

畳み込みニューラルネットワークを用いた 衛星画像における実用的高精度車両検出

Accurate Vehicle Detection using Convolutional Neural Network in Satellite Images for Practical Applications

学籍番号 47-166734
氏名 古賀 洋平 (Koga, Yohei)
指導教員 柴崎 亮介 教授

1. はじめに

1. 1. 背景

近年、衛星画像の高解像度化及び機械学習、特に深層学習の進歩により、衛星画像が様々な実用途に用いられつつある。その一つが、地表面の経済活動量推定である。衛星画像はパブリックなデータのため、直接のデータが得られない（もしくは取得にコストがかかる）事象や現地調査が難しい場所においても利用可能である。これを利用して地域の状況を正確に把握することで、効果的な意思決定等に資することができる。既往研究では、時系列衛星画像の分析による都市の広がりの変化の把握（宮崎ら（2014）[1]）や、夜間光画像、日中画像およびセンサデータを用いた地域の貧困度の推定（Jeanら（2016）[2]）がある。また民間では、米のベンチャー企業 Orbital Insight [3] が有名小売チェーンにおける駐車場の車両数を全国的にカウントすることで、売り上げ指数の予測を行っている。

こういった推定を行うための特徴量として、車両は経済的活動とよく関連しており、有望である。また、交通量推定等のアプリケーションも多いため、高精度な車両検出は非常に重要である、近年では深層学習の

進歩により高精度が達成されている。しかし、実用的には、トレーニングデータの地域と関心地域が異なり、関心地域のトレーニングデータが存在しない場合が多い。そのような場合、性能は大幅に低下する。コストの観点も考え、効率的に高精度を達成可能な方法を検討する必要がある。

1. 2. 研究の目的

本論文では、衛星画像における実用的な高精度車両検出に向け、深層学習の手法を検討する。精度の検証にあたっては、トレーニングデータの地域と異なる関心地域を想定し、そこでの性能について主に検証する。そして、性能低下を改善する手段として、ドメインアダプテーション (DA) の手法を用いる。さらに、DAに加え、少量の教師データを用いたとき、どの程度まで性能が向上できるか検討する。最後に、車両検出の実用例として商業施設における駐車場車両数の推定を行う。

2. 手法

2. 1. 深層学習を用いた物体検出手法

深層学習は脳を模した構造である。その中でも、畳み込みニューラルネットワーク (Convolutional Neural Network; CNN) と呼ばれる手法が深層学習発展の火付け役とな

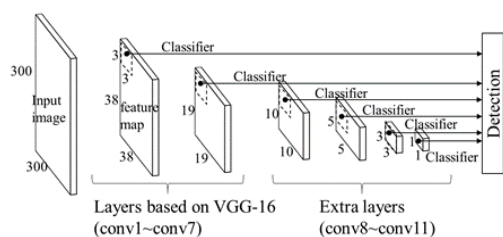


図 1: SSD の概要.

った. 深層学習における物体検出は, 主にこの CNN を用いて実現されている.

2. 2. 車両検出手法

車両検出には, 高速かつ高精度な手法である Single Shot MultiBox Detector (SSD) [4] (図 1) と呼ばれる手法を採用し, 小さなオブジェクト (車両) を検出しやすいよう, チューニングを行った.

2. 3. ドメインアダプテーション手法

教師データが得られる場所をソースドメイン (SD), 得られない (少ない) 場所をターゲットドメイン (TD) と呼ぶ. DA 手法は, 教師無しで TD, すなわち関心地域における性能を向上させる手法である. 本論文では, Deep-CORAL [5] 及び Adversarial Discriminative Domain Adaptation [6] をベースとした CORAL DA 及び Adversarial DA を設計し, 性能を検証した. CORAL DA は, 各ドメインのデータそれぞれの分散共分散行列を求め, その 2 乗距離を小さくすることで, 統計的な特徴が似た共通の特徴を学習する手法である. Adversarial DA は, 敵対的学習と呼ばれ, ドメイン間の差異を区別するネットワークと, それに区別されないような特徴抽出器のトレーニングを交互に行うことで, ドメイン間共通の特徴を学習する手法である (図 2).

