

AIを活用した 地物自動抽出に関する研究

(平成30年度～令和4年度)

地理情報解析研究室

大野 裕幸

1. 研究の位置づけと概要
2. 検証データの作成
3. CNNと目的関数の検討
4. 学習データの構成の検討とAIの構築
5. 目標達成状況の自己評価
6. 本研究の出口戦略
7. 現時点の課題と今後の研究内容

1. 研究の位置づけと概要

地図作成が直面する課題

地図作成工程の流れ

写真撮影

(現地調査)

標定

判読・図化

(編集)

地図情報

(ステレオモデル構築)



- ・基盤地図情報
- ・電子国土基本図 (地図情報)

【課題】

- ・労働集約型
- ・時間を要す

【背景】

コンピュータの能力向上
AI技術の実用化

研究目的

「判読・図化」の工程のブレークスルーを実現する地物抽出AIの構築

本研究におけるAIとは

- ・ つまるところ、取り組むのは土地被覆分類問題である
- ・ その手段にいわゆるA I による画像認識技術を用いる



【A I による土地被覆分類※】

画像に写っているピクセル1つ1つを、何が写っているかあらかじめ定義されたカテゴリに分類していく。



入力画像



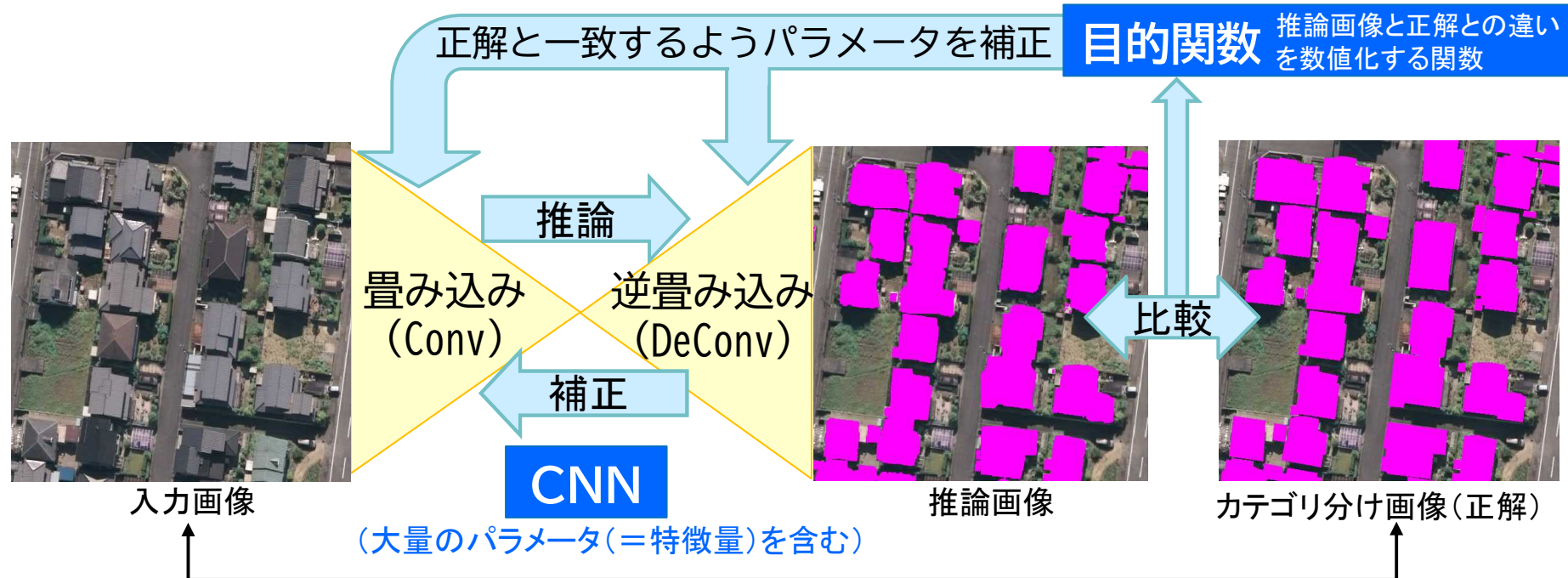
カテゴリ分け画像

入力画像と、カテゴリ分け画像のセット
(=学習データ)を使って、「**深層学習**」により
AIを学習させる。

※画像認識の分野で「セマンティックセグメンテーション」と言われる手法

本研究におけるAIとは(深層学習とは)

- ・ 深層学習には「学習データ」「目的関数」「CNN※」を用いる
- ・ 欲しいのは「入力画像をより正解に近く変換できる特徴量」



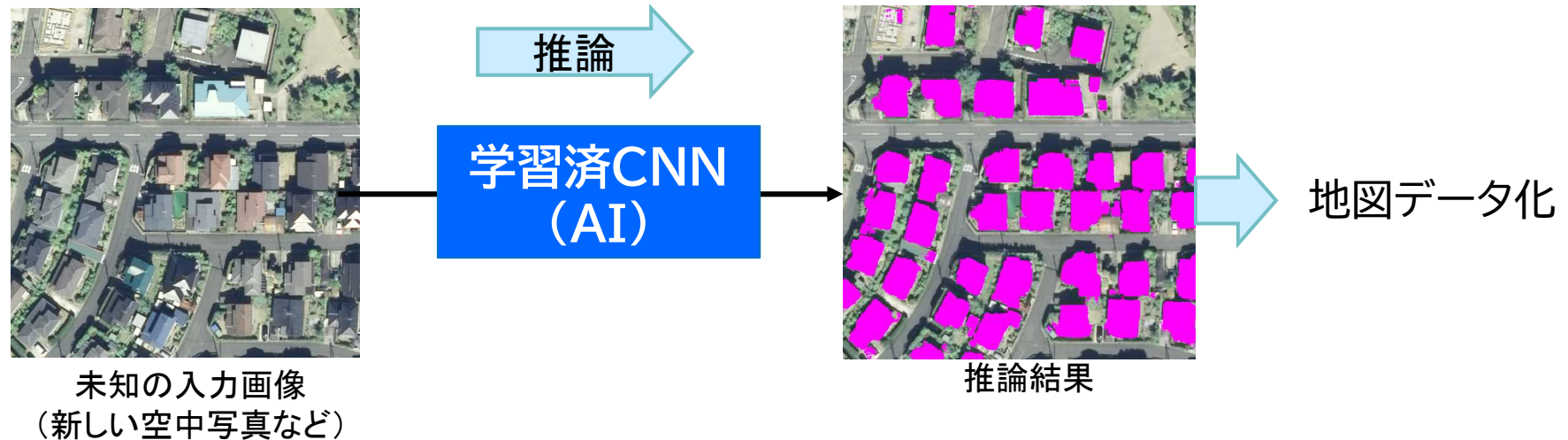
大量の学習データを使って目的関数の出力を0に収束させる (= 正解と一致させる) 方向に推論とCNNのパラメータの補正を繰り返す。パラメータの値は次第に様々な特徴量を反映した値となっていく。

深層学習

本資料では、「深層学習したCNN (学習済モデル)」のことを「AI」と呼ぶ。

※CNN: Convolutional Neural Network (畳み込みニューラルネットワーク)

- ・ 深層学習を終えたCNN (AI) には、様々な特徴量が習得されており、そのまま推論に使用できる。

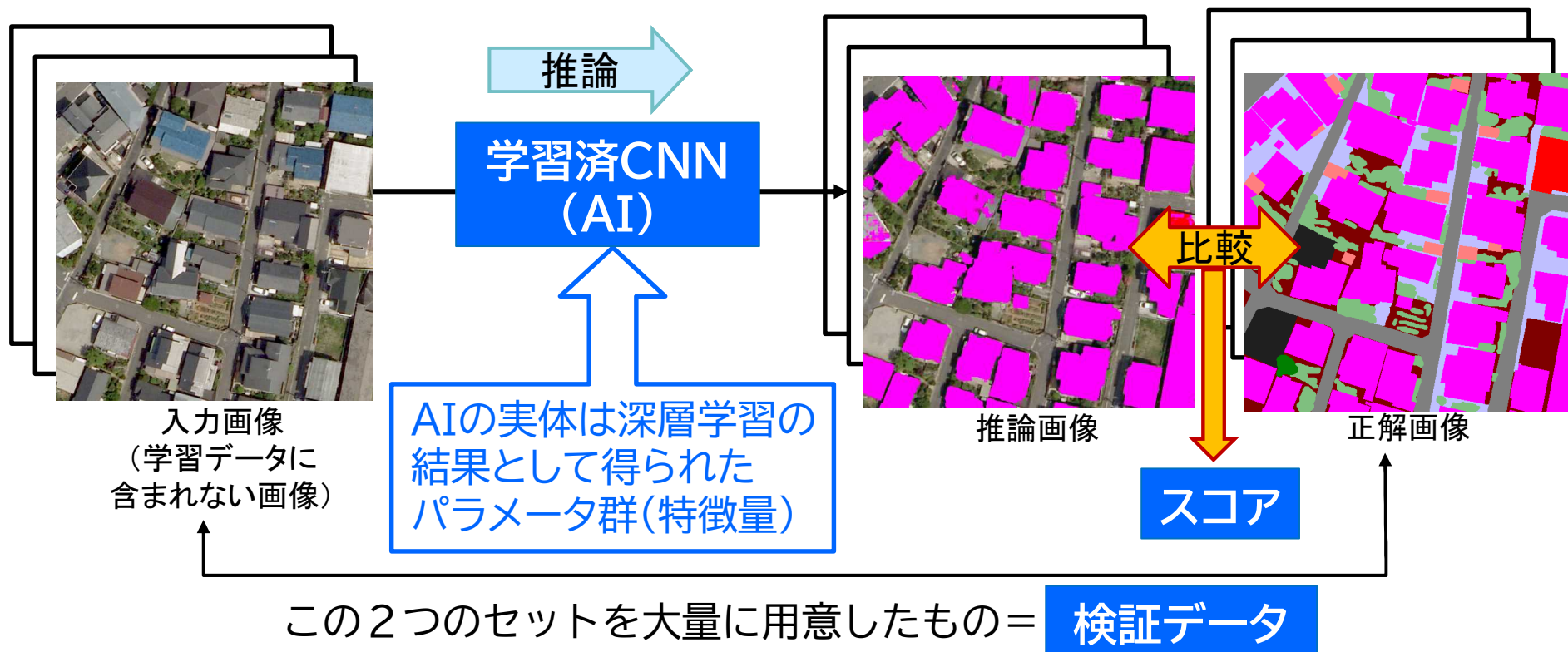


AIを使う際に、学習は原則不要※。
使用時は、推論のみが行われるため、処理速度が非常に高速になる。
かつ、深層学習したモデルの推論精度は、従来法より高くなることが多い。

