

## 再学習不要で新規物体を高精度に検出できる 特徴ベクトル登録型 Few-shot 物体検出技術

Few-Shot Object Detection Technology Using Feature Vector Registration for Highly Accurate Detection of New Objects without Retraining

小林 大祐 KOBAYASHI Daisuke

深層ニューラルネットワーク(DNN)などを用いて画像から特定の物体を検出する技術は、生産性の向上や自動化・省力化のために使用され、重要性が高まっている。しかし、導入時に存在しなかった新規物体を検出するには、DNNに再学習させる必要がある。Few-shot物体検出技術は、数枚の画像があれば再学習可能だが、高速なハードウェアを使用しても数時間以上掛かっていた。

そこで東芝は、新規物体の画像から抽出した特徴ベクトルを活用することで、数枚の新規物体の画像と正解情報があれば再学習を不要にすることが可能な特徴ベクトル登録型 Few-shot 物体検出技術を開発した。更に、画像内の学習対象以外の物体の形状を自動的に学習する独自の自己教師あり学習方式を開発して適用し、新規物体の検出精度を従来手法より向上させた。

Technologies for detecting specific objects in images using deep neural networks (DNNs) and other techniques are becoming increasingly important to improve productivity, promote automation, and save labor. However, retraining of a DNN is necessary in order to detect new objects that did not exist at the time of its introduction. The few-shot object detection technique makes retraining possible using only a few images, but this process takes at least several hours even when high-speed hardware is used.

Toshiba Corporation has developed a few-shot object detection technology using feature vector registration that requires only a few images of a new object together with the correct information, eliminating the need for retraining. We have also developed and applied an original self-supervised learning method that automatically learns the shapes of objects in an image other than those targeted for training, improving the accuracy of detection of new objects compared with conventional methods.

### 1. まえがき

近年、DNNなどにより画像から物体を検出する技術は、人物の行動や周辺環境の認識に必要不可欠になり、社会インフラにおける保守点検作業の効率化や、物流現場における業務効率化など、多種多様な分野で実利用が進んでいる。

DNNは導入時に、検出対象の画像と正解情報を用意して学習してから、検出を実施する。しかし、実際の環境では、新しい部品や製品を扱う場合など、未学習の新規物体も検出できるようにすることが求められる。新規物体の検出には通常、DNNの再学習を行うが、大量の画像と正解情報を用意する必要があることに加え、学習の時間が長く掛かるため、現場での活用は難しかった。

Few-shot物体検出技術<sup>(1)</sup>は、数枚の新規物体の画像と正解情報だけで、DNNを再学習させて新規物体を検出できる技術である。しかし、GPU (Graphics Processing Unit) などの高速なハードウェアを使用しても、学習に数時間以上掛かっていた。また、DNNの再学習をしなくても新

規物体の検出が可能な手法として、新規物体の画像を検出に用いる Attention-RPN<sup>(2)</sup>があるが、検出精度が十分とは言えなかった。

そこで東芝は、新規物体の画像から抽出した特徴ベクトルを活用することで、新規物体の数枚の画像と正解情報を用意するだけで、DNNの再学習なしで新規物体の検出が可能な特徴ベクトル登録型 Few-shot 物体検出技術を開発した<sup>(3)</sup>。また、独自の自己教師あり学習方式を開発して導入することで、新規物体の検出精度の向上を実現した。

ここでは、特徴ベクトル登録型 Few-shot 物体検出技術の概要及びネットワーク構造と、公開データセットを用いた検出精度の評価実験の結果について述べる。

### 2. 特徴ベクトル登録型 Few-shot 物体検出技術の概要

特徴ベクトル登録型 Few-shot 物体検出技術のキー技術として、特徴ベクトルの活用と自己教師あり学習方式について述べる。

#### 2.1 特徴ベクトルの活用

特徴ベクトル登録型 Few-shot 物体検出技術では、事前

学習で同じ種類の物体の特徴ベクトルを平均化したものをプロトタイプ(検出すべき物体を代表する特徴ベクトル)としてメモリーに登録し、推論時にプロトタイプを用いて、物体の分類及び物体の矩形位置と大きさを求める(図1)。新規物体を検出するときは、事前学習したDNNから出力される新規物体の特徴ベクトルをプロトタイプとしてメモリーに登録する。これにより、再学習しなくても新規物体に対応できる。

## 2.2 自己教師あり学習方式の導入

一連の仕組みが効果的に動作するためには、特徴空間の中で、同じ種類の物体の特徴ベクトル同士は近くに位置し、異なる種類の物体の特徴ベクトルは遠くに位置していることが必要である。図2に示すとおり、教師あり学習では、学習用データ内で正解情報が付与されていない物体は背景とみなされるため、特徴空間の学習に寄与せず、特徴空間内で適切な位置を占めることができない。そこで、正解情報

