壊	2	0	1	2	5	20.0
	合計	1,485	358	1	233	2,077	-
適合率		95.4	86.3	100	-	-	-
F値		90.3	79.2	33.3	-	正答率	83.1
被覆		予測					再現率
		無し	有り	他		合計	
正解	無し	1,433	55	196		1,684	85.1
	有り	53	303	37		393	77.1
	合計	1,486	358	233		2,077	-
適合率		96.4	84.6	-		-	-
F値		90.4	80.7	-		正答率	83.6

表-20 全画像における予測結果（評価用データ）

被害区分		予測					再現率
		無被害	損傷	倒壊	他	合計	
正解	無被害	14,236	365	259	1,268	16,128	88.3
	損傷	558	1,696	280	204	2,738	61.9
	倒壊	106	101	583	176	966	60.4
	合計	14,900	2,162	1,122	1,648	19,832	-
適合率		95.5	78.4	52.0	-	-	-
F値		91.8	69.2	55.8	-	正答率	83.3
被覆		予測					再現率
		無し	有り	他		合計	
正解	無し	16,044	287	1538		17,869	89.8
	有り	262	1,591	110		1,963	81.0
	合計	16,306	1,878	1648		19,832	-
適合率		98.4	84.7	-		-	-
F値		93.9	84.7	-		正答率	88.9

続いて、画像毎の予測結果の特徴について述べる。図-21 は本モデルを使用し、熊本地震直後の西原村における評価用画像を用いて被害予測を行った画像の一部である。上図（正解）と下図（予測）を比べると倒壊（赤）、被覆有（橙）、損傷（黄色）等、被害建物の分布を概ね捉えられていることが分かる。一方、建物以外の農地やビニールハウス等を無被害と誤判定するケースや、損傷等の被害建物を無被害と誤判定するケースが一部確認できる。また、下図ではマッピングする際に使用したパッチ画像の端部において、隣接パッチと予測クラスが異なっているケースが確認できる。このようなパッチ画像端部の予測クラスの不一致が誤分類に影響を与えている可能性がある。



図-21 World View-3 画像における建物被害予測結果（西原村）（上：正解，下：予測）



図-22は大阪府北部地震におけるテスト用画像を用いて被害予測を行った画像の一部である。無被害（緑）、被覆有（橙）建物の分布は概ね捉えられていることが分かる。



図-22 Pleiades 画像における建物被害予測結果  
（大阪府北部地震）（上：正解，下：予測）

図-23は台風 15 号におけるテスト用画像を用いて被害予測を行った画像の一部である。ビニールハウスや農地等、建物以外を無被害に誤分類している箇所が若干確認できるものの、被覆有（橙）建物の分布は概ね捉えられていることが分かる。

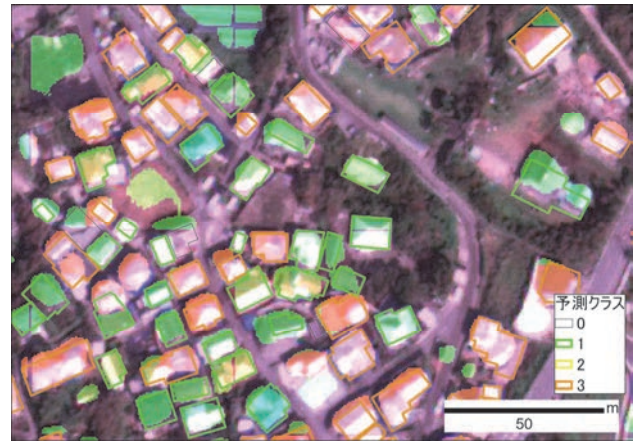


図-23 GeoEye-1 画像における建物被害予測結果  
（台風 15 号）（上：正解，下：予測）

以上の結果から、Mask R-CNN を用いた本モデルの特徴として、特に大阪北部地震等、建物が密集する市街地における建物検出およびブルーシート被覆有りの検出においては有効である一方で、農地等において建物以外の部分を無被害と判断する誤検出や、被害区分の誤分類を一部で確認した。しかし、各被害区分において F 値で 6 割程度以上の分類精度を示し、被害の概要把握に十分な水準であることを確認した。このことから、本モデルを活用することにより、事前に建物の位置が与えられていない状況においても、建物の分布を確認し、被害建物の概要を即時に把握することが可能であり、本研究において開発したモデルが有効であることが示された。

## 7. まとめおよび今後の課題

発災直後の被害状況把握への活用を目的として深層学習によるインスタンスセグメンテーションモデルである Mask R-CNN を用いた建物被害自動分類モデルを開発した。高解像度光学衛星である WorldView-3, GeoEye-1, Pleiades の各画像を用いて学習・評価を行ったところ、既往研究において構築したセマンテックセグメンテーションモデルである U-net を用いたモデルと概ね同様に、全体的な被災程度や被害集中地域等を自動抽出することが可能であることを確認した。また、従来モデルでは予測結果がピクセル毎に分割されていたのに対し、個々の建物毎に分かれたオブジェクトとして無被害・損傷・倒壊・被覆有に分類された建物のデータを抽出することが可能になり、目視による被害判読作業を要さず、被害状況把握の効率化に寄与することが可能になった。

ただし IoU, F 値等の精度評価指標に基づく予測精度は U-net モデルよりも若干劣っており、出力結果画像から推定した特徴として、Mask R-CNN モデルでは特に、