※ファインチューニングという学習を短時間行って、個別データへの適合性をより高めたうえで推論を行う場合や、転移学習と言って、CNNのパラメータ(習得された特徴量)の一部を転用して別のCNNに変更する学習を行う場合もある。

AIの性能評価（スコアの算出）

- AIの性能は、検証用のデータの推論結果によって客観評価する
- 検証用のデータが途中で変わると性能が比較できないため、本研究では最初に「検証データ」を作成し、それを性能評価に用いる



検証データから推論した画像とその正解を比較※することでスコアが算出される。スコアの算出方法には、F値（F1-Score）、IoUなどがあるが、本研究では主にF値を用いて評価する。

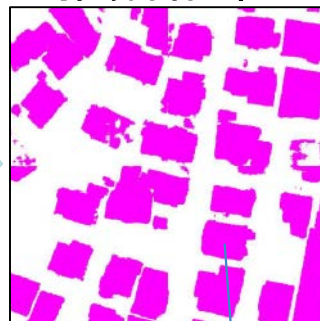
※検証時の比較方法は、学習時の目的関数とは関係ない。

カテゴリ分類の定量評価指標(F値)

検証データ入力画像



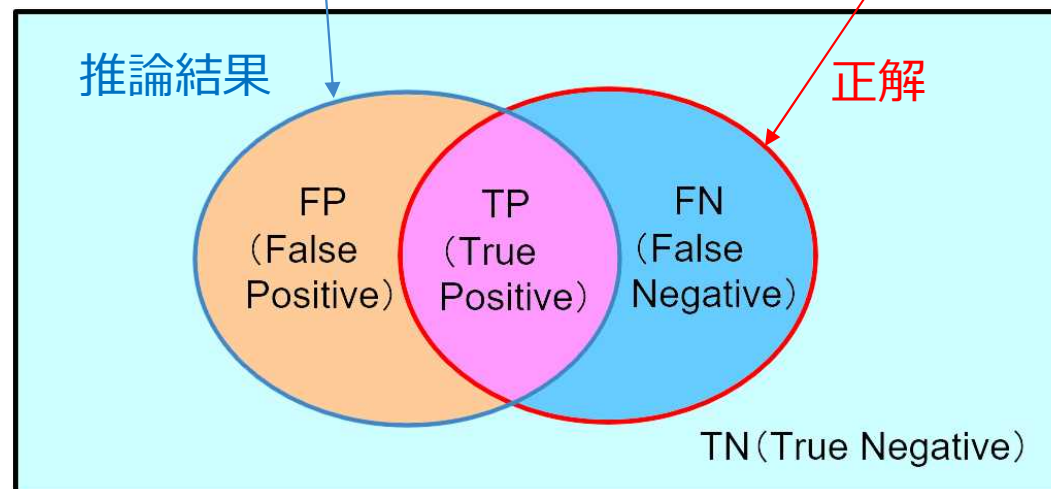
推論結果



検証データ(正解画像)



評価値
算出※



※評価値の算出方法

① 適合率(Precision) = $TP / (TP + FP)$

② 再現率(Recall) = $TP / (TP + FN)$

③ F値(スコア) = $\frac{2 \times \text{Recall} \times \text{Precision}}{(\text{Recall} + \text{Precision})}$

	正解が1	正解が0
推論結果が1	TP	FP
推論結果が0	FN	TN

研究の流れ(着手時の想定)

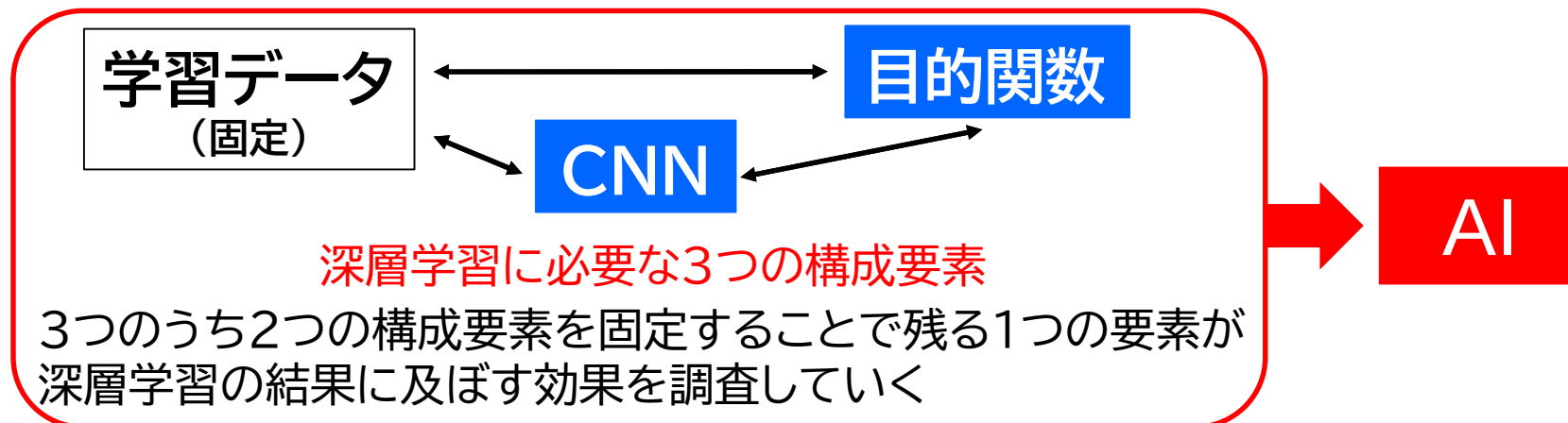
研究着手時に想定していた流れ

① 大量の **学習データ** と **検証データ** の作成

深層学習に使用する学習データを固定
 その中でスコアを向上できるCNN構成や目的関数を工夫



② 深層学習を反復してスコアを参照しながら ・最適な **CNN** 構成とより良い **目的関数** の検討 一方を固定し、もう一方の構成や内容を変更しながら深層学習を繰り返す



研究着手時の想定

地理座標上に各種データを集約することで大量の学習／検証データを研究初期に効率的に作成することを計画

地物教師データベース

オルソ画像



■ 歪みを取って
地図と重なるよう
処理された画像

既存の地図情報や地形情報



3次元点群

■ レーザスキャナに
よって取得された
点群データ

衛星画像

■ 先進光学衛星

自動的に作成

検証データ

自動的に作成

学習データ

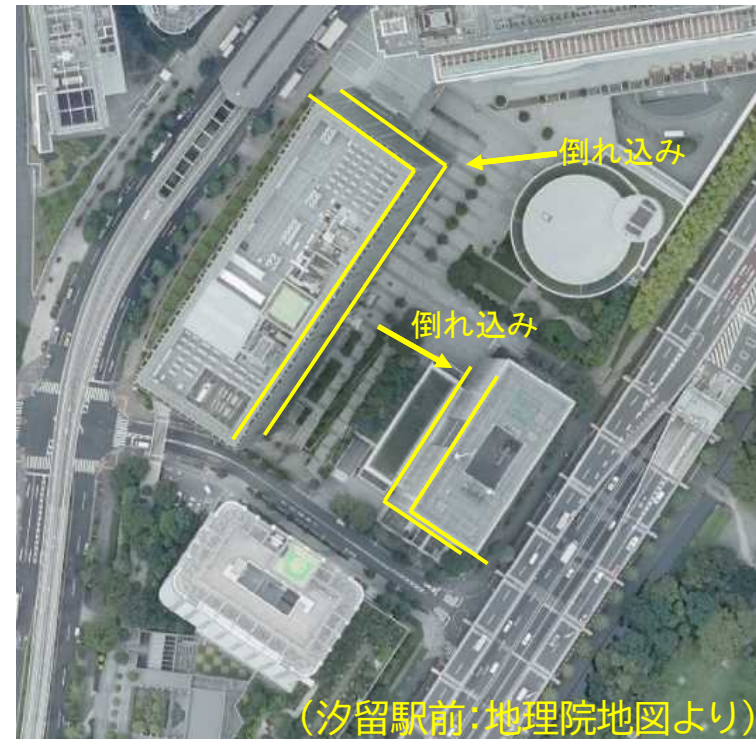
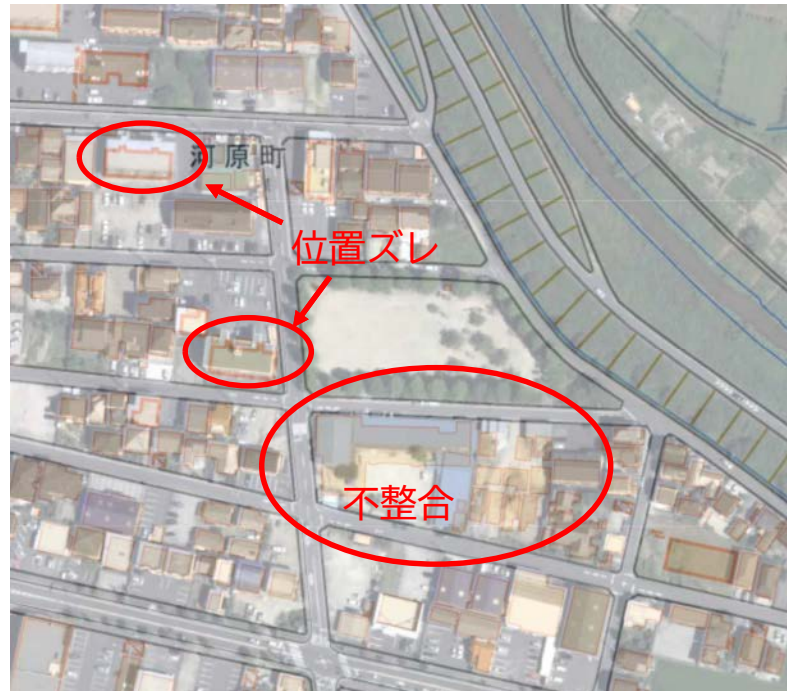
目的関数

