

AI を活用した地物自動抽出に関する研究（第 2 年次）

実施期間	平成 30 年度～令和 4 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	白石 喬久	
	遠藤 涼	中埜 貴元	
企画部国際課	岸本 紀子		

1. はじめに

本研究は、労働集約的な業務構造となっている地図作成に、近年画像認識の分野にブレイクスルーをもたらした畳み込みニューラルネットワーク（CNN）と深層学習の技術（以下総称して「AI」という。）を導入することで、長期的には地図作成の自動化を、短期的には地図作成における「判読」及び「図化」の工程の生産性向上を図ることを目的としている。5 年計画の初年度にあたる平成 30 年度に、過去に国土地理院がデジタル航空カメラによって撮影したすべての空中写真撮影地区を母集団とした評価用データセットを作成した。これにより、令和元年度から抽出性能の数値評価が可能となった。

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（トレーニングデータと検証用データ）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。令和元年度は②を中心に取り組み、その検討結果に基づき①を実施した。

空中写真や衛星画像から特定の地物を抽出しようとする AI の研究は、主に、既存のトレーニングデータの集合を対象とし、それに対する抽出性能が最も高くなる CNN の構成やハイパーパラメータを検討するという手法が用いられる。しかし、本研究では、対象となる空中写真が毎年度の撮影の実施により漸次増加すること、国土地理院が保有するすべての空中写真を母集団、いわゆるビッグデータとして使用し、トレーニングデータセットを自由に作成できる環境にあること、本研究の目的に合致した既存のトレーニングデータが存在しないことなどから、多くの研究のアプローチとは逆に、使用する CNN を固定し、抽出性能を高められるトレーニングデータの構成や内容を検討するアプローチを採っている。これは、一般的に機械学習では「良質な」トレーニングデータを大量に使用するのが好ましいとされている中であって、いかなる要件が「良質」であるかを明確にし、より少ないトレーニングデータでより高い性能を引き出せるようにするというアプローチとも言える。

一方、空中写真画像から地物を抽出する方法は、(a) 特定の種別（道路や建物等）の地物が写った画素範囲のみを抽出する方法と、(b) すべての画素を何等かの地物項目に分類する方法の 2 種類に大別される。これらは、いずれも画像認識におけるセマンティック・セグメンテーション問題として捉えることができるが、そのためのトレーニングデータの作成方法や評価方法はそれぞれ異なる。

既存のデータセットとしては、(a) をテーマとしたものは、オルソ化された衛星画像から道路、建物、土地被覆を分類する DeepGlobe 2018 (Demir et al., 2018) で用いられた SpaceNet Dataset や、オルソ化された空中写真から建物を抽出する性能を競う Inria Benchmark Dataset (Maggiori et al., 2017) な

どが知られている。(b)については、自動運転における物体認識等を行うことを目的として車両内から前方を撮影した映像の全画素を30クラスに分類した Cityscapes Dataset (Cordts et al., 2016) などが知られているが、衛星画像や空中写真を対象とした地物分類データセットの存在は確認できていない。

性能の評価は、(a)の場合はF値(F1 Scoreとも言う)で、(b)の場合はmIoUで実施する。

本研究では、トレーニングデータセットは当面の間、(a)、(b)の両方のアプローチで作成し、相当数のデータが揃った段階でどちらの手法がより高い地物抽出性能が得られるかを比較評価する予定であり、令和元年度は主に(a)のトレーニングデータの作成に比重を置いて実施した。

2.2 深層学習に用いるCNNの決定

セマンティック・セグメンテーション問題で用いられるCNNは様々な構成のものが提案されている。本研究では、平成30年度にまず Conditional GAN (Isola et al., 2017) を用い、次いでその発生器(Generator)に用いられているU-Net構造のネットワークのみを用いて両者の抽出性能を比較した結果、すべてのケースでU-Net構造のみの抽出性能が良いことを確認していた。DeepLabやResNetなど、U-Net構造よりもさらに高い抽出性能を謳うネットワークの使用も検討したが、トレーニングデータの仕様が異なることから、CNNはU-Net構造のネットワークに固定することとした。U-Net構造の層の数は、Encode側、Decode側とも5層~9層までのネットワークを用いて同じトレーニングデータセットを使ってトレーニングした結果から、最も抽出性能が良かったEncode側、Decode側とも8層、合計16層のネットワークとした。これは、Conditional GANの発生器と全く同じ構成のネットワークである。以後、本研究に用いたCNNは、すべてこのU-Net構造16層のネットワークである。

