

# 再帰構造を用いた全層畳み込みニューラルネットワークを利用した 航空写真からの道路領域抽出

杉田 尚弥<sup>1</sup> 尾島 修一<sup>1, a)</sup> 植村 匠<sup>1, b)</sup>

**概要：**近年，インターネットやスマートフォンの普及により，デジタル地図を伴うサービスが普及し，デジタル地図の活用がますます進んでいる．しかし，デジタル地図の作成は未だに人の手によるものが多く，更新速度の遅さなど時間的なコストが掛かるなどの問題が挙げられる．そこで筆者らは航空写真から道路領域の自動認識に関する検討を行う．手法として，全層畳み込みニューラルネットワークに再帰構造とグローバル畳み込みネットワークを用いることで従来手法よりも高い精度で，道路領域の認識を可能とした．

キーワード：機械学習，地図情報，画像認識

## Extraction of road areas from aerial photographs using full layer convolution neural network with recursive structure

Naoya Sugita<sup>1</sup> Ojima Shuichi<sup>1, a)</sup> Uemura Takumi<sup>1, b)</sup>

**Abstract:** In recent years, with the spread of the Internet and smart phones, services accompanying digital maps have become widespread, and the utilization of digital maps is progressing more and more. However, the creation of digital maps is still many people's hands, and problems such as the time required for updating such as slow updating speed are raised. Therefore, the authors examine the automatic recognition of the road area from the aerial photograph. As a method, by using the recursive structure and the global convolution network for the all - layer convolution neural network, it is possible to recognize the road area with higher accuracy than the conventional method.

**Keywords:** Machine Learning, Map information, Image recognition

### 1. はじめに

紙面からデジタルへの地図媒体の移行により，利用者の目的に合わせて詳細かつ柔軟に情報を提供することが可能になった．デジタル地図情報は古くからカーナビゲーションシステムから始ま

り，交通情報の管理，都市計画，防災・災害対策など広く利用されてきた．そしてインターネットの普及に伴いデジタル地図を利用した地図案内サービスが始まり，スマートフォン等の携帯端末の普及により多様なサービスが展開されるようになった．Google mapのように地図と航空画像や車載カメラで撮影した実写画像を重ね合わせて表示するサービスなども出ており，今後ますますデジタル地図が活用されていくと思われる．そこで，

---

<sup>1</sup> 崇城大学

a) ojima@cis.sojo-u.ac.jp

b) t\_uemura@cis.sojo-u.ac.jp

デジタル地図は今まで以上にデータの新鮮さや精度の高さが要求されるようになった。

現在、一般的に実在するデジタル地図は、既存の紙面地図からのトレースといった人の手による作業で作成されている。そのため、地図を更新するにあたって、現地測量より紙面地図を作成する作業も含め、多くの時間もしくは人手が必要であった。現地測量を航空写真測量にすることで作成期間の短縮を図ることができたが、測量結果からデジタル地図データを作成する際には依然として人の手による作業が必要であり、測量時からデジタル地図データの完成までには短縮されたとはいえ、長い期間が必要となる(図1)(a-d)。

これらの問題を解決する手法として、画像認識技術を用いて紙地図や航空画像からデジタル地図を自動的に作成、および更新する技術の実現に期待が高まっている。

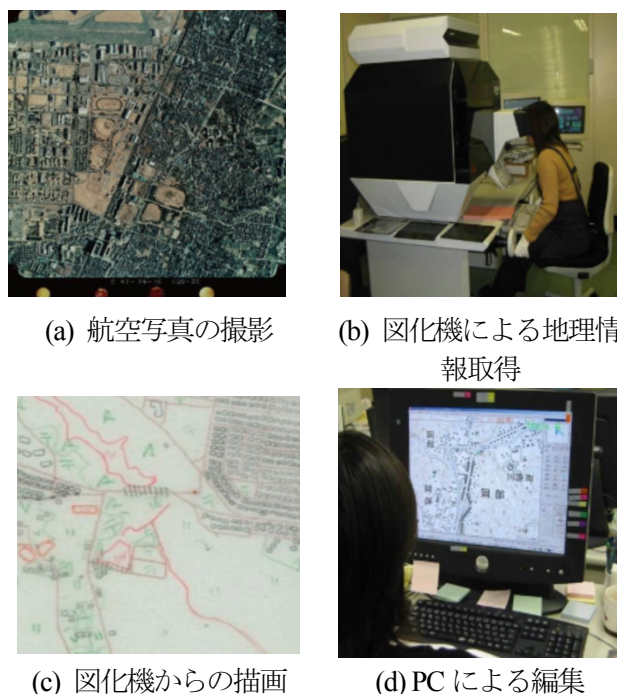


図1 デジタル地図作成の流れ

## 2. 従来手法

デジタル地図を作成する前処理として、航空写真や衛星画像から道路領域を抽出する手法が数多く提案されている。近年では、深層学習を用いて地物を認識・抽出する手法が数多く提案されており、高い成果を収めている。