CNN

AI

深層学習に必要な3つの構成要素

研究初期に直面した問題点



オルソ画像でも倒れ込みがみられる
（倒れ込みの陰に隠れた地物は抽出不能）

- ・ 既存の地図情報と空中写真／オルソ画像の内容の不整合
- ・ オルソ画像でも建物の倒れ込みが見られる
- ・ 本研究では許容できない量の位置ズレ

が判明。地物教師データベースを構築して大量の
学習／検証データを効率的に作成することが不可能に。



実際(現在)

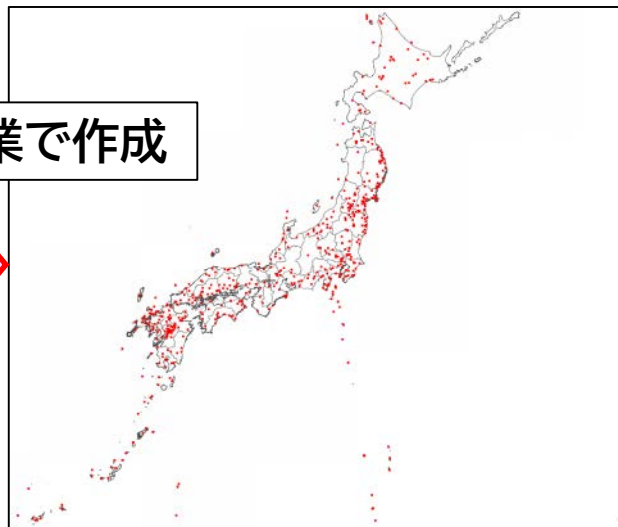
地理教師データベースの構築が頓挫したことにより、
学習データ、検証データとも手作業で作成する必要が生じた
➡しかし、これがいくつかの知見の発見につながることに。

空中写真



■オーバーラップがあり、同じ地物を多方向から撮影

手作業で作成



検証データ

AIに処理させたい母集団
(=日本全国)における
性能評価が可能なデータ
セットの作成

手作業で作成

学習データ

目的関数

CNN

AI

深層学習に必要な3つの構成要素

研究の流れ(確定後)

確定した研究の流れ

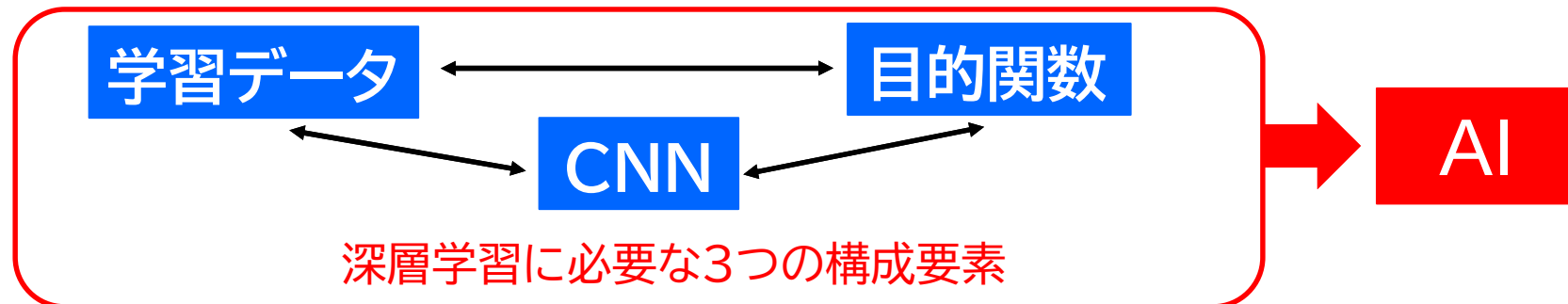
- ① 性能評価のスコア算出に必要な **検証データ** の作成
検証用のデータを固定することで、性能評価の基準に一貫性を持たせる



- ② 最適な **CNN** 構成とより良い **目的関数** の検討
大量の学習データが用意できなくなったことから、暫定的に検討用の学習データを作成し、深層学習を繰り返しながらCNNと目的関数を検討



- ③ 効率よくスコアを上げられる各項目の **学習データ** の作成
②の検討結果からCNNと目的関数を固定し、目標スコアをより少ない数の学習枚数で達成できる学習データの構成内容(画像選択方法)の検討



【目標（数値目標）の設定】

①抽出成功率の達成度判定の目標

40項目以上の地物に対し、基盤地図情報の精度の範囲内※1で
人が判読した場合に匹敵する抽出成功率の達成
(参考：自動抽出の対象となる地物は93項目※2)

②実用的なレベルへの到達度判定の指標

抽出対象となる地物数全体の**80%程度**の抽出

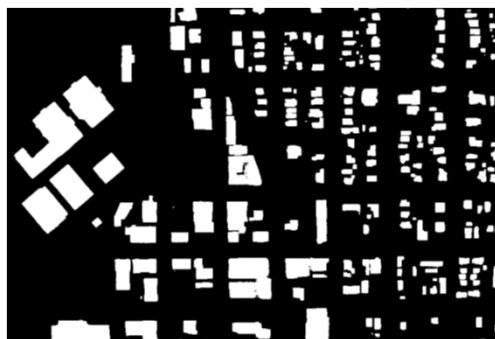
技術的なブレイクスルーとすることを標榜していることから、
非常に高い目標を設定。

※1 基盤地図情報項目以外は、電子国土基本図で必要とされる精度とする

※2 地図情報レベル2500の項目のうち、行政界や建物記号などの、本質的に画像やセンサーで観測できない項目と等高線の分類項目を除外した項目数

研究開発の目標(研究着手当時のスコア)

- Inria Aerial Image Labeling Dataset(対象は建物のみ)
最高スコア Bodhiswatta Chatterjeeによる 0.7873(2018年11月当時)



(参考)
2021-02-15時点の最高スコアは
Eugene Khvedchenya
による 0.8106 (2020-09-03)

出典: <https://project.inria.fr/aerialimagelabeling/leaderboard/>

- DeepGlobe 2018
CVPR2018の一環で開催された衛星画像を用いた地物抽出の国際コンペ



道路抽出 0.6577
(Alibaba iDST and AMAPチーム)



建物抽出 0.7391
(Hamaguchi 他, 2018)
(PASCO MLTチーム)



土地被覆分類 0.5384
(Alibaba iDST and AMAPチーム)

← 各部門の
優勝スコア

画像出典: like Demir他(2018)、A Challenge to Parse the Earth through Satellite Images

2. 検証データの作成

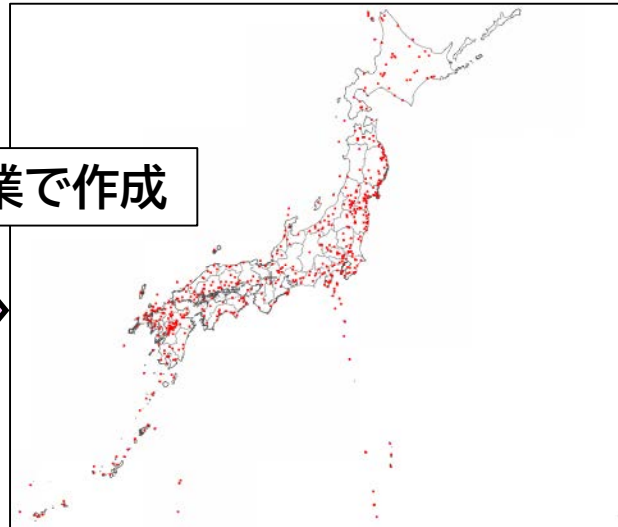
① 検証データの作成

空中写真



■オーバーラップがあり、同じ地物を多方向から撮影

手作業で作成



検証データ位置の地理的分布
(赤点が検証データの位置)

検証データ

- ・2007※～2017年度に測量用デジタルカメラで撮影された全549地区から機械的に抽出
- ・画像中央の572×572pixelの範囲を切り出して36カテゴリに分類

検証データは、AIの性能を数値評価する(スコアの算出)ために用いる。

- ・母集団の分布とできるだけ一致させる
- ・有意な評価値を算出できるだけのボリュームを有す必要がある。

学習データの一部を性能評価用の検証データとする研究が多いが、本研究では、検証データは、学習データとは全く別の独立したデータセットとして構築。

※測量用デジタルカメラは2007年から導入

① 検証データ

入力画像

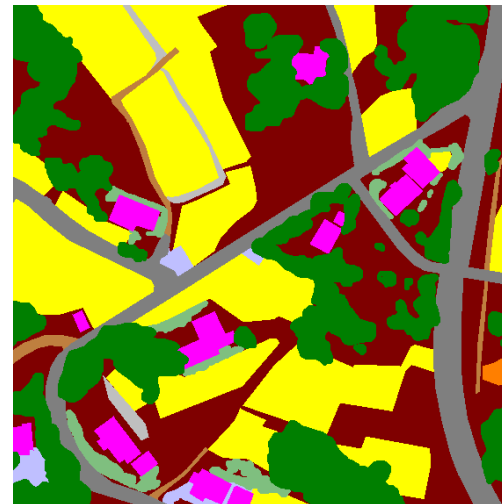


カテゴリ分け画像(正解)