2.3 トレーニングデータが抽出結果に与える影響の調査

一般的に、機械学習においては、トレーニングデータの数は多いほど良いと言われる。一方、良質なトレーニングデータが重要とも言われる。しかし、空中写真を対象としたトレーニングデータを作成してきた中で、特に前者の論調に疑問を呈するようになった。そこで、道路を対象とし、2007年から2017年までの間に国土地理院がデジタル航空カメラで撮影したすべての空中写真画像を母集団(以下「母集団」という。)として、(1)無作為に抽出した画像群、(2)恣意的に選択した画像群、(3)(1)と(2)を合わせた画像群、の3種類のトレーニングデータセットを作成し、抽出性能を評価した。

2.4 トレーニングデータの作成

2.2及び2.3の評価結果から、一定の枚数を作成する毎にトレーニングデータの効果を確認しながらトレーニングデータの作成を継続した。

2.5 検証用データの作成

各項目のトレーニング済モデルの地物抽出性能の評価は、トレーニングデータとは全く別に母集団から機械的なルールにより平成30年度に作成した検証用データセット(以下「GSI検証データ」という)を用いる。しかし、GSI検証データには、「15自然植生」「07水部」「23荒地」の3項目はピクセル数が過大でPrecision(適合率)が適切に評価できない点、逆に「06軌道」「13水制」「27太陽光発電設備」などピクセル数が過少でRecall(再現率)が適切に評価できない項目が存在するといった課題がある。前者は対象地物以外のピクセル数を増加させること、後者は対象地物のピクセル数を増加させることで解消可能であるため、個々の項目毎に2018年以降に撮影された空中写真撮影地区からフルサイズの空中写真画像を検証用データに追加する作業を実施した。新たに追加された検証データの画像サイズは使用されるカメラが地区によって異なるため一定ではないが、例えば国土地理院が使用しているDMCII e230型航空カメラの場合、1枚の画像サイズは14144×15552pixelとなり、1枚あたりGSI検証データセット各画像(572×572pixel)の約672倍のピクセル数が確保できる。これは、GSI検証データ全体のピクセル数の約1.23倍に達するピクセル数であるため、前者の課題を解消する

ことができる。また、後者についてもある程度のピクセル数を確保することが期待できる。ただし、追加する検証用データは地区数が限定されかつ1枚あたりのピクセル数も多いため、GSI検証データが持つ地域多様性が阻害される恐れがあることから、追加するデータ数が一定程度の地区数に達するまでの間は、得られるF値の取扱いには注意が必要となる。なお、これに伴い、2018年以降に撮影された空中写真は、一切トレーニングデータ作成に使用しないこととした。

3. 得られた成果

3.1 トレーニングデータが抽出結果に与える影響の調査

道路（02 舗装道路，03 未舗装道路の2項目）を対象として，作成するトレーニングデータの元画像の選択方法が抽出性能にどのような影響を及ぼすかを調査した結果は次のとおりである。

まず，母集団から（1）無作為に抽出した画像1080枚，（2）恣意的に選択した画像1080枚，（3）（1）と（2）を合わせた2160枚の画像の3種類の画像群を用意した。（1），（2）は，それぞれ抽出されるべき道路の範囲をPhotoshop上で塗り分け，元画像と道路を塗り分けた範囲の画像のセットをトレーニングデータとした。（3）は（1），（2）で作成したものをそのまま使用した。各画像は572×572pixelのサイズで，これを286×286pixelのサイズに4分割した画像と，286×286pixelのサイズに縮小した画像の合計5枚の画像に加工し，さらに各画像を90度，180度，270度に回転させた画像を追加することで合計（1）21600枚，（2）21600枚，（3）43200枚のトレーニングデータが作成できた。

次に，この3種類をそれぞれ同じハイパーパラメータで深層学習させたトレーニング済モデルを用いて，道路の抽出性能を比較した。推論結果の性能評価にはF値を用い，評価用の画像はGSI検証データを用いた。評価結果を表-1に示す。

