EasyChair Preprint

№ 8610

Building Detection from Ortho Images Using Digital Elevation Maps

Shunsuke Konagai, Naoto Abe, Masakatsu Aoki and Jun Shimamura

August 6, 2022

標高データを利用した空中写真からの建物抽出
Building Detection from Ortho Images Using Digital Elevation Maps

小長井 俊介　阿部 直人　青木 政勝　島村 潤

Shunsuke KONAGAI　Naoto ABE　Masakatsu AOKI　and Jun SHIMAMURA

日本電信電話株式会社 人間情報研究所
NTT Human Information Laboratories

1. はじめに

畳み込みニューラルネットワーク（以下 CNN）の地理情報分野への応用として，衛星写真や空中写真を人間がトレースして地図を作成する作業の一部を CNN により代替する試みが行われており[1], 国際的な標準データセットの公開や地物検出精度コンテスト開催[2]なども進んでいる．筆者らは電子地図の更新作業の一環である空中写真から建物を抽出するタスクにおける都市部に特徴的な課題への対応を目的として，連続する垂直空中写真から画像幾何学処理によりオルソ画像（合成正射画像）を合成する過程で生成されるデジタル標高データとオルソ画像とを併せて CNN の入力とする建物抽出手法を考案し実験を行ったので結果を報告する．

2. 都市建物抽出の課題

デジタル地図データの更新に関して，一般に都市部は建造物の更新頻度が高いため，地図データも高い更新頻度が求められる。このため人手によるトレース作業を機械化する要求も高くなっている．

CNN による建物抽出を行う技術は大きくセマンティックセグメンテーションとインスタンスセグメンテーションの二種類が存在する．セマンティックセグメンテーションは処理対象の画像を画素単位でクラス分類するのに対し、インスタンスセグメンテーションは画像中のオブジェクトの抽出とそのオブジェクトのクラス分類とを併せて行うという違いがある．地図データの更新を目的とするタスクにおいては，建物等の地物をオブジェクトとして弁別できる点でインスタンスセグメンテーションがより好ましい．しかしながら一般に郊外エリアにおいては，建物は周囲を画像特徴が大きく異なる植栽や道路等に囲まれているためオブジェクト抽出の難易度は低い一方，都市部では建物が密集していることによりオブジェクト抽出の難易度が高くなるという問題がある．

また深層学習により高精度な CNN を訓練するためには多量の教師データが必要となる．この教師データの作成は空中写真のトレースによる地図の作成と同様に人手によって行われるため高コストである．近年発展している半教師あり学習の枠組みを適用して教師データ作成コスト低減を図ることが考えられ，今回の筆者らのタスクに直接適用はできないがインスタンスセグメンテーションに対する半教師あり学習手法も提案されている．[3][4]

3. 提案手法

3.1 概要

筆者らが開発したセマンティックセグメンテーション向けの半教師あり学習フレームワーク「教師ラベル補正技術」[5]により画素単位のマルチクラス分類 CNN の訓練を行い、CNN によって空中写真の画素分類を行う．電子地図用途に必要となる建物の弁別は，分類クラスの一つとして建物境界を示すクラスを導入し，分類結果の建物クラス画素の連続領域を個々の建物とするという後処理を追加することで実現した．

3.2 教師ラベル補正技術（半教師あり学習）

今回利用した半教師あり学習フレームワーク「教師ラベル補正技術」はクラス毎の正例・負例・不明の各画素から抽出する特徴ベクトルが特徴空間上で離れるよう Triplet Loss [6] に基づく目的関数で特徴抽出モデルを訓練し，特徴空間で十分に正例・負例に近い不明画素にそれぞれ正例・負例の疑似ラベルを付与することで教師データを補正する．その補正した教師データを用いて特徴抽出 CNN の再訓練を行うことを繰り返すことで, 漸次 CNN の精度向上を図るものである．

