

## 深層学習を用いた屋外設置物を対象とした特定物体検出手法 Specific Object Detection for Outdoor Objects using Deep Learning

重富 達哉<sup>†</sup> 赤嶺 有平<sup>†</sup> 根路銘 もえ子<sup>‡</sup>

Tatsuya Shigetomi Yuhei Akamine Moeko Nerome

### 1. はじめに

AR を用いて情報提示を行う場合、対象の画面上の位置を取得する必要があり、屋外の物体を対象とした場合は、極端な照明変動に対してロバストな特定物体検出手法が求められる。

特定物体検出では主に SIFT 特徴量を用いたものが一般的である。しかし屋外環境下においては、より顕著となる天候要因による照明変動、隠れといった画像の変動などの「見えの変化」により、複雑な形状をもつ物体など日照の影響を受けやすい物体を対象とした場合に、局所特徴量が変化しマッチングの精度が低下するといった問題がある。

深層学習を用いた特定物体検出においても、屋外環境下での見えの変化による影響は無視することができない問題である。理論上では十分なデータがあれば見えの変化に対しても頑健な識別モデルを作成することは可能であるが、特定物体を対象にしており、かつそれが屋外に存在してある場合、見えの変化による分布の広がりに対して十分な学習サンプルを揃えることが困難となる。

そこで本研究では、学習に十分なデータを収集することが困難な場合に有効な特定物体検出手法の開発を目指す。

### 2. 提案手法

提案手法では、物体検出部とインスタンス識別部の二段構成で処理を行う。概略図を図 1 に示す。

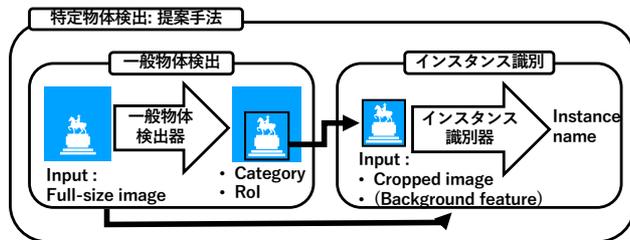


図 1 提案手法 概略図

まず、前段である物体検出部ではフルサイズの画像を入力とし、一般物体検出を行い、対象物体の RoI(Region of Interest)とカテゴリを推定する。一般物体検出手法は、たとえば YOLO[1]のようなハイパフォーマンスなモデルを使用することを想定している。

推定された物体が対象カテゴリに属していれば後段のインスタンス識別が行われる。この際、推定された RoI の範囲を切り取り、対象物体と背景を分離し、それぞれをインスタンス識別の入力とする。これは、背景情報がインスタンスの識別を行う上で重要な特徴となり、特にインスタンス間で形状の差異が見られないパターンにおいて有効に

働くという仮定に拠るものである。この背景情報と対象物体のクロップ画像を入力とするインスタンス識別器を CBO-Net(Combination of Background Image and Object Image - Network)とする。CBO-Net の構造を図 2 に示す。

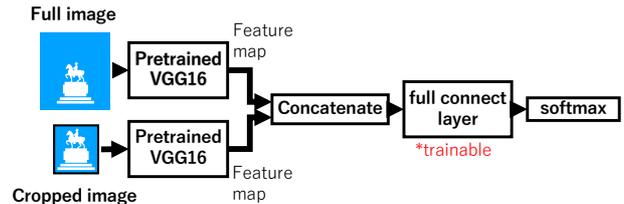


図 2 CBO-Net

次節の 3. 検証実験にて、背景情報を用いることの特定物体検出への貢献について検証を行う。

### 3. 検証実験

実験では、主に背景情報を用いることの特定物体検出への貢献と、最低限のデータセットを用いた場合の屋外での見えの変化への頑健性について検証を行う。

RoI 推定の精度については、一般物体検出に依存しているため、今回は評価対象外とする。

インスタンス識別部では、対象物体のみ切り取られたクロップ画像とフルサイズの背景画像の両方を入力とする CBO-Net と、クロップ画像のみを入力とした VGG16[2]を用いて検証を行う。

#### 3.1 データセット

検証用に 2 つの対象物のデータセットを用意した。

ひとつは道路標識であり、インスタンス間にまったく形状の差異がないが、別の場所に設置してあるものを用いた。形状がシンプルかつインスタンス間で一定であることから、日照変化の影響を受けにくいと、背景情報がどれ程インスタンス識別の精度に貢献できているかを検証することが可能であると想定した。

もうひとつの対象として陶器の像を用いた。これは比較的複雑な形状をしており、表面が滑らかで光を反射しやすいため、日照変化による細かい陰影の変化や光の反射による見えの変化が発生しやすい対象物である。このデータセットから、主に屋外での見えの変化への頑健性を検証することが可能であると想定した。

これら 2 つの対象に対して複数のインスタンスを用意し、さまざまな日照条件下で撮影を行う。学習に最低限のデータセットを収集している状況を想定し、撮影した画像から、特定の日照条件で撮影された画像群を学習に用い、他の日照条件で撮影された画像群を学習時のバリデーションに用いる。これらの学習データ及びバリデーションデータセットと異なる日照条件で撮影されたすべての画像群をモデル評価に用いる。