【画像の諸元】

- ・サイズ 572×572pixel
- ・ラベル形式 RGB及びIndex
- ・画像形式 PNG
- ・作成方法 測量技術者の
立体視による
- ・分類基準 公共測量作業
規程の標準図式
に基づく
- ・地上画素寸法 20～40cm

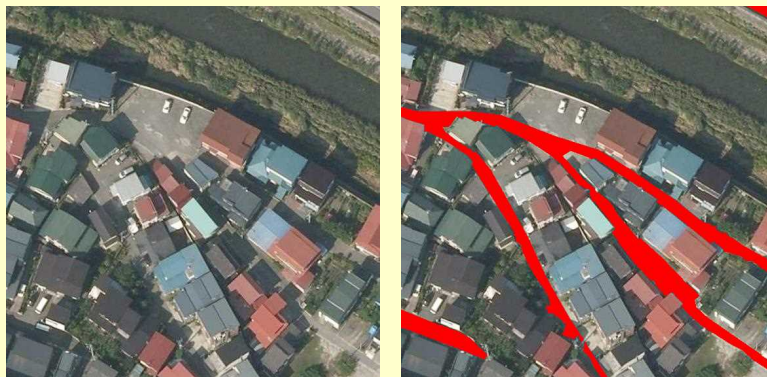
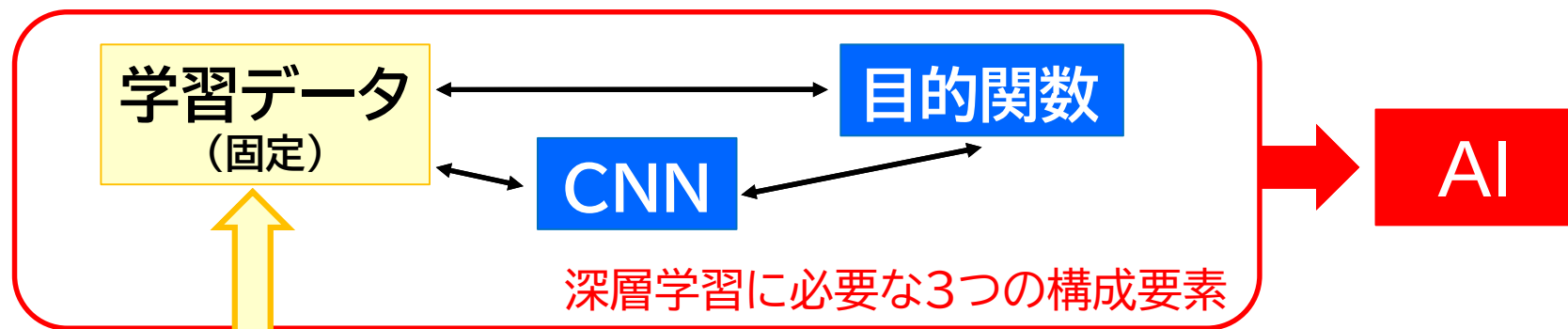


3. CNNと目的関数の検討

② 深層学習によるAIの開発

深層学習に必要な構成要素は「学習データ」「目的関数」「CNN」

- ・学習の効果を見るため、3つのうち2つの要素を固定して比較
- ・まず「道路」の学習データを5400枚作成して学習データを固定



無作為に選択した空中写真から
道路の学習データを5400枚作成

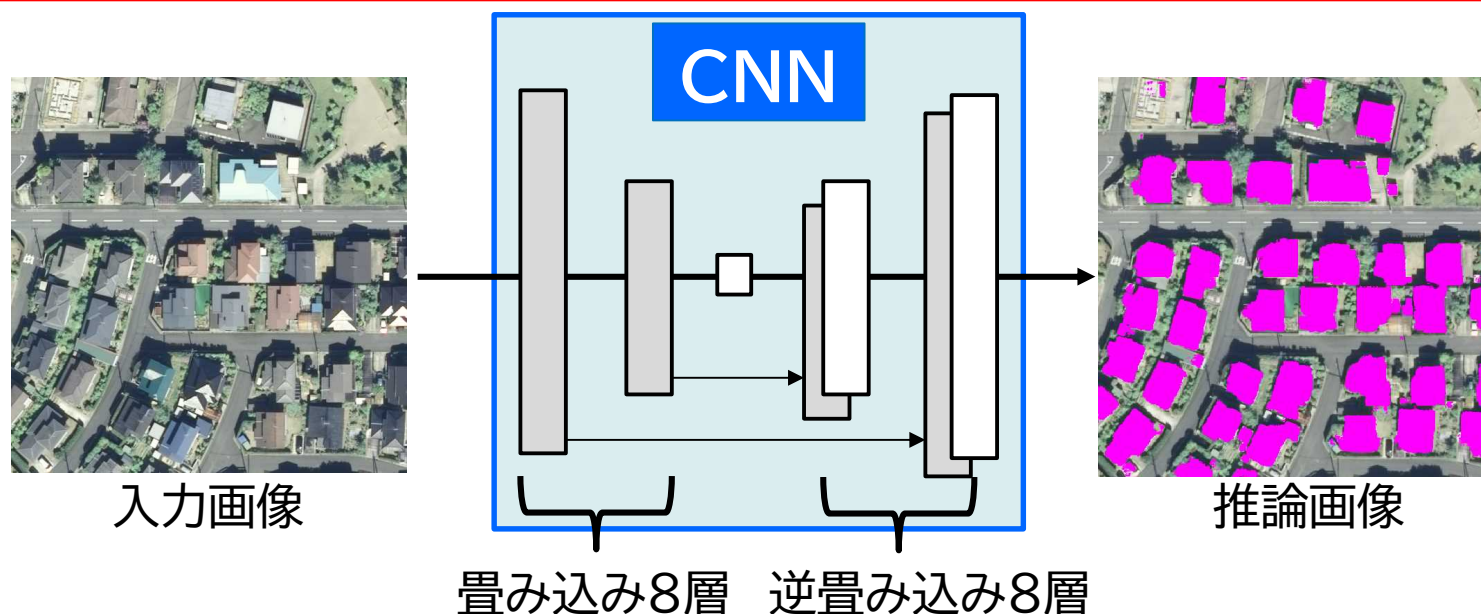
- ・目的関数とCNNの検討のために、学習データを終始、固定
- ・目的関数とCNNの構成を1つずつ変更しながら、その都度深層学習

- ・性能評価は、学習済モデルで **検証データ** を推論させた結果から逐一スコア(F値)を算出して実施

② 深層学習におけるCNNと目的関数の検討

CNNは、複数のニューラルネットワークを比較して選定。

- ・ 研究着手時に最も高性能であったネット(U-Net構造)を採用
- ・ その後、さらに高性能といわれるCNNも出現、CNNの構造に反映予定



- ・ CNNの構成は、畳み込み、逆畳み込みとも各8層構成を採用。
 - スコアのみでなく、学習時間や学習データ作成効率なども勘案して決定
- ・ 目的関数は、各画素値の差の絶対値の総和(L1ノルム)を採用。

