

AI を活用した地物自動抽出に関する研究（第 1 年次）

実施期間 平成 30 年度～平成 34 年度
地理地殻活動研究センター
地理情報解析研究室 大野 裕幸 中埜 貴元
遠藤 涼

1. はじめに

地図作成は、労働集約的な産業である。近年では地図の高精度化や高頻度な更新が強く求められているが、従来の地図作成工程のままでは対応が困難になりつつある。一方、画像認識の分野では深層学習技術によりブレークスルーが生じ、いわゆる AI を用いた画像認識技術が次々と実用化されている。そこで、本研究は、地図作成工程にこれらの画像認識技術を導入することで労働生産性を高め、効率的な地図作成につなげることを目的とし、まずは従来の「判読」「図化」の工程を研究対象として 5 年計画で研究に着手した。本研究の長期的な目標は地図の自動作成の実現である。

2. 研究内容

2.1 研究開発の概要

研究内容は、①深層学習に用いる教師データの作成、②深層学習を実行する畳み込みニューラルネットワーク（CNN）の構成の検討とハイパーパラメータの探索、③自動抽出した地物を自動的に地図データベースに反映させるアルゴリズムの検討、に大別される。本年度取り組んだ内容は主に①と、②の試行である。

研究の対象とする画像は、測量成果である空中写真及びそれから作成されたオルソ画像とし、衛星画像は平成 32 年度打ち上げ予定の先進光学衛星の画像を対象としている。また、精度等の目標とする地図情報レベルは、空中写真を原典とする場合は 2500、先進光学衛星を原典とする場合は 25000 とする。自動抽出の対象とする地物は少なくとも 40 項目、抽出率は全体で 80%以上を目指すという数値目標を掲げている。なお、空中写真は、デジタル航空カメラにより撮影されたものを対象とする。

2.2 深層学習に用いる教師データの作成

深層学習は、機械学習の範疇に含まれるため、機械学習に用いる教師データと同じく、学習用のデータと検証用のデータによって構成される。両者の仕様は基本的に同じものであるが、検証用のデータの画像サイズは学習用データより大きくても良い。

2.2.1 基盤地図情報等のデータ及び地理院地図オルソ画像を用いた教師データ作成の検討

まず、既存の基盤地図情報及び電子国土基本図のデータと、地理院地図のオルソ画像を用いた教師データ作成に関して検討したが、データとオルソ画像の間の作成（撮影）年次の不一致や位置ズレが激しく、教師データの原典として用いるのは適当でないとの結論を得た。また、地理院地図のオルソ画像は ZL18 のデータが最大であるが、ZL18 では地上画素寸法が約 50cm と、研究の対象としている測量成果としての空中写真の地上画素寸法 20cm 以下とは大きく異なっている。空中写真から ZL19（地上画素寸法約 25cm）又は ZL20（同約 12cm）のオルソ画像を作成しなおすことも検討したが、2007 年以降にデジタル航空カメラで撮影された地区は少なくとも 549 地区あり、これらすべてのオルソ画像を作成しなおすのは困難で、地理院地図を含むオルソ画像を原典データとして用いることも適当でないとの結論づけた。

2.2.2 空中写真を用いた教師データ作成の検討

次に、オルソ画像に代わる原典画像として空中写真を用いる検討を行った。空中写真を原典画像とする場合、問題となるのは参照データとして既存の地図データを使用できない点である。したがって、参照データは判読しながら手作業で作成しなければならず、作成コストが高い。作成コストが許容できるかを検証するため、道路の教師データを数百枚試作してみたところ、200枚/人日以上の作成が可能であること、判読をしたうえで参照データを作成するため、既存の地図データとオルソ画像で作成した場合にある程度含まれる不整合をほぼ完全に排除できること、副次的な効果として教師データ作成者の判読スキルの向上が図れること、などから、空中写真を原典とし、判読によって教師データを作成する手法を採用した。さらに、検証データの作成に当たっては、より判読の確実性を確保するため、立体視による判読を実施することとした。立体視による判読が可能である点も、空中写真を原典とするメリットである。

