

## Faster R-CNN を用いた空撮画像からの車両検出

杉本 将来† 谷口 行信† 早瀬 和也††

†東京理科大学 ††日本電信電話株式会社 NTT メディアインテリジェンス研究所

### 1 はじめに

空撮画像とは、飛行機、衛星等に搭載されたカメラから地上を撮影した写真のことである。近年、小型の地球観測衛星が大量に打ち上げられるようになり、地球上のほとんどの地点において、日々の空撮画像が取得できる。空撮画像の活用事例として、空撮画像からの車両検出による交通量調査、安全目的のための車追跡、駐車場の車のカウントが挙げられる。現在、これらの作業は、人の目視でおこなわれており、人的コストの削減が大きな課題となっている。そのため、画像認識技術を用いて空撮画像からの車両を自動検出・カウントする手法が提案されている [1]。しかし、これらの手法は、天候や日照条件によって精度に差がでたり、背景を車両として誤検出することが多く、実用化には至っていない。本研究では深層学習を用いた物体検出手法を空撮画像に適用することで車両検出の精度向上を目指す。

### 2 従来手法

空撮画像からの地物検出の従来手法として、畳み込みニューラルネットワーク (Convolutional Neural Network, CNN) を用いた手法 [1] がある。本研究で対象とする、航空画像の場合、車両の大きさは約  $30 \times 30$  [px] と非常に小さいため、畳み込みの過程で小さな車両の特徴が失われたり、隣り合う複数の車両の特徴が区別できなくなってしまう問題が存在する。また、車両と類似した地物 (家の屋根、倉庫など) を車両として誤検出する問題がある。

### 3 提案手法

前述した二つの問題を解決するために、本研究では、CNN に基づく物体検出手法の一つである Faster R-CNN [2] を車両検出に適用する。提案手法の特徴は、(1) 訓練データの中の誤検出した背景領域に“背景”というラベルを付与して訓練データに追加することで、車両と類似した地物の誤検出を減らすこと、また、(2) 小さく写る車両

の特徴を捉えるために、ResNet\_skip とよぶネットワークを採用したことにある。

#### 3.1 Faster R-CNN の構造

Faster R-CNN は、畳み込みネットワークから生成された特徴マップを入力として Region Proposal Network (以下、RPN) とよばれる CNN を用いて物体候補領域を抽出する。ベースネットワークに ResNet [3] を用いる。抽出された、物体候補領域の特徴マップを入力として検出ネットワークにより物体の種類を識別する。本研究では、空撮画像に含まれる車両の外接矩形を付与した独自データセットを作成し、Faster R-CNN の全ての層の重みを再学習した。図 1 に Faster R-CNN を用いて検出された車両と、車両として識別され誤検出された背景の領域を示す。このように、空撮画像内には車両と類似する地物が数多く存在していることがわかる。

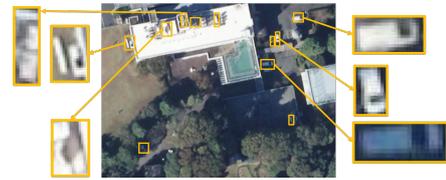


図 1: FasterR-CNN を用いた検出結果 (左:背景, 右:車両)

#### 3.2 背景ラベルの学習

Faster R-CNN に背景ラベルを学習させる処理を図 2 に示す。(i) 車両の学習: 上述したデータセットを用いて Faster R-CNN の学習を行う。(ii) 背景ラベルの生成: 学習済み Faster R-CNN を用いて訓練データセットに対して車両検出を行う。車両として誤検出された背景領域に“背景”というラベルを付与し、元のデータセットに追加する。(iii) 車両と背景の学習: (ii) で生成した新しいデータセットを元に Faster R-CNN をランダムに初期化して学習させる。誤検出しやすい領域の特徴を学習させることによって誤検出を減少させる。

#### 3.3 ResNet\_skip

図 3 に示すように、提案する ResNet\_skip ネットワークは畳み込みブロック 5 と畳み込みブロック 4 の特徴マップを足し合わせ検出ネットワークに入力する。畳み込みブロック 4 には、畳み込みブロック 5 によって複数の車両が一つになる以前の、もしくは、消失する以