### 2.1 畳み込みネットワーク

中村らは、深層学習を用いた衛星画像における地物認識手法を提案した。この手法では、一般物体認識の分野で従来手法に比べて圧倒的に高い性能を収めている Convolutional Neural Network (以下 CNN) と呼ばれる深層学習モデルを用いた。CNN は、空間分解能の高い航空写真や衛星画像において既存の特徴量と同等以上の性能を示すことが知られている。そこで、この研究では、中程度の空間分解能(~30m)の大規模な衛星画像データに対して地物の自動認識を行った。実験の最初のステップとして、比較的大型の太陽光発電所すなわちメガソーラーの自動検出・識別を試みた。この結果、ほとんどのメガソーラーが検出できたが、使用した画像がメガソーラーと周辺の色が分離し易い画像であった[1]。

### 2.2 全層畳み込みネットワーク

高橋らは、再帰構造を用いた全層畳み込みニューラルネットワーク(FCN: Fully Convolutional Network)により、航空写真における建物領域をセグメンテーションする手法を提案した。この手法では、すべての層を畳み込み層のみで構成したモデルを使用することで、入力画像と同じサイズの確率マップ出力が得られるようにしている。これにより、入力画像に対して pixel 単位で建物認識が可能となる。さらに、中間層にショートカット接続を利用した畳み込み層のブロックを構築し、このブロックを再帰的に配置している。この結果、Intersection Over Union による評価で既存手法を上回る性能を示した[2]。

### 2.3 グローバル畳み込みネットワーク

Chao Peng らは、画素レベルでそれが何かを認識するセマンティックセグメンテーションの精度改善のため、大規模カーネルの畳み込みを有するエンコーダ・デコーダアーキテクチャを提案した。セマンティックセグメンテーションではセグメント化されたオブジェクトの分類と局在化の2つのタスクが問題となる。そこで、Chao らはこれらの問題を解決する手法として Global Convolutional Network(以下 GCN)を提案した[3]。

### 3. 提案手法

本研究では、高橋らのモデル構造に変更を加えることで、より高い精度で航空写真から自動的に道路領域を認識することを目的とする。そこで、本研究では図2に示すモデルを用いて航空写真における道路領域を高い精度で分類する手法を提案する。このモデルでは、従来手法で検証がされていなかったFCNのエンコーダ・デコーダ部分のモデルの構造に、近年セグメンテーションの分野で著しい成果を挙げたGCNを用いている。図2に示すupsampleでは特徴マップの大きさを最近傍補間により2倍に拡大している。また、各畳み込み層の出力に対して活性化関数にReLU、デコーダの最終層の出力に対してはシグモイド関数を用い、活性化関数の初期値としてそれぞれHeの初期値、Glorotの初期値を用いた。モデルの出力に対し誤差関数を最小二乗誤差、モデルの重みの更新にはADADELTA[4]を用いた。