予測に使用したパッチ画像の端部で誤判定が多いことが確認された。精度向上の為の工夫として、パッチ画像から画像の結合・マッピングを行う際に、一定の幅でオーバーラップさせた後に結合し、各画像の中心から一定の領域に属する建物の予測結果のみをマッピングに使用することにより予測精度向上が可能であると考えられる。

また、現状は建物抽出と被害推定を同時に **Mask R-CNN** により行っているが、建物抽出と被害推定のモデルを分割し、被害推定については **U-net** による予測結果を併用する等、アルゴリズムの改良が予測精度向上に寄与する可能性がある。今後も教師データの拡充やモデル性能向上にむけた機能強化を進めるとともに、実際の災害対応における活用に向けて各機関と連携した取り組みを続けていきたい。

加えて、本稿では目視により一定の解像度や品質が保たれていることが確認された光学衛星画像を使用した。撮影条件によって画像の品質に差が生じる可能性がある。本稿で使用した輝度の正規化や、解像度の統一等は品質の差を埋める一つの工夫である。一方で、民間による衛星技術の活用促進に伴い、小型衛星を活用したコンステレーション等、今後より多様なセンシングデータが取得されることが見込まれるため、**RGB** だけでなく様々な周期帯における、多様な特徴を持ったリモートセンシングデータを活用した技術開発を進めていきたい。

**謝辞**：本研究の一部は 2018 年度～2022 年度にかけて実施された「戦略的イノベーション創造プログラム (SIP) 第2期」の課題「国家レジリエンス (防災・減災) の強化」に位置付けられた「II. 被災状況解析・共有システム開発」において実施されました。

## 参考文献

- 1) 国際災害チャータホームページ, <https://disasterscharter.org/ja/web/guest/home> (2023-05-17 閲覧)
- 2) センチネルアジアホームページ, <https://sentinel-asia.org/> (2023-05-17 閲覧)
- 3) 田口 仁, 酒井 直樹, 白田 裕一郎: 衛星リモートセンシングデータ等の災害情報共有・流通のためのプラットフォーム開発, 写真測量とリモートセンシング, 小特集「空間情報データプラットフォーム (その3)」, Vol.59, No.1, pp.10-13, 2020.
- 4) 酒井 直樹, 田口 仁, 六川 修一: 災害時にも使えるフェーズフリーな衛星データの利活用, **AI・データサイエンス論文集**, Vol.4, No.L1, pp.9-18, 2023.
- 5) 内藤 昌平, 土屋 美恵, 水澤 弘充, 田口 仁, 藤原 広行: 熊本地震における WorldView-3 画像を用いた U-Net による建物被害自動判別手法の開発, **AI・データサイエンス論文集**, Vol.3, No.J2, pp.255-267, 2022.
- 6) He, K., Gkioxari, G., Dollar, P. and Girshick, R.: Mask R-CNN, *Proc. of the IEEE International Conference on Computer Vision (ICCV)*, pp.2961-2969, 2017.
- 7) Ghosh, S., Das, N., Das, I. and Maulik, U.: Understanding Deep Learning Techniques for Image Segmentation, 2019.
- 8) Bergstra, J. S., Bardenet, R., Bengio, Y. and Kégl, B.: Algorithms for Hyper-Parameter Optimization, *Proceedings of the 24<sup>th</sup> International Conference on Neural Information Processing Systems (NIPS11)*, December 2011, pp.2546-2554, 2011.
- 9) <https://qiita.com/tetsuro731/items/96df6b923c616080021c> (2023-06-02 閲覧)
- 10) <https://github.com/masc-it/Mask-RCNN> (2023-06-02 閲覧)
- 11) [https://github.com/fchollet/deep-learning-models/releases/download/v0.2/resnet50\\_weights\\_th\\_dim\\_ordering\\_th\\_kernels\\_notop.h5](https://github.com/fchollet/deep-learning-models/releases/download/v0.2/resnet50_weights_th_dim_ordering_th_kernels_notop.h5) (2023-06-02 閲覧)

(Received June 30, 2023)

(Accepted August 31, 2023)

## Automatic Building Detection and Damage Classification Model Using Mask R-CNN with Multiple Types of High-resolution Satellite Images

Shohei NAITO, Misato TSUCHIYA, Hiromitsu TOMOZAWA and Hitoshi TAGUCHI

In this study, we developed a model that aims to generate information to support disaster response by quickly identifying damage caused by natural disasters such as earthquakes and typhoons. This model uses high-resolution optical satellite images to automatically extract buildings in the image using Mask R-CNN, an instance segmentation method based on deep learning, and automatically classifies building damage into three levels: undamaged, damaged, and destroyed, and two levels: with or without blue sheet coverage. As a result, the accuracy of building extraction (IoU) was about 35%, and the accuracy of damage classification (Fmeasure) for each building was about 52%, which was slightly lower than that of the semantic segmentation model of U-Net. However, it was confirmed that the model has a certain level of performance as a model that can simultaneously perform building extraction and damage classification. The building detection and damage classification model was constructed using three types of high-resolution satellite images: WorldView-3, Pleiades, and GeoEye-1. The accuracy of building extraction was about 39%, and the accuracy of damage classification was about 92% for no damage, 69% for damage, 56% for collapse, and 85% for covered buildings, indicating that the model has a certain degree of generalization performance and can be used for early damage assessment.