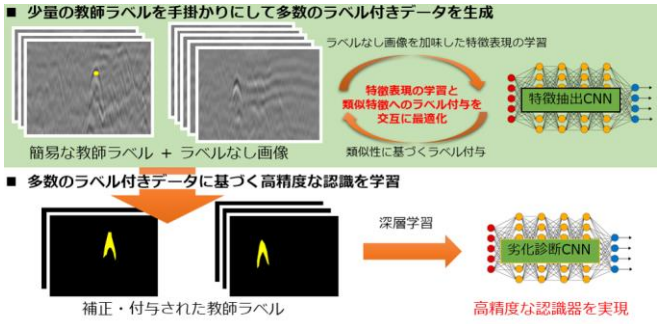

図 1.教師ラベル補正技術概要

3.3 建物境界クラス

特に都市部において密集する複数の建物がセマンティックセグメンテーションによっては分離できない問題への対処として，建物と建物との境界を分類クラスの一つとして採用した．CNN により出力される各クラスの推論結果マスクの論理演算を行うことで個々の建物の分離を図った．

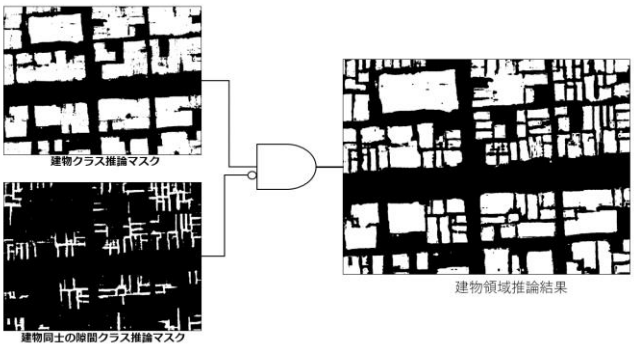

図 2. クラス推論マスク演算

3.4 デジタル標高データ

空中写真は画像の辺縁部に向けて建物の倒れこみや地形由来の歪みがおこる．これらを解消するために連続する複数の空中写真に重複して存在する標定点群や画像処理の場合は画像局所特徴量に基づく特徴点群の幾何計算によりデジタル標高モデルを作成する．[7] このデジタル標高モデルを利用し，空中写真上の像の位置を補正する正射変換を行ったオルソ画像が地図作成のためのトレース対象となる。

この過程で作成されるデジタル標高モデルは原理的に空中写真に撮影されている建物の高さ情報を含むため，都市部で建物が密集している場合であっても，隣接する建物の高さが異なる場合には個々の弁別に有効に働くことが期待できる．

このデジタル標高モデルを利用した建物抽出には、２つの方法が考えられる。

標定点群や画像特徴点群から幾何計算によって生成される 3D 点群に対して Pointnet[8]に代表される 3 次元物体検出を適用する方法と、RGB-D データに変換してから適用する方法[9]である。前者は、高精度な検出のためには高密度な 3D 点群が必要となり、そのようなモデル生成は計算量が膨大となる。後者は、透視投影カメラモデルを前提としており、今回の処理対象のようなオルソ画像とデジタル標高モデルの組合せにはそのまま適用できない。そこで我々は、オルソ画像とデジタル標高モデルの組合せから、既存の 2 次元画像処理 CNN を使って 3 次元物体検出を検出する手法を開発した。本手法では、まずオルソ画像を RGB 色空間から HSV 色空間に変換する。次に、建物の弁別への寄与度が低いと予想される S チャネルの彩度の代わりに、デジタル標高モデルの標高値を入れる。この V：明度，E：標高，H：色相の 3 チャネル画像を既存の 2 次元画像処理 CNN の入力とすることで、RGB3 チャネル画像を入力とする既存の各種 CNN をそのまま利用することができる．