2.2.3 教師データの仕様

2.2.3.1 学習用データの仕様

学習用データの仕様は、当初 256×256 の画像サイズで作成を開始したが、並行して実施した CNN の検討結果等から 2 度の仕様変更を行い、画像サイズ 286×286、RGB 形式で作成し、抽出対象の全項目を定義した 37 カテゴリ分類のもの及び主要な個別項目を対象とした 2 カテゴリ～4 カテゴリ分類のものに分けて、それぞれデータ構築を実施した。表-1 に全項目 37 カテゴリ分類のカテゴリ定義を示す。個別項目のデータ仕様（塗分け色）は、全項目の指定色とは必ずしも同じではない。学習データの原典は、閲覧用に作成された JPEG 形式の空中写真を用いた。

表-1 全項目のカテゴリ分類定義

番号	色 (#RGB)	項目	番号	色 (#RGB)	内容	番号	色 (#RGB)	内容
2	#7F7F7F	舗装道路	15	#007F00	自然植生	44	#FFBFBF	防波堤
3	#BF7F3F	未舗装道路	16	#FFFF00	畑	54	#BFFFBF	芝地
4	#5F5F5F	駐車場	17	#00FFFF	水田	58	#FFBF00	砂礫地
5	#FF007F	分離帯	18	#007F7F	収穫後の水田	61	#BFBFFF	たたき
6	#3F0000	軌道	21	#7FFF7F	茶畑	62	#7F3F3F	岩がけ
7	#0000FF	水部	22	#FF7F00	果樹園	63	#7F00FF	高塔
8	#FF0000	堅牢建物	23	#7F0000	荒れ地	65	#00007F	湿地
9	#FF00FF	普通建物	24	#3F3F3F	空地	66	#BF00FF	材料置場
10	#FF7F7F	無壁建物	27	#BFBF00	太陽光発電設備	67	#7FBF7F	園庭
11	#7F7F00	被覆	28	#BF00BF	輸送管	68	#7F7FFF	墓地
12	#BFBFBF	土堤	36	#BF7FBF	雪覆い等	255	#FFFFFF	未分類
13	#007FFF	水制	38	#7FBFBF	プラットフォーム			
14	#7FBF00	歩道橋	39	#FF7FBF	タンク			

2.2.3.2 検証用データの仕様

検証用データの仕様は基本的に学習用データと同じとしたが、次の 3 点の仕様が異なる。①画像サイズは 572×572 又はそれ以上、②測量成果として保存されている TIF 形式の画像が原典、③正式な検証用データは、立体視による判読、確認が必須。また、検証用データ作成に用いた空中写真とそれに

隣接する空中写真は、学習用データから除外した。

2.3 深層学習の試行と CNN の検討

空中写真画像から各ピクセルのカテゴリ分類を行って地物が占める範囲を特定するという処理は、「領域分割 (Semantic Segmentation)」という分野に属する。この分野は、2010 年から ILSVRC という画像認識のコンテスト (Berg et al., 2010) が開催され、2012 年には深層学習が注目される契機となった画像全体をカテゴリ分類する「画像分類 (Classification)」と比較すると、分類の難易度が高く、PSP Net (Zhao et al., 2017), U-Net (Ronneberger et al., 2015) などが提案されているものの、画像分類ほどの分類精度を達成しているものは見られない。

本研究では、表現の異なる画像の相互変換に高い性能を示している pix2pix (Isola et al., 2017) 及び U-Net の 2 種類の CNN と、道路、建物、水部の 3 種類の個別項目の教師データを用いて CNN の性能評価を実施した。

3. 得られた成果

3.1 深層学習に用いる教師データ

図-1 に、作成した教師データの例を、表-2 に作成した教師データの枚数を示す。検証用データは、地域多様性を確保するため、2007 年以降にデジタル航空カメラで撮影された全 549 地区から機械的に 1 枚ずつ対象画像を抽出した (図-2)。個別項目「道路」の検証データは、さらにフィルムカメラにより 1967 年以降に撮影されたほぼ全ての地区から機械的に 1 枚ずつ対象画像を抽出して作成したものを含む。個別項目「道路」以外の検証データは、作成中である。



図-1 作成した教師データの例 (全項目)