3. 実験

3. 1. 使用するデータ

SD には, 車両検出用データセットである,

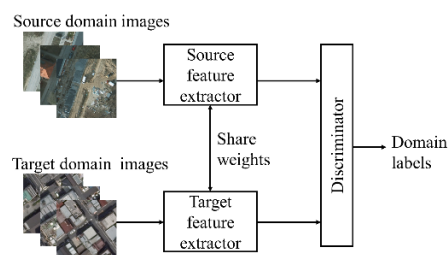


図 2: Adversarial DA の概要.

Cars Overhead with Context (COWC) データセット [7], TD には NTT 空間情報 [8] 提供の東京の空撮画像を用いた. 本論文のターゲットは衛星画像であるため, 共に解像度を 0.3m/pixel にダウンサンプリングして用いた.

3. 2. テスト方法及び性能評価基準

車両検出は SSD の検出方法に準じ, 主な性能評価基準に Average Precision (AP) 及び F1 値またはその平均を用いた.

3. 3. ソースドメインのみ用いた車両検出器の性能

SD のデータのみ用いて SSD のトレーニングを行った. SD のテストでは AP 及び F1 の平均は 84.1% に達したが, TD のテストにおいては 66.3% と大幅に低下した.

3. 4. ドメインアダプテーションの適用

CORAL DA 及び Adversarial DA を適用した結果, AP 及び F1 の平均は 76.8% 及び 75.9% に回復した.

3. 5. 性能評価

CORAL DA 及び Adversarial DA により, 性能低下の半分以上を回復させることができた (図 3). 本実験においては, CORAL DA のほうが若干良い結果となった. DA 無しでは検出されていなかった車両が, DA 適用によって多く検出されたことが, 性能向上の主な要因となった. また, 具体的な車両検出結果画像を比較したところ, 特に密集した

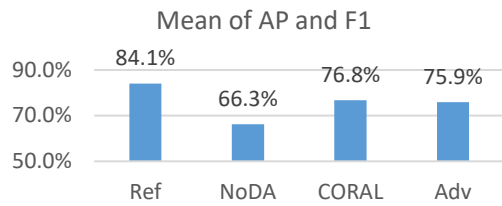


図 3: DA による精度改善.

地域で性能が大きく向上しており、TD の画像特有の密集した特徴が性能低下を引き起こしていたのが、DA によって改善されたことを示している。

3. 6. 少量のラベル付きデータを用いた精度向上

さらなる精度向上のため、TD の少量のラベル付きデータを用いて性能向上を試みた。用いたのは、SD のトレーニングデータの約 1/12 (SD は 4 地域のデータから成るので、1 地域あたりでは約 1/3) である。結果、AP 及び F1 の平均は 81.0% まで向上した。データの量を変えて実験したところ、改善幅はデータの増加に従ってゆるやかになっており、DA と少量のラベル付きデータで十分な精度に達している可能性がある。ただ、これを確認するにはさらにデータを増やして実験する必要があり、今後の課題である。最終的に、Data Augmentation (データを人工的に加工することで増やし、精度を向上させる手法) を用いることで、AP 及び F1 の平均は 81.9% になった。

4. 車両検出器を用いた駐車場の車両数推定

最後に、車両検出器の実用例として、商業施設の駐車車両数推定を行った。具体的には、ある時間における立体駐車場の屋上の車両数を、画像から車両検出によって計測し、そこから 1 日の駐車数推定を行った。

推定は 1 時間おきの車両数それぞれを変数とする単回帰モデルを構築し、そのいずれかを用いた。トレーニングデータとして、土日 4 日間屋上の車両数を 1 時間おきに実際に計測したデータと、計測日の総駐車車両数のデータを用いた。(匿名性のため、データ提供元及び施設名は伏し、車両数は指標化する。) 決定係数は 15 時及び 16 時のとき最も高くなり、予測の性能が高くなった。これは、実地調査において 14~16 時台に最も駐車車両数が多かったこととも一致する。