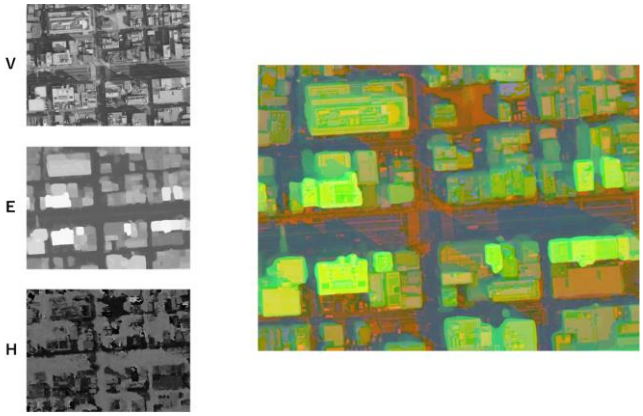

図 3. VEH データ例

3.5 後処理

今回の実験の目的は地図更新作業の一部機械化であり，出力データはそのまま，またはオペレータによる簡易な修正作業のみで電子地図として通用するものとしたい．しかしオルソ画像の解像度が十分に高くない場合にはセマンティックセグメンテーションでの領域境界がいびつになりがちで，建物輪郭情報を表現するベクトルデータのデータサイズが大きくなるだけでなく，オペレータによる頂点編集が困難になるという問題がある．このため地図データとして許容できる誤差の範囲で建物の外形を単純化することが望ましい．今回はセマンティックセグメンテーションで抽出した建物領域の外接矩形を出発点としたベクトル図形処理によって抽出領域の外形単純化を図り，単純化後の建物領域と正解ラベルとの IOU を算出・評価した．

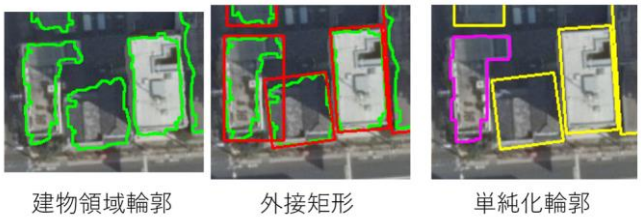

図 4. 輪郭単純化処理例

4. 実験

4.1 データセット

実験対象は国土基本図，地図情報レベル 2500 の図郭コード 09LD191：東京都蔵前橋周辺の東西 2km×南北 1.5km とし，この範囲を連続撮影した空中写真から，市販の写真測量ソフトウェアによりオルソ画像を作成した．これを 9×9 で分割し，画像サイズの 1333×1000

画素の画像 81 枚を生成した．この内 74 枚に対しておおよそ 50%の画素に正解ラベルを付与して訓練データとし，残り 7 枚に対しては全画素に正解ラベルを付与して精度評価データとした．

分類するクラスは「建物」「建物境界」「道路」「その他（水域，駐車場，植生域，etc.)」の 4 クラスとした．デジタル標高データは上記区画のオルソ画像作成過程で浮動小数点数値として生成されたものを 8bit 整数に正規化したうえでヒストグラム平滑化処理を行って画像の 1 チャンネルとした．

4.2 実験設定

事前に RGB オルソ画像に対して複数のセマンティックセグメンテーションモデルを適用し，そこで建物クラスの mIOU の精度が高かった DeeLab v3+ [10] によって，RGB 3 チャネルのオルソ画像を入力した場合と E:デジタル標高データを併用した VEH の 3 チャネルを入力した場合とで，それぞれ教師ラベル補正技術による再訓練を行い，精度評価用オルソ画像のクラス分類および建物領域の外形単純化後処理後の mIOU を算出した．教師ラベル補正技術による再訓練は訓練データのクロスバリデーションによる精度変化が規定条件を達成するまで行った．

4.3 実験結果

通常の RGB オルソ画像およびデジタル標高データを併せた VHE オルソ画像での処理結果 mIOU を表 1 に示す．

「再訓練数」は教師ラベル補正技術によって疑似ラベル付与によって補正した教師データを用いた CNN トレーニングの繰り返し回数である．

「クラス演算」列には

建物クラス推論結果マスク∧￢建物境界クラス推論マスク∧￢道路クラス推論マスク

と建物 GT との mIOU を，「輪郭単純化」列にはクラス演算結果の推論結果マスクに対して 2.3 に記述の後処理を適用した結果と建物 GT との mIOU をそれぞれ掲載する．

表 1. 実験結果 mIOU

データ	再訓練数	クラス演算	輪郭単純化
RGB	再訓練無	0.5661	0.6421
	30	0.7500	0.7467
	71	0.6817	0.7237
VEH	再訓練無	0.5696	0.6619
	30	0.8119	0.7996
	63	0.8196	**0.8460**