→ Conditional GAN(Isolaら 2017)のGenerator部分とほぼ同じ構成

- ・ 最終年度に高性能なCNNのアイデアを本研究のCNN構成に反映予定

・ CNNの構造図はConditional-GAN (Isolaら 2017) を元になっている。

4. 学習データの構成の検討 とAIの構築

③ 学習データ構成方法の検討

学習データは、項目毎に作成する。

- ・「抽出性向」を意識して対象画像を選定する
- ・ 計算資源は有限であるため、無駄な学習データはできるだけ排除する

学習データは、単に数を増やすだけでは性能向上への効果は少なく、計算資源を浪費するだけ。良質なもので少ない枚数でも性能は向上する。

学習データの実例。塗分け色は検証データと異なることがある。

●道路



●人工斜面・土堤



●太陽光発電設備



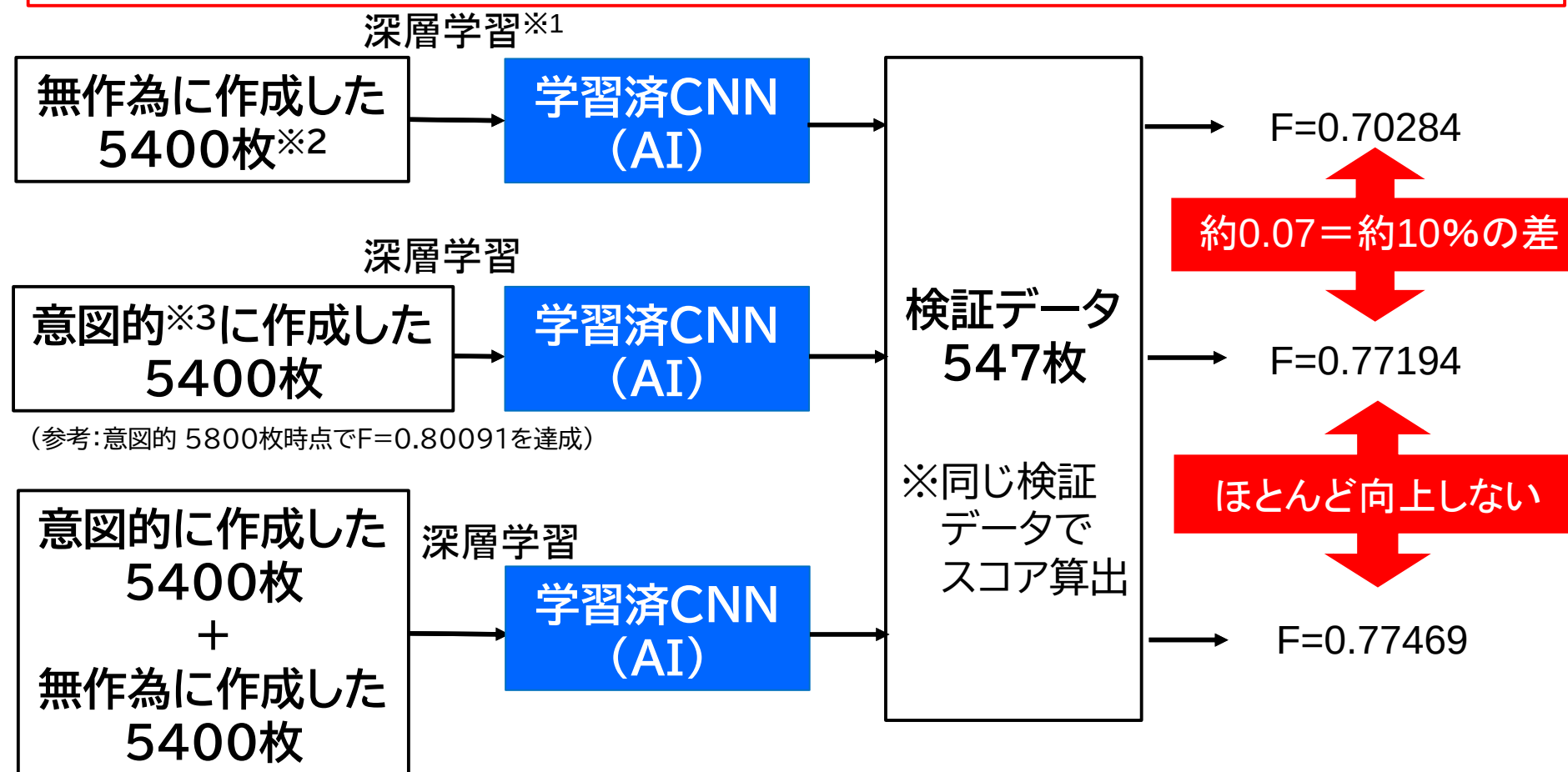
●普通建物



③ 学習データとAIの性能の関係

道路の学習データの作成方法で比較。CNNの構造と目的関数は固定。

- ・ 本研究で醸成したノウハウに基づき意図的に作成すると、より少ない枚数でF値が向上することを確認



※1 深層学習の各種条件は同じ。

※2 実際の学習に用いる枚数は方向別の回転を加えて4倍の21600枚になる。意図的なものも同じ。

※3 「意図的」の中身は、本研究の過程で得た画像選択のノウハウ群であり、「スコア算出時の検証データの推論画像の不良部を見て、それを改善する」等の不正な手段でスコア向上を図るものではない。

③ AI の抽出性向 (の制御)

AND型とOR型の2種類の抽出性向を持つAIに学習し分けている。

■ AND型 (正射影を抽出する必要がある地物用)

- ・ 複数の方向から撮影された画像の抽出結果の論理積(AND)を取るもの
- ・ 適合率より再現率を高めに誘導するよう学習データを構築
 - 誤抽出があっても良いのでできるだけ漏れなく抽出するよう育成

特に工夫せず学習データを作成すると、このタイプのAIに育つ傾向。

実装のイメージ

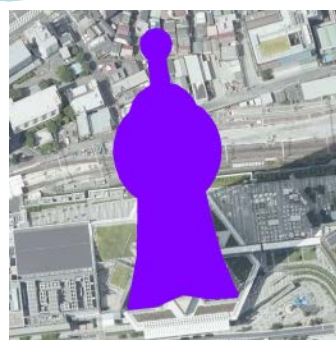
東京スカイツリー
の場合

CKT20191-C23-52
(90度回転)



+

CKT20191-C23-46
(90度回転)

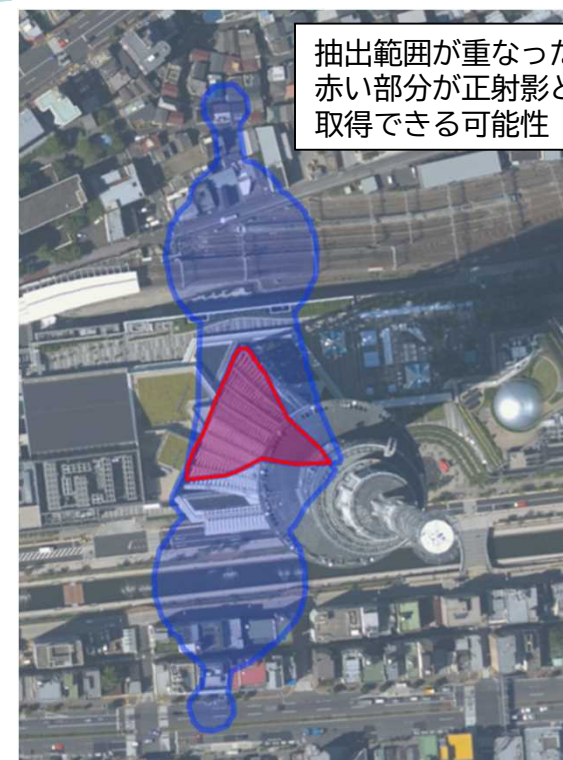


+



正射変換

論理積



抽出範囲が重なった
赤い部分が正射影として
取得できる可能性

高塔又は堅牢建物として抽出
(イメージ)

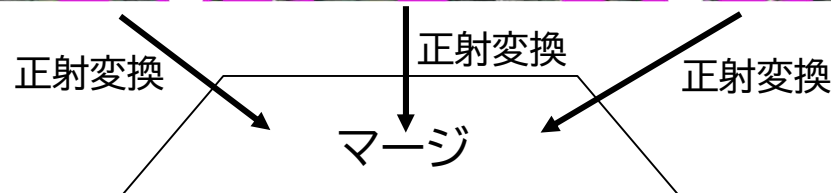
③ AI の抽出性向 (の制御)

■ OR型 (一般的な地物用)

- ・ 複数の方向から撮影された画像の抽出結果の論理和(OR)を取るもの
- ・ 再現率より適合率を高め誘導するよう学習データを構築
 - 漏れが生じてても良いので、誤抽出をできるだけ少なくするよう育成

→ 濁水抽出用AIで実証済。生の空中写真を使用するメリットが大きい。

実装 (次ページにも道路での実証例)