<sup>†</sup> 琉球大学 工学部 Ryukyu University, Fac. of Engineering

<sup>‡</sup> 沖縄国際大学 経済学部 Okinawa International University, Fac. of economics

教師データとして用いるクロップ画像は作業を半自動にするためにトラッキングツールを用いて用意した。

また学習に用いるデータセットは random contrast, random brightness, random rotate, gaussian noise といったデータオーグメンテーションを行い、元のデータセットサイズの 5 倍に拡張した。データオーグメンテーションはすべて標準正規乱数に従い、CBO-Net に入力する背景画像とクロップ画像のセットは同じオーグメンテーションを行う。

最終的なデータセットの画像数は、道路標識が[train: 599\*5, valid: 613, test: 3314], 陶器の像が[train: 334\*5, valid: 200, test: 435]となった。

データセットの一部をそれぞれ図 3 に示す。

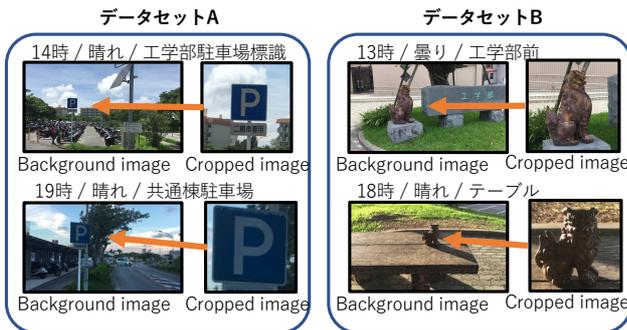


図 3 データセットのサンプル

### 3.2 学習

2 つのデータセットを、それぞれ 2 つのインスタンス識別器で学習させるため、計 4 つの学習モデルを作成する。それぞれの学習時のバリデーションの値が一番高かった時点での学習モデルを用いてモデルの評価を行う。

## 4. 実験結果

以下の図 4 に各学習モデルの平均識別率を示す。また、図 5, 6 にそれぞれ道路標識と陶器の像の日照状況毎の識別率を示す。

	標識(二輪専用駐車場)	陶器の像(シーサー)
CBO-Net	0.64	0.57
VGG16	0.42	0.53

図 4 モデル評価の平均識別率

condition	晴れ	夕雲	夕雲
recode time	06/21/14時頃	06/23/19時頃	06/24/19時頃
CBO-Net	0.51	0.75	0.68
VGG16	0.61	0.32	0.32

図 5 日照状況毎の識別率(道路標識)

condition	晴れ	くもり	くもり	くもり	くもり	晴れ
recode time	06/21/14	06/22/16	06/22/18	06/23/19	06/24/19	06/28/11
CBO-Net	0.67	0.32	0.51	0.60	0.59	0.62
VGG16	0.66	0.89	0.70	0.79	0.72	0.26

図 6 日照状況毎の識別率(陶器の像)

陶器の像を対象としたデータセットでは、図 6 中の 06/28/11 時頃に撮影したサンプルのみ、極端な照明変化が発生している。

図 4 から、どちらの対象物でも背景情報を用いる CBO-Net が高い平均識別率を示した。また、CBO-Net では陶器の像データセットよりも標識データセットを対象にした場

合の方が高い識別率を示しているのに対し、VGG16では標識データセットよりも陶器の像データセットを対象にした場合の方が高い識別率を示している。

図 5 より、照明変化による極端な見えの変化が発生している状況では VGG16 の識別率が低下しているのに対して、CBO-Net では識別率の低下は見られない。比較的見えの変化が小さい晴れの日照状況下(06/21/14 時頃)においては、CBO-Net よりも VGG16 の方が識別率は高い結果を示している。

図 6 より、曇りの日照状況下ではいずれも CBO-Net よりも VGG16 の方が識別率は高いが、晴れの日照状況下においては CBO-Net の方が識別率は高い結果を示している。特に、06/28/11 時頃に撮影した極端な日照変化が発生しているサンプルにおいては、VGG16 では極端な識別率の低下が見られるが、CBO-Net では識別率の低下は見られない。

## 5. 考察

背景情報がインスタンスの識別を行う上で重要な特徴となり、特にインスタンス間で形状の差異が見られないパターンにおいて有効に働くという仮定は、実験結果の図 4 より、CBO-Net の方が比較的識別率が高く、インスタンス間で形状の差が小さい標識データセットを対象とした場合大きく識別率の差がみられることから、概ね正しいと考えられる。

また、背景情報を用いる CBO-Net では、最低限の学習データセットを用いたにもかかわらず、極端な照明変動による見えの変化が発生している状況においても識別率の低下は見られない。このことより、背景と対象物体の画像をそれぞれ有効に扱うことによって、屋外の見えの変化による分布の広がりに対して不十分な学習データでも頑健なインスタンス識別モデルを作成することが可能であると考えられる。

## 6. おわりに

本研究では、学習に十分なデータを収集することが困難な場合に有効な特定物体検出手法の提案と検証を行った。実験結果より、背景情報を有効に扱うことによって、屋外の極端な見えの変化による分布の広がりに対して不十分な学習データでも、頑健なインスタンス識別モデルを作成できる可能性を示した。

### 参考文献

- [1] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi, "You Only Look Once: Unified, Real-Time Object Detection", <https://arxiv.org/abs/1506.02640> (2016)
- [2] Karen Simonyan, Andrew Zisserman "VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION", <https://arxiv.org/pdf/1409.1556.pdf> (2015)