表-1 元画像抽出方法別の抽出性能評価値（F値）

種別・枚数	(1)無作為 21600枚	(2)恣意的 21600枚	(3)両方 43200枚
F値	0.70284	0.77194	0.77469

この結果から，F値は，（3）>（2）>（1）となっており，トレーニングデータの数が多（3）のF値が最高との結果が得られた。しかし，（2）恣意的21600枚と（3）合算43200枚のケースでF値の差は0.00275に過ぎず，（2）恣意的21600枚のケースに対し，無作為に21600枚のデータを追加してもその性能向上は0.00275に留ったと言い換えることもできる。同じ枚数の場合に画像選択方法（1）無作為と（2）恣意的の差が0.0691に及ぶことを考慮すれば，適切な画像の選択ができれば，トレーニングデータの枚数の半減が可能と考えられる。では，「恣意的」としてどのような選択方法が適切かに関しては，定量的な分析結果を得るに至っておらず，今後の課題として取り組んでいく。

3.2 トレーニングデータの作成状況

本研究は，「作業規定の準則」の標準図式に規定された地図情報レベル2500の地物のうち，少なくとも40項目についてF値0.8以上の抽出性能を達成することを数値目標として掲げている。そのための項目別トレーニングデータの作成数（令和2年3月31日時点）を表-2に示す。実際のトレーニングは，これに90度，180度，270度に回転させた画像を追加した4倍の数で実施することとなる。

3.3 検証用データの作成状況

空中写真画像そのものを対象とした検証用データは，「05 分離帯」用に1枚，「06 軌道」用に3枚，「07 水部」用に3枚，「13 水制」用に1枚の合計8枚を作成した（令和2年3月31日時点）。

表-2 項目別トレーニングデータ作成数 (286×286pixel)

No	項目	作成数	No	内容	作成数	No	内容	枚数
02	舗装道路	恣意的 6000	15	自然植生	200	44	防波堤	-
03	未舗装道路	無作為 5800	16	畑	1310	54	芝地	0
04	駐車場	400	17	水田	415	58	砂礫地	60
05	分離帯	350	18	収穫後の水田		61	たたき	0
06	軌道	500	21	茶畑	20	62	岩がけ	15
07	水部	1900	22	果樹園	20	63	高塔	20
08	堅牢建物	3100	23	荒地	0	65	湿地	0
09	普通建物		24	空地	0	66	材料置場	0
10	無壁建物		27	太陽光発電設備	120	67	園庭	0
11	被覆	20	28	輸送管	0	68	墓地	15
12	土堤	0	36	雪覆い等	76	他に広葉樹, 針葉樹, 竹林, はい松地, しの地, やし科樹林がある (これらは画像分類で対応)		
13	水制	60	38	プラットホーム	0			
14	歩道橋	0	39	タンク	0			

※本表のほかに No00 全種別混合 905 枚がある.

4. 結論

平成 30 年度に GSI 検証データが使用可能となったことで, 令和元年度から研究の本丸に着手している. 使用した U-Net 構造の CNN よりも抽出性能が高いとされるネットワークが存在するが, 敢えて相対的に低い性能の CNN で目標値の達成に挑戦することで, 相対的な性能がより高い CNN を用いた際にさらに高い性能に到達できる可能性がある. 周り道かもしれないが, 「良質」なトレーニングデータの要件を明らかにすることで, 総体的により少ないコストで高い性能を発揮する地物抽出方法の開発を目指し, さらに研究を進めていく.

参考文献

- Emmanuel Maggiori, Yuliya Tarabalka, Guillaume Charpiat, Pierre Alliez (2017): Can Semantic Labeling Methods Generalize to Any City? The Inria Aerial Image Labeling Benchmark, IEEE International Symposium on Geoscience and Remote Sensing (IGARSS).
- Ilke Demir, Krzysztof Koperski, David Lindenbaum, Guan Pang, Jing Huang, Saikat Basu, Forest Hughes, Devis Tuia, Ramesh Raskar (2018): DeepGlobe 2018: A Challenge to Parse the Earth through Satellite Images, arXiv:1805.06561v1.
- Introducing the SpaceNet Road Detection and Routing Challenge and Dataset, <https://medium.com/the-downlinq/introducing-the-spacenet-road-detection-and-routing-challenge-and-dataset-7604de39b779>, (accessed 06 Apr, 2020).
- Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth and Bernt Schiele (2016): The cityscapes dataset for semantic urban scene understanding, CVPR2016 (Oral and Poster).
- Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, Alexei A. Efros (2017): Image-to-Image Translation with Conditional Adversarial Networks, CVPR2017(Poster).