デジタル標高データを利用した VEH データでは，通常の RGB データに対して明確に mIOU の向上が確認できた．また RGB，VEH ともに教師ラベル補正技術の再訓練による精度向上が確認できた．ただし RGB の場合，再訓練回数 30 回から 71 回で mIOU が低下しており過学習が疑われる．以下の左列にクラス演算による推論マスクを，右列に輪郭単純化後のマスクをそれぞれ画像として可視化した例を図 5 に示す．

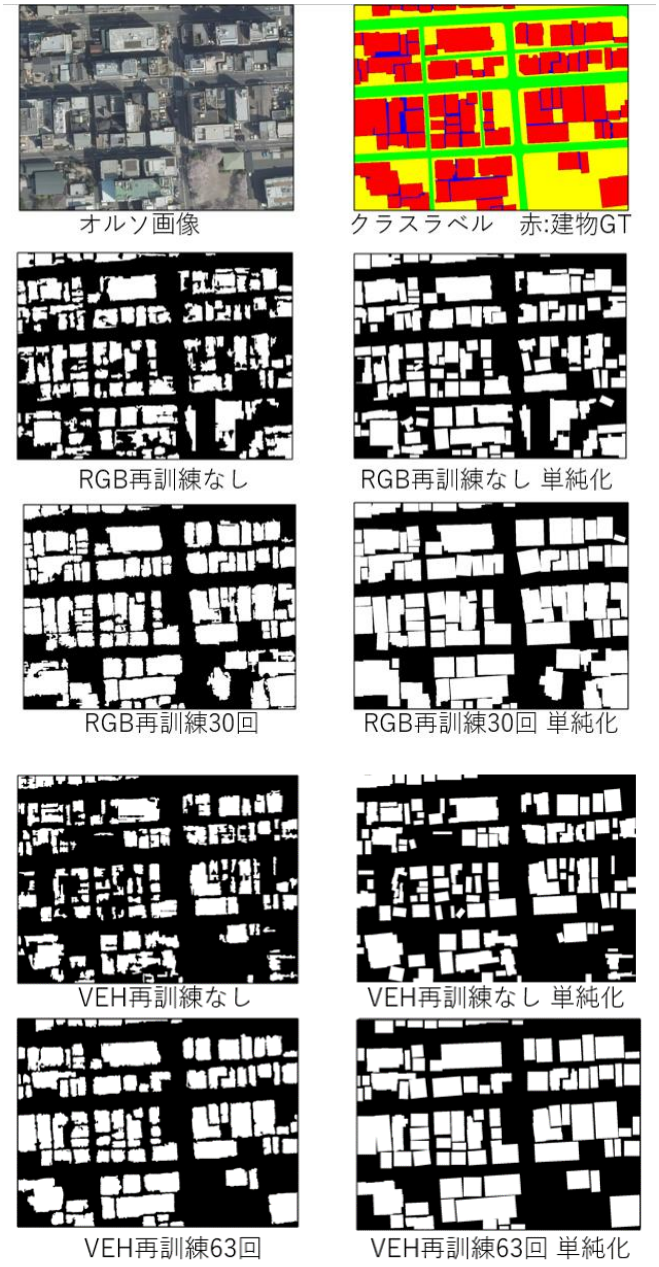

図 5. 実験結果例可視化画像

一部の試行において輪郭単純化により mIOU が向上しているケースが存在した．都市部では矩形や辺間の平行性を保つフットプリントを持つビルの割合が多いという特性が，今回採用した領域の外接矩形に基づく輪郭単純化処理と合致して mIOU 向上につながっている可能性があるが，より多くのデータでの検証が必要である．

VEH 再訓練 63 回の評価画像の 1 枚についてオルソ画

像に単純化後の輪郭を重畳表示した例を図 6 に示す. 建物と推論された領域の面積とその領域の外接矩形の面積とが十分に近いものは外接矩形を単純化輪郭として採用し黄色で表示した.

上記面積の差が大きいものは外接矩形の各辺のベクトルを保存しながら細分化して領域の内側に向けて移動させるという単純化を行った. これらの輪郭は紫で表示した.