Vehicle Detection in Aerial Images with Faster R-CNN

†Masaki Sugimoto, Tokyo University of Science

†Yukinobu Taniguchi, Tokyo University of Science

††Kazuya Hayase, NTT Media Intelligence Laboratories, NTT Corporation

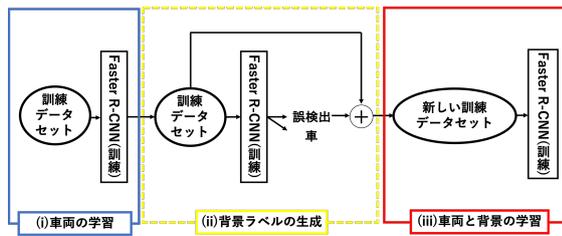


図 2: 背景ラベルの学習

前の特徴を保持しているからである。具体的には、畳み込みブロック 5 を畳み込みブロック 4 の幅高さに合うように拡大した後に、畳み込みブロック 4 に足し合わせた。

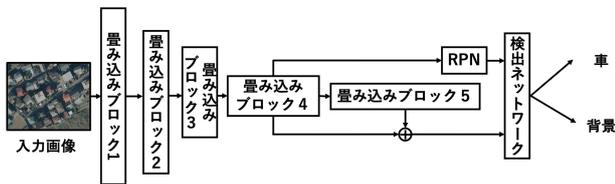


図 3: ResNet\_skip の構造

## 4 実験方法

提案手法による車両検出の精度を評価するために二つの実験を行なった。

### 4.1 実験条件

実験では、2016、2017 年に撮影された同地点の空撮画像（画像提供：NTT 空間情報株式会社）を用いた、 $8000 \times 6000$ [px] の元画像を  $400 \times 300$ [px] に分割した後に、Faster R-CNN の入力サイズである  $600 \times 600$ [px] に近くなるように、 $800 \times 600$ [px] に拡大した。正解ラベルとして車両を囲む外接矩形を目視で付与した。2016 年に撮影された画像をテスト用に 100 枚、2017 年に撮影された同地点の画像を学習用に 764 枚用意した。学習用に含まれる車両の総数は 11347 台になった。

### 4.2 評価方法

本実験では、“背景ラベルの学習”と“ResNet\_skip”の有無を切り替えた 4 通りの方法で、車両検出の精度を測定した：(1)ResNet, (2)ResNet+背景, (3)ResNet\_skip, (4)提案手法 (ResNet\_skip+背景)。提案手法により出力される矩形領域 B が正解か否かを以下のように判定した。矩形領域 B と重なりを持つ正解矩形領域 A とするとき、 $\text{IoU}(A, B) \geq 0.3$  のとき正しく車両を検出したと判定する。ここで、 $\text{IoU}(A, B)$  は

$$\text{IoU}(A, B) = \frac{A \cap B}{A \cup B}$$

と定義する。精度評価には、

$$\begin{aligned} \text{再現率} &= \frac{\text{正しく検出された車両の数}}{\text{正解の車両の総数}}, \\ \text{適合率} &= \frac{\text{正しく検出された車両の数}}{\text{車両として検出された領域の数}}, \\ \text{F 値} &= \frac{2 \times \text{再現率} \times \text{適合率}}{\text{再現率} + \text{適合率}}, \end{aligned}$$

を用いた。

## 5 実験結果

精度評価の結果を表 1 に示す。“(2)ResNet + 背景”は“(1)ResNet”と比較して再現率が 5.6% 低下したが、適合率が 20.4% 向上し、F 値は 4.1% 向上した。背景ラベルの学習によって、誤検出を低減する効果が示された。“(3)ResNet\_skip”と“(1)ResNet”を比較すると、適合率と再現率がとも低下したが、背景ラベル学習と組み合わせた“(4)提案手法”は再現率と適合率が共に向上した。

表 1: 提案手法による検出精度の変化

	(1)ResNet	(2)ResNet+背景	(3)ResNet_skip	(4)提案手法
再現率	<b>0.641</b>	0.587	0.569	0.620
適合率	0.705	<b>0.909</b>	0.696	0.876
F 値	0.672	0.713	0.626	<b>0.726</b>

## 6 おわりに

本研究では、空撮画像を用いた車両検出の自動化と精度向上を目的とし、背景ラベルの学習と ResNet 内の特徴マップを足し合わせることで特徴の欠失を抑える手法を提案した。その結果、両手法を用いない場合と比べて F 値が 5.4% 向上することを示した。今後は、Faster R-CNN に用いられるネットワークのさらなる改善により、より多くの車両の特徴を捉え、さらなる精度向上を目指す。

## 参考文献

- [1] J. Carlet and B. Abayowa, “Fast Vehicle Detection in Aerial Imagery,” *arXiv:1709.08666*, 2017.
- [2] S. Ren, K. He, R. Girshick and J. Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” *Adv in Neural Inf Process Syst*, pp. 91–99, 2015.
- [3] K. He, X. Zhang, S. Ren and J. Sun, “Deep Residual Learning for Image Recognition,” *arXiv:1512.03385*, 2015.