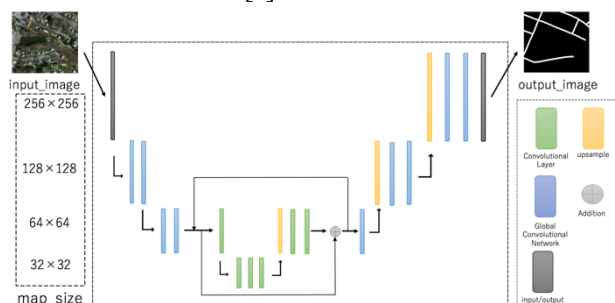


図2 本研究で用いたモデルの概要

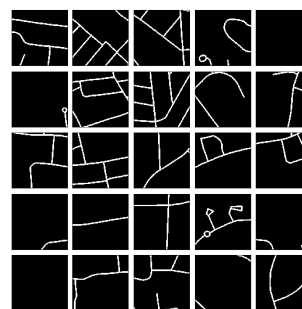
## 4. 実験と結果

### 4.1 実験データ

本実験ではデータセットとして Massachusetts Roads Dataset[5]を用いた。本データはRGB航空写真と道の領域が白で表された2値画像からなり、Training Set 1108枚、Test Set 49枚、Validation Set 19枚で構成され、画像サイズ1500×1500 pixel、空間解像度0.7 m/pixel、濃度諧調数8bit×3chのデータで構成されている。学習のために、1500×1500 pixelの航空写真と同じ大きさで道路にラベル付けされた正解画像から、それぞれ256×256 pixelの部分を切り出した。切り出した画像は図3に、画像の詳細は表1にそれぞれ示す。



(a) 航空写真



(b) ラベル付けされた正解画像

図3 学習用データの例

表1 使用した画像の詳細

|                |                 |
|----------------|-----------------|
| Training Set   | 27,700 枚        |
| Validation Set | 350 枚           |
| Test Set       | 1225 枚          |
| 画像サイズ          | 256×256 [pixel] |
| 空間解像度          | 0.7 [m/pixel]   |
| 濃淡諧調数          | 8 [bit]         |
| チャンネル数         | RGB 3 チャンネル     |

### 4.2 実験方法

本実験では図3に示す学習用データを用いてバッチサイズを8、epoch数を最大300として学習させた。GCNの有用性を示すために従来手法のモデルのエンコーダ・デコーダ部分をそれぞれGCNに置き換えたモデルを学習させた結果の比較を行った。モデルの構造は図4に示すとおり、再帰ネットワークはすべてのモデルで同じ構造とし、GCNを用いていない従来手法のモデル(OE/OD model)、エンコーダのみGCNを用いたモデル(GE/OD model)、デコーダのみGCNを用いたモデル(OE/GD model)、エンコーダとデコーダの両方に

GCN を用いたモデル(GE/GD model)の計 4 つである。

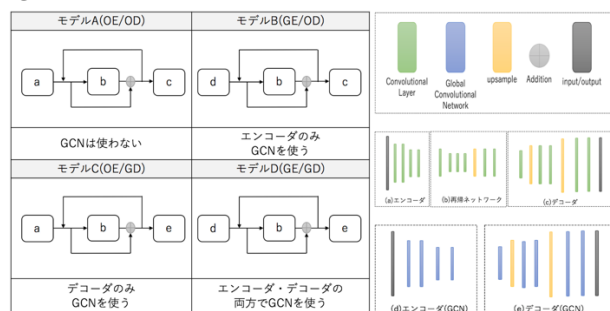


図 4 比較したモデルの概要図

### 4.3 評価方法

評価方法は C.Heipke らにより提案された評価指標[6]を用いて精度の比較を行った。この評価手法では抽出精度を完全性、正確性、品質の 3 種類の指標により評価を行う。ピクセル単位の抽出結果を true positive (TP), true negative (TN), false positive (FP)および false negative (FN)の 4 種類に区分する。これらは式(1)-(4)で定義され、式中の  $S_e$  は抽出された道路領域,  $S_t$  は正解データの領域を表す。

$$TP = S_t \cap S_e \quad (1)$$

$$TN = \bar{S}_t \cap \bar{S}_e \quad (2)$$

$$FP = \bar{S}_t \cap S_e \quad (3)$$

$$FN = S_t \cap \bar{S}_e \quad (4)$$

完全性は、正解データのうち正しく抽出された pixel の割合として式(5)で定義される。

$$\text{完全性} = \frac{TP}{TP+FN} \quad (5)$$

正確性は、抽出結果のうち正解データが含まれる pixel の割合として式 6 定義される。

$$\text{正確性} = \frac{TP}{TP+FP} \quad (6)$$

品質は、完全性と正確性を総合的に評価する指標であり、式 7 で定義される。

$$\text{品質} = \frac{TP}{TP+FP+FN} \quad (7)$$

### 4.4 実験結果

実験結果として、正解データのうち正しく抽出されたピクセルの割合を示す **Completeness** を評価指標として求めた各モデルの推論結果を表 2 に示す。