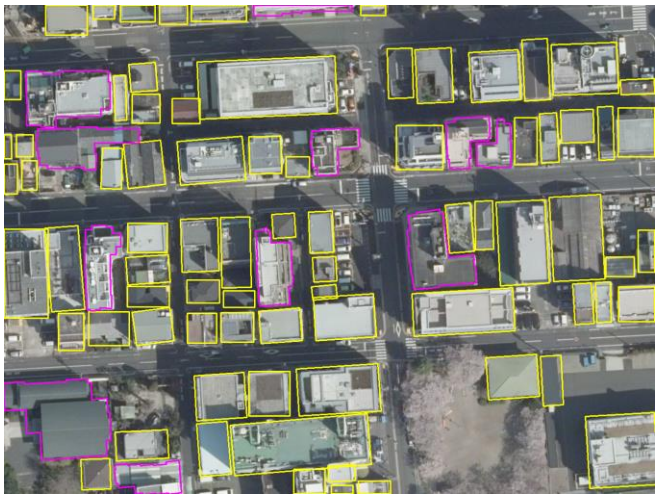

図 6. 単純化輪郭重畳例

一見して各輪郭の傾きを微調整すればそのまま地図として利用して違和感の無い抽出結果を実現できている.

5. まとめ

空中写真からオルソ画像を作成し, それに基づく地図更新を行うというワークフローにおいて, 副産物として生成されるデジタル標高データを利用することで CNN による建物抽出において mIOU による評価で明確な精度向上を確認した. また処理結果を地図としてそのまま利用可能とすることを目的とした輪郭単純化後処理を組み合わせることで定性的には有望な結果を確認できた. 一方で密集した狭小住宅の分離が不十分である等の傾向がみられることを含めて, IOU と地図としての品質に主観的齟齬が生じることが観測された. 地図として利用可能性の判断に資する定量的評価指標の確立が今後の課題と考える.

今回半教師あり学習のフレームワークを建物抽出に適用し, 実験の範囲での有効性を確認したが、実験データに含む未ラベルデータ量が少ない試行であったため, 教師データ量および未ラベルトレーニングデータ量の変動による精度変化の検証もフューチャーワークとしたい.

文　献

[1] Iglovikov, V., Mushinskiy, S., & Osin, V. (2017). Satellite imagery feature detection using deep convolutional neural network: A kaggle competition. arXiv preprint arXiv:1706.06169.

[2] I. Demir, K. Koperski, D. Lindenbaum, G. Pang, J. Huang, S. Basu, F. Hughes, D. Tuia, and R. Raskar. Deepglobe 2018:A challenge to parse the earth through satellite images. arXiv preprint arXiv:1805.06561, 2018.

[3] Khoreva, A., Benenson, R., Hosang, J., Hein, M., & Schiele, B. (2017). Simple does it: Weakly supervised instance and semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 876-885).

[4] Cheplygina, V., de Bruijne, M., & Pluim, J. P. (2019). Not-so-supervised: a survey of semi-supervised, multi-instance, and transfer learning in medical image analysis. Medical image analysis, 54, 280-296.

[5] Murasaki, K., Ando, S., & Shimamura, J. (2022). Semi-Supervised Representation Learning via Triplet Loss Based on Explicit Class Ratio of Unlabeled Data. IEICE Transactions on Information and Systems, 105(4), 778-784.

[6] Hoffer, E., & Ailon, N. (2015, October). Deep metric learning using triplet network. In International workshop on similarity-based pattern recognition (pp. 84-92). Springer, Cham.

[7] 国土地理院 "オルソ画像について" 国土地理院ホームページ https://www.gsi.go.jp/gazochosa/gazochosa40002.html（参照 2022-08-04）

[8] Qi, C. R., Su, H., Mo, K., & Guibas, L. J. (2017). Pointnet: Deep learning on point sets for 3d classification and segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 652-660).

[9] Gupta, S., Girshick, R., Arbeláez, P., & Malik, J. (2014, September). Learning rich features from RGB-D images for object detection and segmentation. In European conference on computer vision (pp. 345-360). Springer, Cham.

[10] Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European conference on computer vision (ECCV) (pp. 801-818).