表-2 作成した教師データの枚数

項目	学習用データ	検証用データ
00 全項目	450	134
02 道路	3300	1372
05 分離帯	300	134
06 軌道	410	134
07 水部	1100	994
09 建物	1900	1036
15 自然植生	2000	134
16 畑	150	134
17 水田	400	134

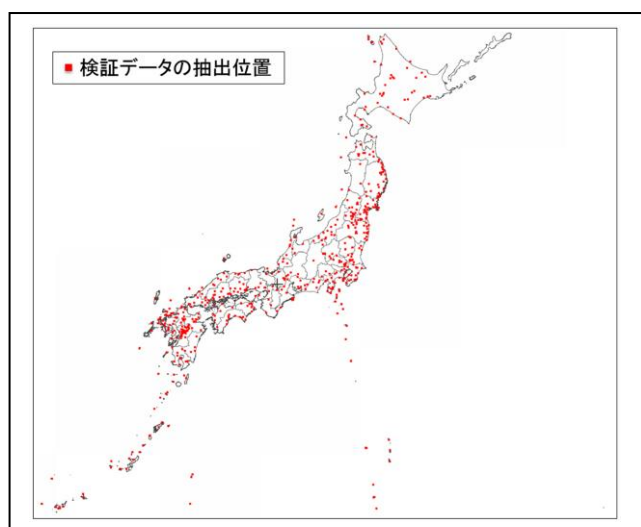


図-2 検証データの抽出位置 (デジタルカメラ 549 地区)

3.2 深層学習の試行と CNN の検討

深層学習には、nVIDIA CUDA 及び Tensorflow, Python を用いた。CNN は、pix2pix と、pix2pix の Generator 部分に用いられている U-Net 構造部分を取り出して独立させた CNN の 2 種類を用いた。

試行を行った地物は、いずれも個別項目の「道路」「建物」「水部」の 3 種類である。学習データは、90 度、180 度、270 度に回転を加えたものを追加し、4 倍の枚数としたものを用いた。使用した学習データの枚数、検証データの枚数、地物認識性能の評価値を表-3 に示す。評価値は、F-Measure (F 値) を用いた。学習データの枚数が多くなるにしたがって、pix2pix よりも U-Net の評価値が高まる傾向が見られている。これにより、教師データを作成し、それによる深層学習を実施した学習済モデルで推論し、認識精度を客観的な数値として把握するという一連の流れを確立することができた。

表-3 深層学習の試行結果

項目	学習枚数	検証枚数	pix2pix	U-Net
02 道路	12,400	549 (Digital Only)	71.3%	76.2%
07 水部	4,400	351 (Digital Only)	81.9%	73.7%
09 建物	6,800	439 (Digital Only)	74.0%	75.6%

4. 結論

本研究では、CNN を用いた深層学習の反復が研究の中核となる。深層学習の主要な構成要素である教師データセットに関する仕様を確定し、地域多様性を考慮した検証用データの作成にも着手し、道路に関しては完成を見た。また、深層学習のもう一つの主要な構成要素 CNN についても、異なる 2 種類の構成による性能比較を行い、評価値を用いた性能評価を実施できることを確認した。これらにより、研究の土台を構築することができたと言ったことができ、次年度以降、本格的な深層学習に着手して評価値の向上につながる最適なハイパーパラメータの探索につなげていく。

参考文献

- Berg, A., J. Deng and F. Li (2010): Large Scale Visual Recognition Challenge 2010, http://www.image-net.org/challenges/LSVRC/2010/pascal_ilsvrc.pdf (accessed 19 Mar. 2019).
- Zhao, H., J. Shi, X. Qi, X. Wang and J. Jia (2017): Pyramid Scene Parsing Network, arxiv, <https://arxiv.org/pdf/1612.01105.pdf> (accessed 19 Mar. 2019).
- Ronneberger, O., P. Fischer and T. Brox (2015): U-Net: Convolutional Networks for Biomedical Image Segmentation, arxiv, <https://arxiv.org/pdf/1505.04597.pdf> (accessed 19 Mar. 2019).
- Isola, P., J. Zhu, T. Zhou and A. A. Efros (2017): Image-to-Image Translation with Conditional Adversarial Networks, CVPR2017.