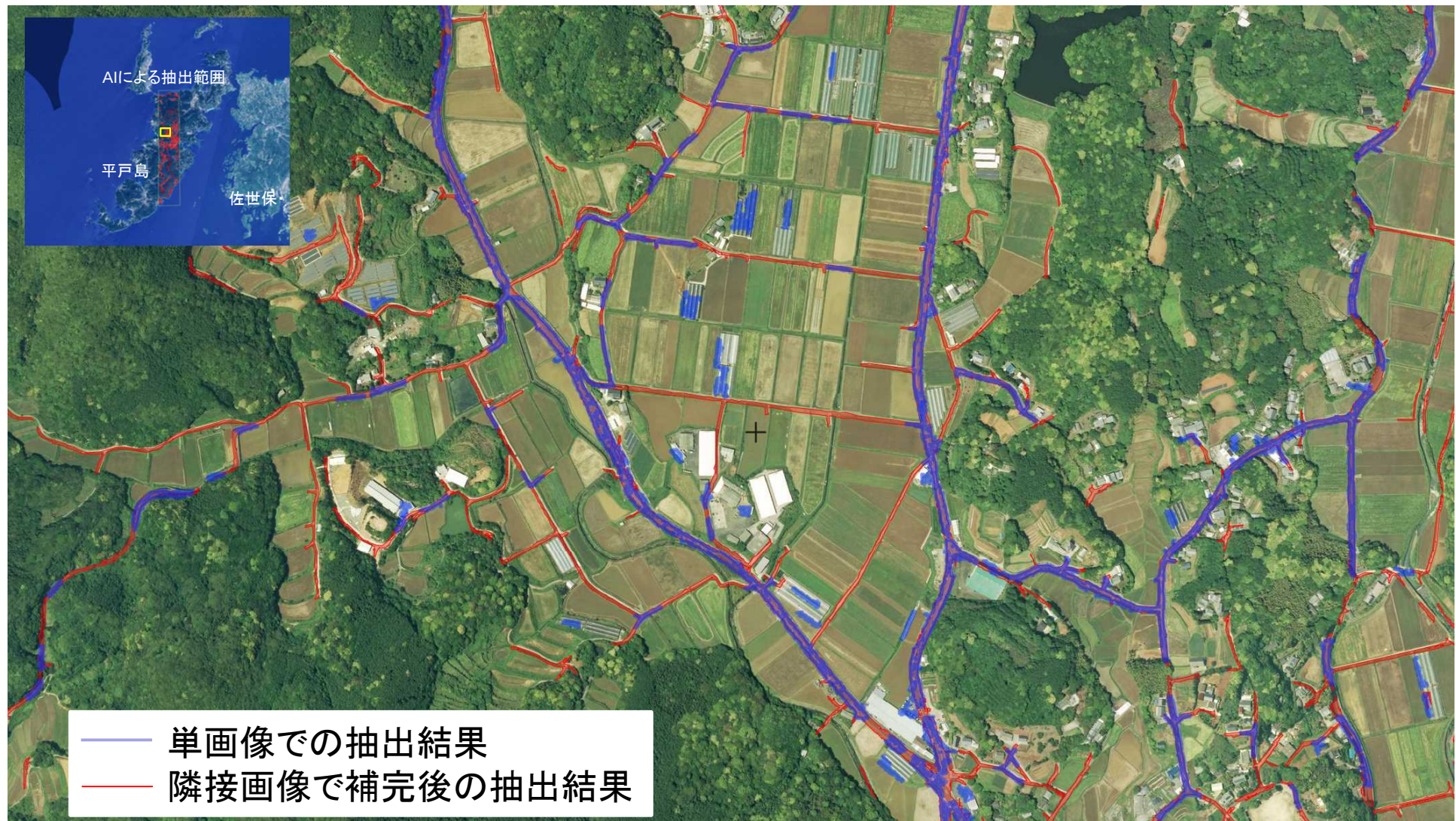


※1つの画像で抽出漏れの部分が、別の画像の抽出結果によって補完される。
 ※各画像の誤抽出が多いとマージ後に誤抽出だらけになってしまう。

複数の抽出結果を地図上でマージすることで抽出漏れを軽減できるのでは、という発想

→ 生の空中写真(とカメラパラメータ)を使うことで可能となった手法

③ OR型AIの抽出例（道路縁）



（単画像）CKU-2017-4-N-C04_0009に、両隣のC04_0008、C04_0010の抽出結果をマージ

※青色のみで赤色と重なっていない部分は、マージ中の処理で誤抽出と判断され、最終結果から削除されたところ

※樹木に完全に覆われたところを除き、赤色の抽出結果での抽出漏れはほぼ見られない → これによるF値の上昇分は未計測

5. 目標達成状況の自己評価

数値目標①

40項目以上の地物に対し、基盤地図情報の精度の範囲内で、F 値 ≥ 0.8

達成できているのは

- | | |
|-------------------------|-----------|
| ・ 道路 (2101道路縁、2106庭園路等) | F 値 0.807 |
| ・ 水部 (5101水涯線、5105湖池) | F 値 0.862 |
| ・ 建物 (3001普通建物) | F 値 0.803 |

の5項目のみ。

➡ **達成率12.5%**

主要項目に絞って学習データの効率的な作成方法などの基礎的な検討を優先していたため(これらは目標には反映されない)、達成率は低い。

しかし、現在の研究の流れにおける個別項目のF値向上は第3段階(Slide14)のタスクである。主要項目では目標をクリアし、おおむね計画どおり。

数値目標②

レコード数ベースの抽出率で、80%以上

対象となる電子国土基本図 (地図情報) のレコード数は、約1億8097万レコード。

- | | | |
|---------------------|-------------|----------|
| ・ 道路 (占有率 18.87%) | × F 値 0.807 | = 15.23% |
| ・ 水部 (占有率 7.17%) | × F 値 0.862 | = 6.18% |
| ・ 普通建物 (占有率 34.54%) | × F 値 0.803 | = 27.74% |

合計値 49.15% ÷ 80.00%

➡ **達成率 61.44%**

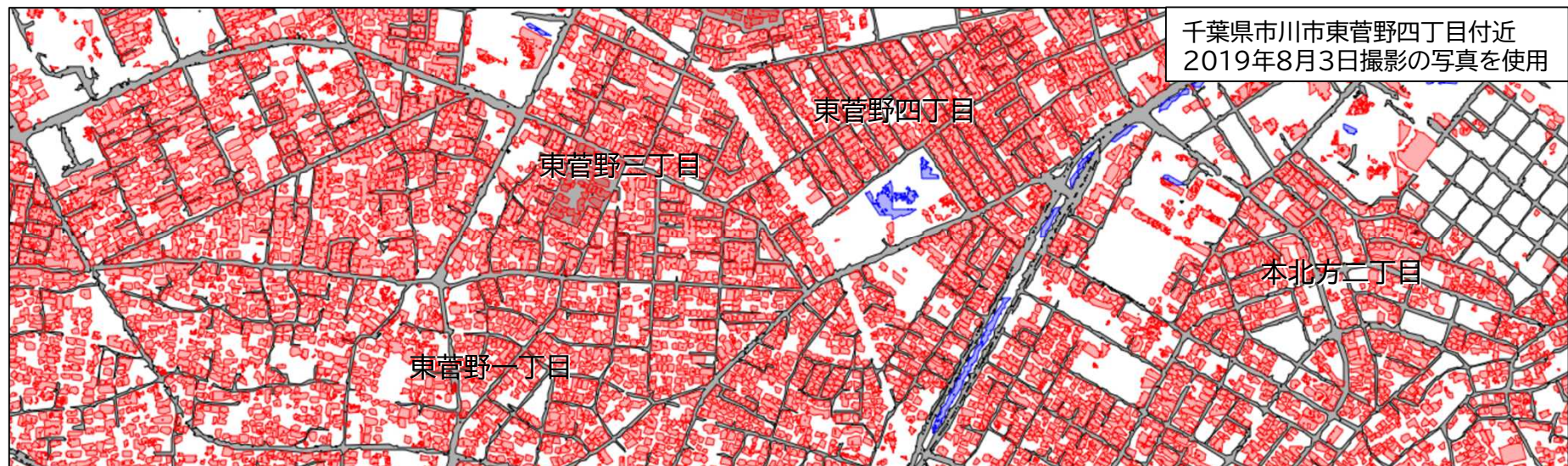
6. 本研究の出口戦略

研究要素を組み合わせると、空中写真から地図データを作成するAIになる

・空中写真1枚(約10km²)あたり処理時間およそ3～20分

(例)普通建物の場合、神奈川県東半分(空中写真1930枚)での抽出が約半日

内訳: AIで約9時間+正射変換約5時間(10並列で処理※)



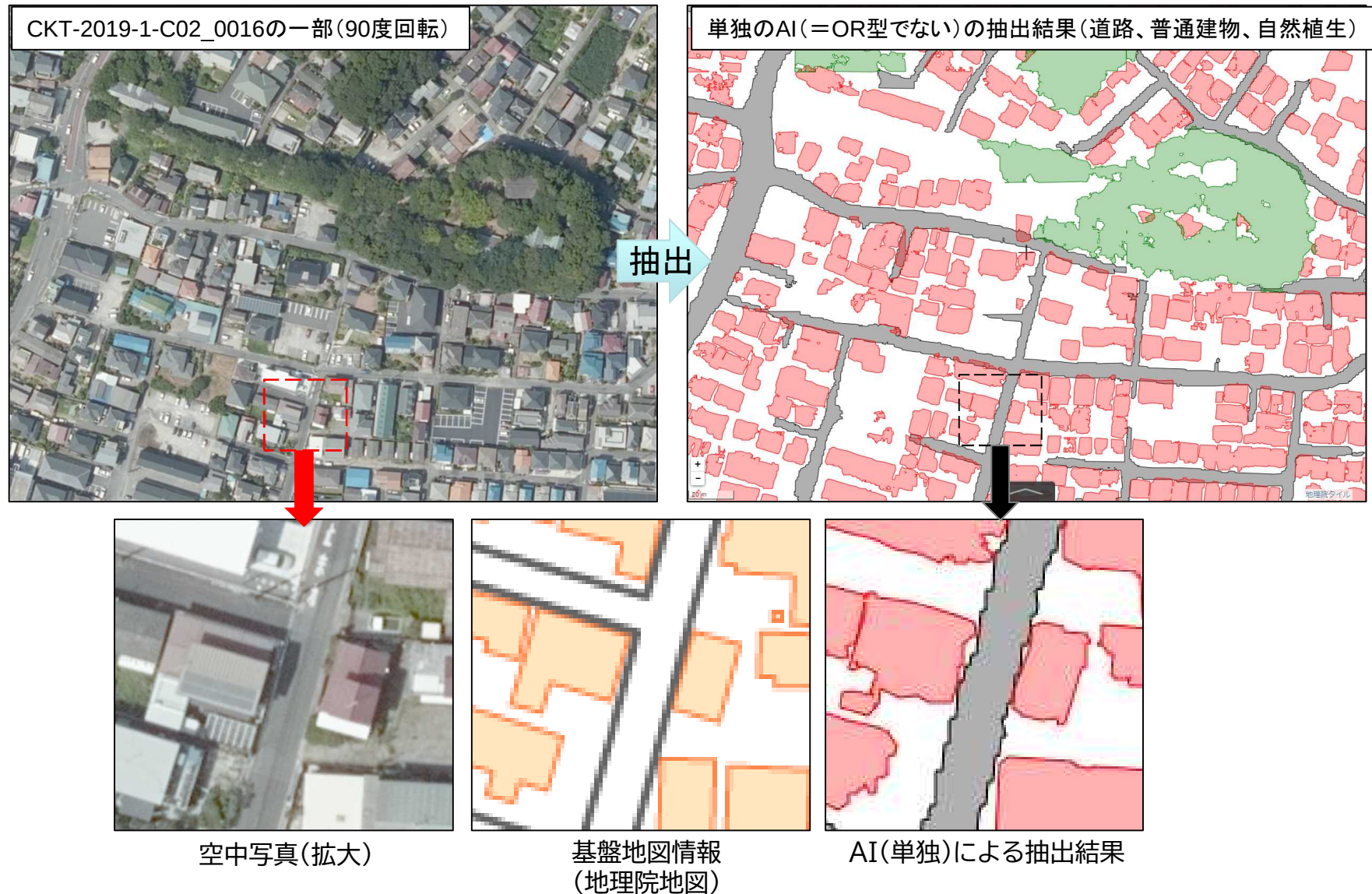
2020/12/14時点のOR型AIによる(道路、普通建物、水部の)自動抽出結果をそのまま地理院地図上で表示したものの一部を抜粋。
街区が灰色に塗りつぶされているところは表示上の問題が生じている箇所、データ自体は正しく生成されている。
中央部にある線状の白部には実際は細い水路があり、水部の抽出漏れ。注記は手入力。

・ただし、測量成果とするには、さらに形状を整形する等の処理(後続の「地図編集」の工程)を加える必要があり、現時点で「測量成果としての地図」の自動作成まで到達しているわけではない。