予測用の画像は、NTT 空間情報提供の空撮画像をダウンサンプリングして用いた。検出数は実際の車両数と等しくなったが、内約を見ると誤検出等も含んでいる。用いた画像は DA で用いた画像の特徴とまた微妙に異なっており、これが性能低下を招いた可能性がある。

この結果を用いて車両数推定を行ったところ、比較的誤差の少ない結果が得られた。ただし、画像の撮影時間は 12 時に近く、モデルの決定係数は低いため、信頼性のある結果ではない。より精度の高い予測のためには、トレーニングサンプルを増やしてモデルの精度を高めることや、予測に適した時間帯 (本実験では、15~16 時) で予測を行うことが考えられる。また、本実験では屋上の車両数から推定を行ったが、屋上の駐車数はそもそも駐車車両が多いときはみ出した値であり、単純な回帰モデルで全体の駐車車両数を推定するのは難しい。まず、より簡単な地上駐車場の推定を試みるのが考えられる。

5. 結論

トレーニングデータ (SD) と異なる関心地域 (TD) で車両検出を行った場合大幅に性能

が低下したが、DA の手法によってその半分以上を回復させることができた。DA は教師無し的手法であり、コストが低くて済むので、実用的な高精度車両検出の達成に非常に有用と思われる。加えて、少量のラベル付きデータを組み合わせて用いることで、更なる精度向上を達成した。また、車両検出の実用例として商業施設の駐車車両数推定を行った。小規模な実験だが、先鞭をつけることができたと思われる。

今後の課題としては、まず DA と少量の教師データで十分な精度に達し得たのか、確認する必要がある。また、今回用いた DA 手法は比較的単純なもののため、より少ないコストで更なる高精度が達成できるような手法の検討の余地がある。最後に、衛星画像の分析による経済活動推定は、投資や政策といった意思決定に非常に有用と思われる。今後、より多くのデータを用いて検証の精度を高める他、車両検出のみならず具体的な売り上げ推定を行う等、さらに研究を深めていくことが必要である。

[1] H. Miyazaki, X. Shao, K. Iwao, R. Shibasaki, “Development of a Global Built-Up Area Map Using ASTER Satellite Images and Existing GIS Data,” in *Global Urban Monitoring and Assessment through Earth Observation*, London: CRC Press, 2014, p. 121.

[2] N. Jean, M. Burke, M. Xie, W. M. Davis, D. B. Lobell, S. Ermon, “Combining satellite imagery and machine learning to predict poverty,” in *Science* 2016, Vol. 353, Issue 6301, pp. 790-794.

[3] Orbital Insight. <https://orbitalinsight.com/> (accessed on 17 January 2018)

[4] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.Y. Fu, A.C. Berg, “SSD: Single Shot MultiBox Detector,” *Computer Vision – ECCV 2016*. ECCV 2016. Lecture Notes in Computer Science, Amsterdam, The Netherlands, 8-16 October 2016; Volume 9905, page 21-37.

[5] B. Sun and K. Saenko, “Deep CORAL: Correlation Alignment for Deep Domain Adaptation,” *ECCV 2016 Proceedings Part III*, pp 443-450, 2016.

[6] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, “Adversarial discriminative domain adaptation,” in *NIPS Workshop on Adversarial Training*, (WAT), 2016.

[7] T.N. Mundhenk, G. Konjevod, W.A. Sakla, K. Boakye, “A Large Contextual Dataset for Classification, Detection and Counting of Cars with Deep Learning,” in *Lecture Notes in Computer Science, Proceedings of the ECCV 2016: Computer Vision—ECCV 2016*, Amsterdam, The Netherlands, 8-16 October 2016; Volume 9907, page 785-800.

[8] NTT 空間情報 <http://www.ntt-geospace.co.jp/> (accessed on 17 January 2018)