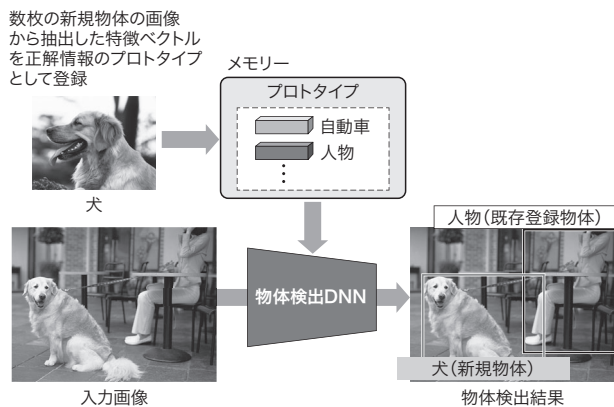


図1. 特徴ベクトル登録型Few-shot物体検出技術の概要

数枚の新規物体の画像から抽出した特徴ベクトルをプロトタイプとして正解情報とともに登録するだけで、再学習することなく新規物体を検出する。

Overview of few-shot object detection technology using feature vector registration

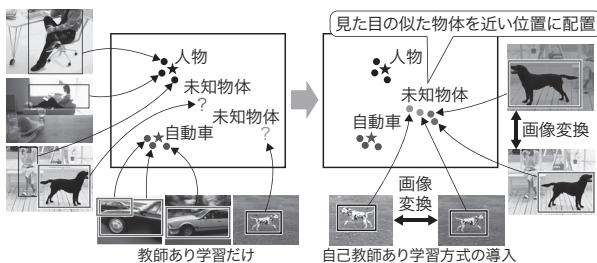


図2. 自己教師あり学習方式の導入による特徴ベクトルの配置の改善

自己教師あり学習方式では、正解情報がない画像の特徴ベクトルも抽出し、画像変換した同一物体の特徴ベクトルが近くに位置するように学習する。

Improvement of feature vector arrangement by introduction of self-supervised learning method

がない、背景として扱っていた物体らしい領域を含めて自動的に学習する、自己教師あり学習方式を開発して導入した。背景にある物体らしい領域から取得した特徴ベクトルと、同一領域の画像に、色の修正や、変形、切り抜きなどの画像変換を施した画像から取得した特徴ベクトルとが、近くに位置するように学習することで、新規物体の検出精度を向上させる。

## 3. 特徴ベクトル登録型 Few-shot 物体検出技術のネットワーク構造

推論に用いる、プロトタイプを用いるネットワーク構造と、事前学習に用いる自己教師あり学習方式を導入したネットワーク構造について述べる。

### 3.1 推論に用いるネットワーク構造

従来のFew-shot物体検出技術も利用する2ステージ方式の物体検出<sup>(4)</sup>に用いるネットワーク構造を基に、登録型の構造を取り入れた。2ステージ方式では、物体検出タスクを、画像中から矩形の物体候補領域(RoI: Region of Interest)を抽出するネットワーク(RPN: Region Proposal Network)、及びRoI内の物体の分類と矩形の位置・大きさの微調整の二つのステージに分けている。

特徴ベクトル登録型Few-shot物体検出の推論に用いるネットワーク構造を、図3に示す。2.1節に述べたとおり、事前学習で、特徴ベクトルを平均化しプロトタイプとしてメモリーに登録してある。推論時に画像が入力されると、バックボーンで特徴抽出し、RPNで画像からRoIを切り出して、特徴抽出ヘッドで特徴ベクトルに変換する。分類層では、特徴ベクトルと、物体ごとのプロトタイプのコサイン類似度に基づいて、それぞれの登録された物体である確率を推定し、分類する。

物体の分類に加え、RoIを微調整して最終的な矩形の位置・大きさを回帰する処理にもプロトタイプを活用する。矩形の位置・大きさ回帰層の前段に、プロトタイプに応じた回帰用フィルターを出力するフィルター生成層を組み込んでいる。動的にフィルターを変更することで、物体ごとに矩形の位置・大きさを適切に調整できる。