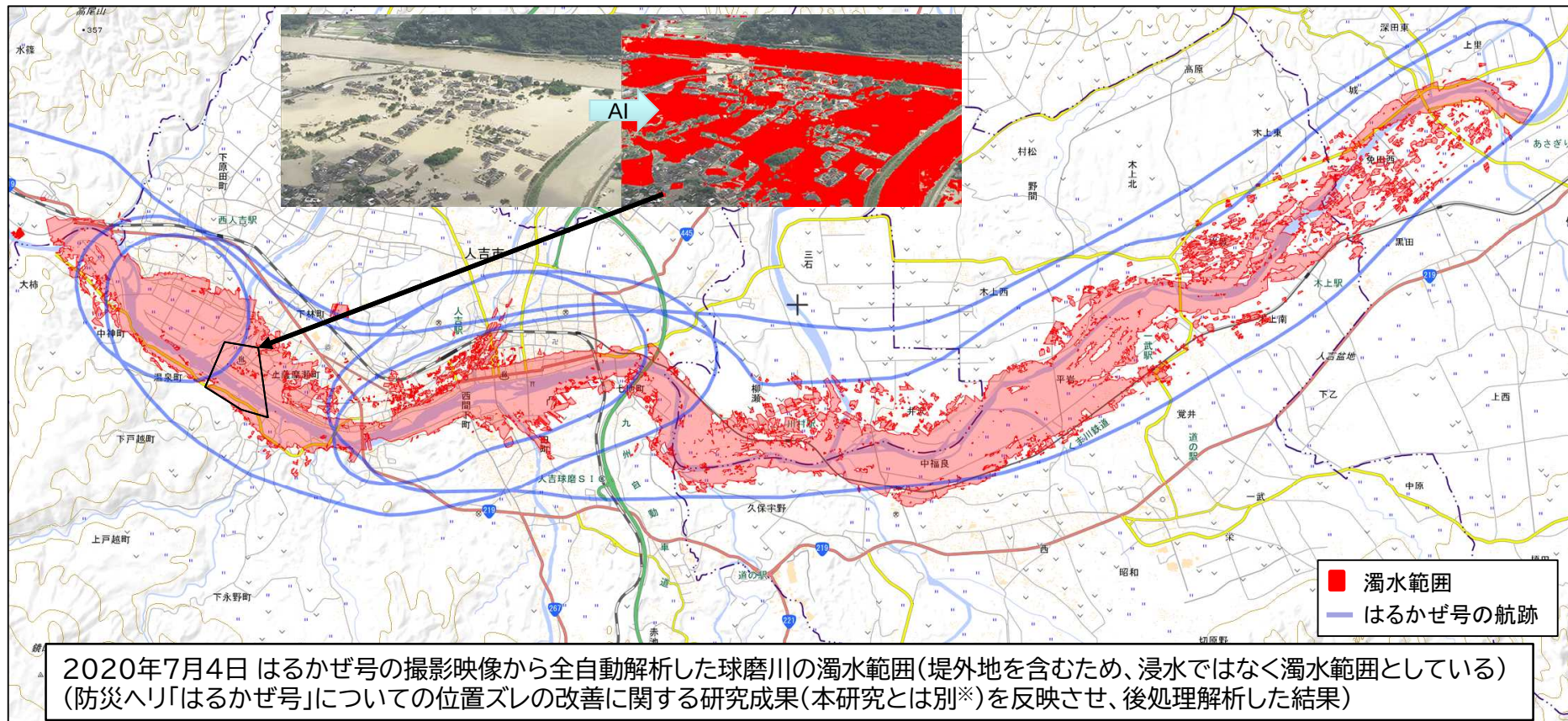
※AIでの抽出は10並列で実施した実時間(約9時間18分)。正射変換は並列処理なしの処理時間(約47時間)を10並列分に換算。

地図自動作成 AI としての動作（中間まとめ）

- ただし、測量成果とするには、さらに形状を整形する等の処理（後続の「地図編集」の工程）を加える必要があり、本研究のみで「測量成果としての地図」の自動作成まで到達するわけではない。



AIの部分を入れ替えるだけで様々なものの図化に対応させることができる。
→OR型AIが、防災ヘリの映像内の濁水抽出に用いられて試験運用中
→撮影時点から30秒遅れで次々に濁水範囲が地図データ化されていくもの



(比較のための参考)

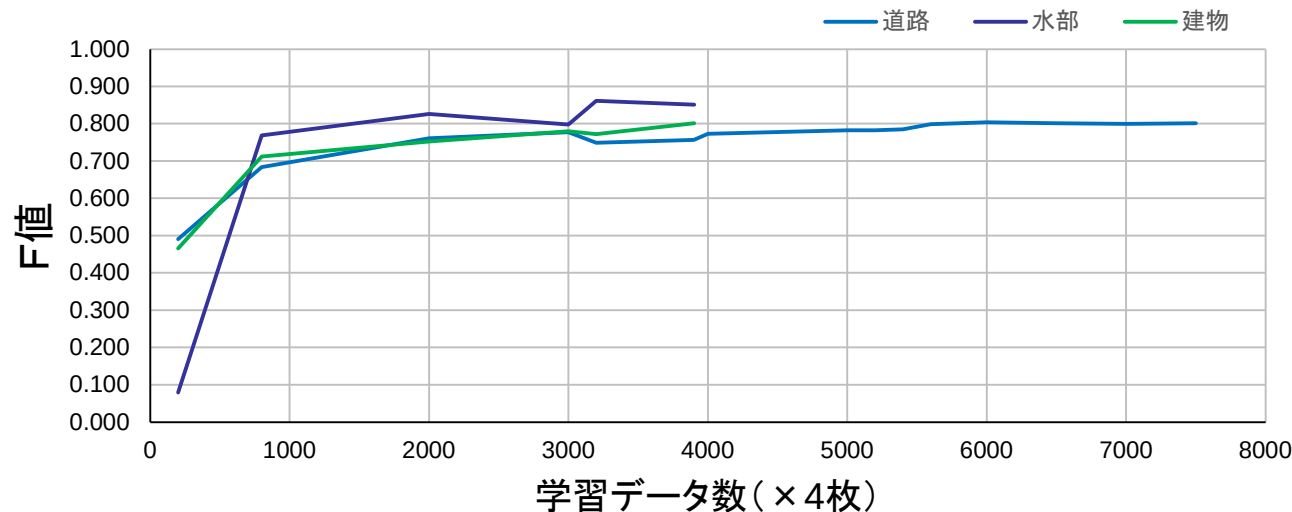
浸水推定図 https://www1.gsi.go.jp/geowww/saigai/202007/shinsui/02_shinsui_kuma_tougou.pdf

※災害状況自動判読システムの開発(総合流域防災対策事業調査費:H30~R3)



7. 現時点の課題と今後の研究内容

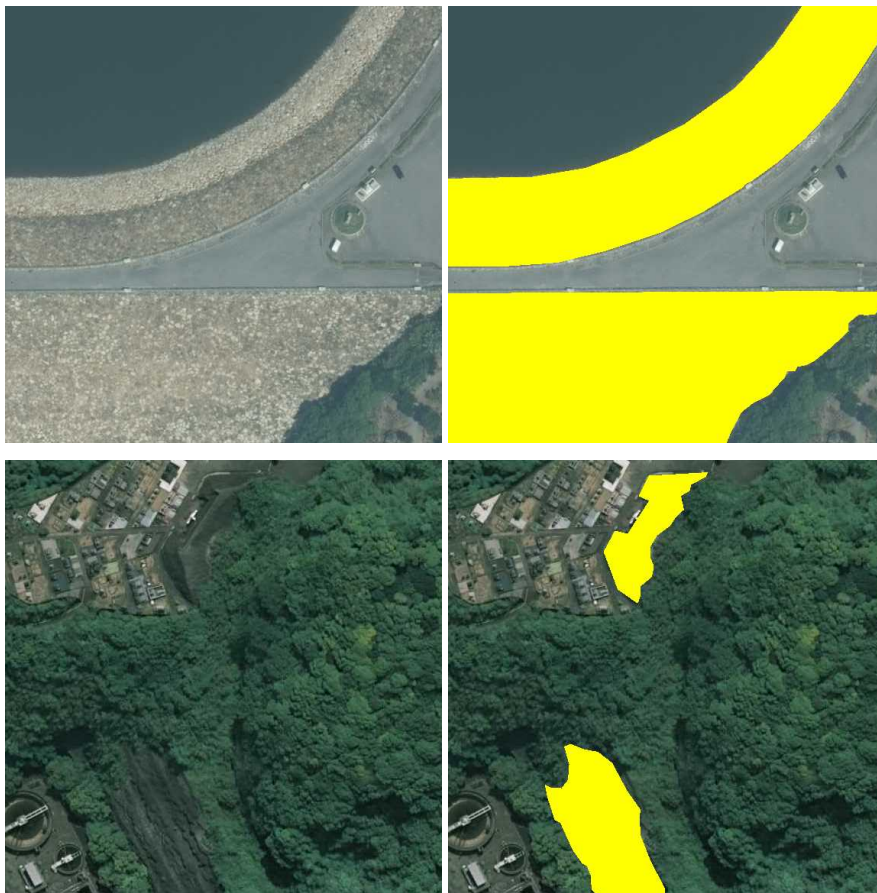
- 項目別数値目標の達成率が低い。
- 学習データの整備促進。F値 ≥ 0.8 の項目の増加を図る。
 - ・AND型AI対応の3項目を早期に整備し、実証
 - ・その他、数値目標未達成の項目別学習データの整備促進



- ・学習データ数が増加するにつれて、F値は緩やかに上昇する傾向がみられる
→ 学習データ数の増加が有効(基本的スタンス)
- ・1000枚以下の段階でF値が向上してこない地物は、本質的に本研究の手法がそぐわない可能性(次スライド)