表 2 からエンコーダ・デコーダの両方に GCN を用いたモデルの評価結果が最も高いことが分かる。

表 2 テストデータにおける評価結果

| モデル      | 結果          |
|----------|-------------|
| A(OE/OD) | 0.96        |
| B(GE/OD) | 0.96        |
| C(OE/GD) | 0.97        |
| D(GE/GD) | <b>0.98</b> |

### 4.4 評価結果

図 5 の c, d で示される評価結果は真値と判別結果画像を 4.3 節で述べた評価手法による評価結果を色分けした図である。図中における画素の色の意味を以下に示す。

表 3 評価画像の説明

|                    |                            |
|--------------------|----------------------------|
| 白 : true positive  | 道路である領域を道路と判別した画素          |
| 黒 : true negative  | 道路でない領域を道路でないと判別した画素       |
| 緑 : false positive | 道路でない領域を道路として判断した画素 (誤検出)  |
| 赤 : false negative | 道路である領域を道路でないと判別した画素 (未検出) |

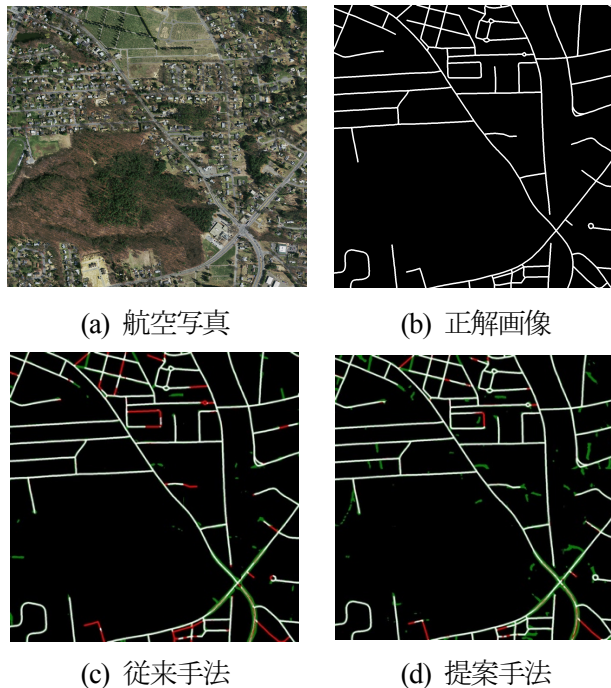


図5 評価画像

## 5 結論

エンコーダとデコーダの各畳み込み層を GCN にしたモデルが最も高い評価を示した. 図5に示す提案手法と従来手法の評価画像を比較すると, 提案手法のほうが多く道路領域を認識することができていることが分かる. GCN をモデルに追加した場合の傾向として, 評価画像において緑で描画されている過認識領域が増加することが分かった. さらに, それらは人間による目視で確認したところ道路として認識することができる場合が多く見られた. このことから, 出力結果を正解画像だけ行った評価は精度として充分でないと言えることが分かった.

## 6 おわりに

本研究では, 航空画像から自動的に道路領域を抽出することを目的とした. 手法として, 再帰構造を用いた FCN に GCN を用いることで, 従来手法よりも高い精度で道路領域の認識を可能とした. 今後の課題として, 今回実験で行うことができなかった GCN 部分のカーネルサイズの拡大が道路領域の認識に及ぼす結果の検証が挙げられる.

## 参考文献

- [1] 中村良介, 他, “地球観測衛星画像上の地物認識”, JSAI, 2016.
- [2] 高橋 宏輝ら他, “再帰構造を用いた全層畳み込みニューラルネットによる航空写真のセグメンテーション”, 第20回 画像認識・理解シンポジウム, 2017.
- [3] Chao Peng, Xiangyu Zhang, Gang Yu, Guiming Luo, Jian Sun. "Large Kernel Matters – Improve Semantic Segmentation by Global Convolutional Network". In *arXiv*, Vol. abs/1703.02719, 2017.
- [4] Matthew D. Zeiler. ADADELTA: an adaptive learning rate method. In *arXiv*, Vol. abs/1212.5701, 2012.
- [5] Volodymyr Mnih. "Machine Learning for Aerial Image Labeling". PhD thesis, University of Toronto, 2013.
- [6] Heipke C., Mayer H., Wiedemann C., and Jamet O., "Evaluation of Automatic Road Extraction, "International Archives of Photogrammetry and Remote Sensing, Vol. 32, 3-2W3, pp. 47-56, 1997.