- 高さに関与する地物の抽出は、これまでの手法で対応できないことが判明
- 高さを認識できる新たな抽出手法の開発
 - ・被覆、人工斜面は現時点の手法のみでは抽出不可能と考えられる
 - ・AND型AIをさらに応用した抽出手法開発が必要

【被覆の学習データの例】



【人工斜面(土がけ)の学習データの例】

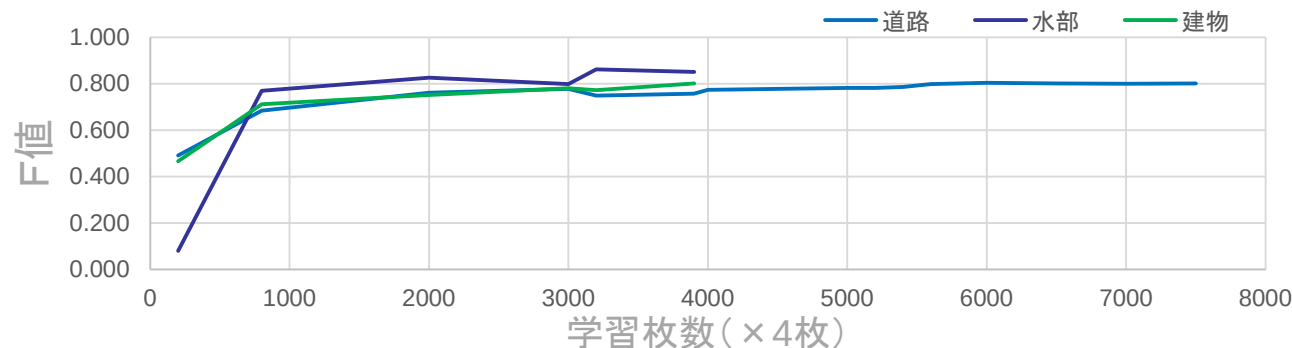


【被覆の最大F値:学習データ400枚時点で0.029】

【人工斜面の最大F値:学習データ500枚時点で0.010】

■ 学習データの整備促進。F値 ≥ 0.8 の項目を増やすために。

- ・AND型AI対応の3項目を重点整備し、実証
- ・その他、数値目標未達成の項目別学習データの整備。



■ 高さを認識できる新たな抽出手法の開発

- ・被覆、人工斜面は現時点の手法のみでは抽出不可能と考えられる
- ・AND型AIをさらに応用した抽出手法開発が必要

■ 先進光学衛星と3次元点群は研究対象から除外

- ・先進光学衛星の打ち上げが延期となったため
- ・3次元点群は地図情報レベル500～1000に適しており、本研究ではオーバースペックのため

- 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, 2018
- Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, Alexei A. Efros(2017): Image-to-Image Translation with Conditional Adversarial Networks, CVPR2017, 2017.
- Ryuhei Hamaguchi、Shuhei Hikosaka(2018): Building Detection from Satellite Imagery using Ensemble of Size-specific Detectors, CVPR2018, 2018.
- 大野裕幸, 遠藤 涼, 中埜貴元, 篠田昌子(2019): 空中写真を用いた地物抽出性能評価用データセットの構築, JSAI2019, 2019.

参考資料

- **畳み込みニューラルネットワーク(Convolutional Neural Network)**

入力側から次の層の値を算出する際に畳み込み演算を用いるニューラルネットワーク。画像を処理する際に用いられることが多く、畳み込み演算に用いるフィルタのサイズや演算間隔によってさまざまな効果が得られる。

- **U-Net構造(のニューラルネットワーク)**

CNNの中身を図にしたときに、入力から出力に向かってU字型に層(レイヤ)が描かれる。U-Netという医療用のCNNがそのような構造を取っていたことから、U-Net構造と呼ばれるようになった。

本研究で用いているU-Net構造は、Conditional-GAN(Isola他, 2017)の発生器に用いられているものとほぼ同じ。

- **GPU(Graphics Processing Unit)**

もともとは画面表示を高速化するための専用部品だったが、深層学習用のライブラリで使えるようになってからは深層学習を高速に実施するために必須の部品となった。

本研究では、学習には主にGeForce RTX3090を、推論にはGeForce GTX1080Tiを、実装系の推論にはQuadro RTX8000を使用している。

- ・GAN(Generative Adversarial Networks:敵対的生成ネットワーク)

入力画像から(偽)画像を生成する発生器(Generator)と、発生器が生成した偽画像と正解画像を比較して偽であることを見破る判別機(Discriminator)で構成されるニューラルネットワーク。

発生器は判別機をだませる(正解にできるだけ近づけた)画像を生成するよう学習され、判別機は、どれだけ正解に近い画像でも発生器が作成した画像は偽と見破るように学習される。発生器と判別機が競い合って学習する様から「敵対的生成」と呼ばれる。

抽出対象の地物一覧

●クラス分けの番号と標準表示色。

番号	全項目色(RGB)	項目名
2	(127,127,127)	舗装道路
3	(191,127,63)	未舗装道路
4	(95,95,95)	駐車場
5	(255,0,127)	中央分離帯
6	(63,0,0)	軌道
7	(0,0,255)	水部
8	(255,0,0)	堅牢建物
9	(255,0,255)	普通建物
10	(255,127,127)	無壁建物
11	(127,127,0)	被覆
12	(191,191,191)	土がけ
13	(0,127,255)	水制
14	(127,191,0)	横断歩道橋
15	(0,127,0)	自然植生(樹木)

番号	全項目色(RGB)	項目名
16	(255,255,0)	畑
17	(0,255,255)	水田
18	(0,127,127)	収穫後の水田
21	(127,255,127)	茶畑
22	(255,127,0)	果樹園
23	(127,0,0)	荒地
24	(31,31,31)	空地
27	(191,191,0)	太陽光発電設備
28	(191,0,191)	輸送管
34	(255,191,255)	石段
36	(191,127,191)	雪覆い等
38	(127,191,191)	プラットホーム
39	(255,127,191)	タンク
44	(255,191,191)	防波堤

番号	全項目色(RGB)	項目名
54	(191,255,191)	芝地
58	(255,191,0)	砂礫地
61	(191,191,255)	たたき(コンクリート舗装された空間)
62	(127,63,63)	岩がけ、露岩
63	(127,0,255)	高塔
65	(0,0,127)	湿地
66	(191,0,255)	材料置場
67	(127,191,127)	園庭
68	(127,127,255)	墓

- ・他に01影、255未分類がある。
- ・24空地、28輸送管は現在は研究対象から除外。
- ・17水田と18収穫後の水田は、研究対象としては合わせて1項目とカウント

項目別の抽出性能(2021-02-24時点)

項目	学習枚数	F値	項目	学習枚数	F値	項目	学習枚数	F値
02・03道路	8400	0.807	17・18水田	2500	0.557	58砂礫地	1400	0.324
04駐車場	1200	0.563	21茶畑	600	-	61たたき	80	-
05分離帯	900	0.113	22果樹園	500	0.053	62岩がけ	400	0.021
06軌道	1500	0.380	23荒地	400	0.132	63高塔	800	0.111
07水部	4200	0.862	27メガソーラー	1600	0.461	65湿地	400	-
08堅牢建物	1200	0.199	32雨裂	1800	0.345	67園庭	30	-
09普通建物	5800	0.803	34石段	400	-	68墓地	900	0.025
10無壁建物	400	0.169	36雪覆い等	400	-	(15自然植生)	1000	0.853
11被覆	500	0.029	38プラットホーム	700	-	19竹林	画像分類	0.652
12土がけ	600	0.012	39タンク	500	-	20針葉樹林	画像分類	0.707
13透過水制	1100	0.470	41風車	900	-	26しの地	画像分類	0.471
14横断歩道橋	950	-	43不透過水制	100	-	40広葉樹林	画像分類	0.690
16畑	3800	0.442	54芝地	400	-	(00全項目)	1500	0.466

※実際のトレーニングに使用する学習枚数は、3方向に回転させたものを加えるため、4倍となる。

※F値欄「-」は、F値未算出。項目欄カッコ書きは参考値。00全項目の評価値はmIoU。画像分類の項目はクロスエントロピーによる top-1集計値。

※F値の算出には検証用のGSI-547データセット(学習枚数換算2188枚分)を使用。

※F値欄、下線のある数値は、GSI-547データセットに加えて、マイナー地物用拡張GSIデータセット-52(学習枚数換算5722枚分)を使用。

項目別の占有率(数値目標2の算出用)

項目	占有率	F値	項目	占有率	F値	項目	占有率	F値
02・03道路	18.87	0.807	16畑	1.76		58砂礫地	1.92	
04駐車場	0		17・18水田	2.25		61たたき	0	
05分離帯	0.18		21茶畑	0.04		62岩がけ	0.29	
06軌道	0.17		22果樹園	0.30		63高塔	0.00	
07水部	7.17	0.862	23荒地	0.41		65湿地	0.03	
08堅牢建物	1.23		27メガソーラー	0		67園庭	0	
09普通建物	34.54	0.803	32雨裂	0.36		68墓地	0.03	
10無壁建物	4.08		34石段	0.02		(15自然植生)	-	
(09建物三種)	-		36雪覆い等	0.00		19竹林	0.04	
11被覆	2.08		38プラットホーム	0.01		20針葉樹林	1.76	
12土がけ	16.32		39タンク	0.00		26しの地	0.06	
13水制	0.09		41風車	0.00		40広葉樹林	1.81	
14横断歩道橋	0		54芝地	0		(00全項目)	(100.0)	

※占有率の単位は(%)。小数点以下3桁目を四捨五入。

※36雪覆い等は、0.0007%。39タンクは、0.0049%。41風車は、0.0008%。

※63高塔は、0.0003%。ただし、国土地理院のデータに送電線の鉄塔は含まれない。

※占有率が「0」の項目は、国土地理院の地図データベースにデータ種別が存在しない項目。