新規物体を検出する場合は、新規物体の特徴ベクトルを平均化し、プロトタイプとしてメモリー登録することで、同様に動作を実行する。

### 3.2 事前学習に用いるネットワーク構造

図4に、事前学習に用いるネットワーク構造を示す。

学習用データの画像の中で正解情報のある物体については、通常の2ステージ方式の物体検出で用いる損失関数に加えて、RoIの特徴ベクトルと対応するプロトタイプを近づけ

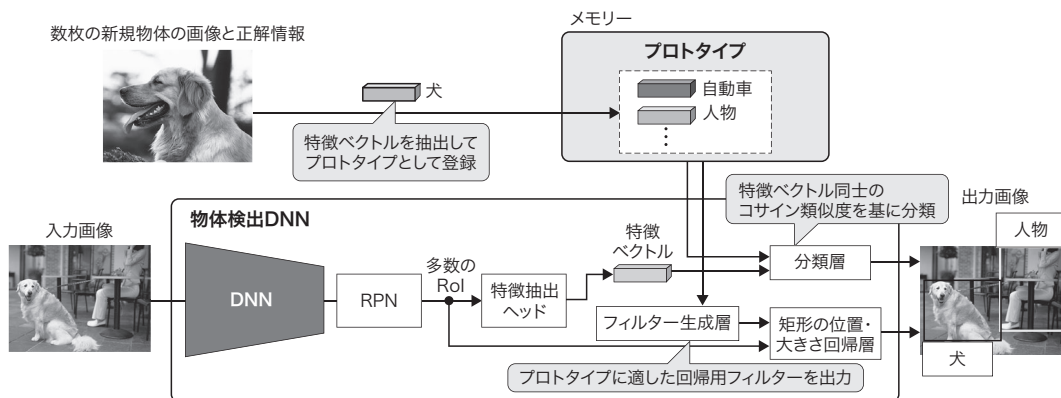


図3. 推論に用いるネットワーク構造

プロトタイプとしてメモリーに登録した新規物体の特徴ベクトルに基づいて、物体の分類と矩形の位置・大きさの回帰を実行する。

Network architecture for inference

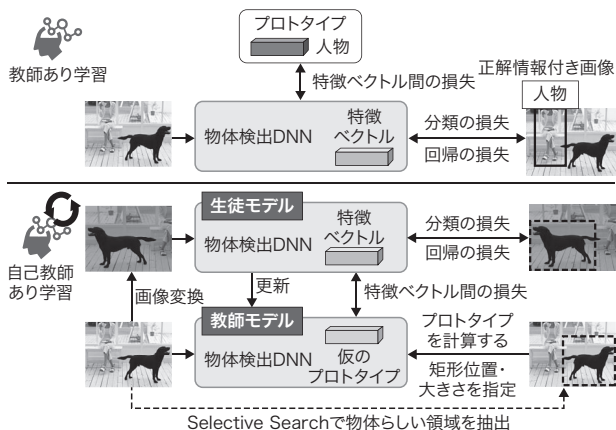


図4. 事前学習に用いるネットワーク構造

通常の教師あり学習に加え、正解情報がない画像も検出対象とする自己教師あり学習を実行する。

Network architecture for pretraining

る損失関数を用いて、両者が特徴空間内で近い位置に分布するように促す。

更に、新規物体の検出能力を改善するために、自己教師あり学習も同時に行う。自己教師あり学習では、正解情報が教示されていない部分に物体らしい領域があった場合に、その特徴ベクトルを仮のプロトタイプとして学習する。これを繰り返すことにより、正解情報がなくても、似たような画像部分が特徴空間内で近い場所に位置するようになる。

自己教師あり学習方式では、教師あり学習で使用するモデルを、生徒モデルと教師モデルに分けて学習する。生徒モデルと教師モデルは同じ構造であり、教師モデルは生徒モデルの学習に使う仮の正解矩形から仮のプロトタイプを作る役割を果たす。仮の正解矩形の生成には、色やテクス

チャーが似た領域を抽出する Selective Search<sup>(5)</sup>を用い、得られた領域からランダムに10か所を選択する。画像と仮の正解矩形を教師モデルに入力し、推定した特徴ベクトルをプロトタイプとみなして、生徒モデルの特徴ベクトルの学習に使用する。また、生徒モデルに入力する学習用データには、教師モデルに入力した画像に対して、画像変換を施したものを用いる。なお、教師モデルの学習対象となる重みのパラメータは、生徒モデルのパラメータを用いて指数平滑移動平均で更新する。

#### 4. 評価実験

評価実験には、Pascal VOCデータセット<sup>(6)</sup>を使用した。Pascal VOCデータセットの20種類の物体を、学習用に15種類、テスト用に5種類に分割し、選択する種類を変更した3セットを評価用データとして用意した。検出精度の評価には、評価指標のmAP (Mean Average Precision) @50を用いた。新規物体の画像を1枚、若しくは10枚使用した場合の検出精度を、特徴ベクトル登録型Few-shot物体検出技術と、1章で述べたAttention-RPN<sup>(2)</sup>と比較した。

新規物体の画像の登録だけを行った場合の検出精度の評価結果を図5(a)に示す。特徴ベクトル登録型Few-shot物体検出技術は、3セットの評価用データの全てで、Attention-RPNを大きく上回った。特に、評価用データ1では、画像を1枚登録した場合の検出精度が、Attention-RPNの21.2%に対して、特徴ベクトル登録型Few-shot物体検出技術は46.0%と、大きく改善した。

また、新規物体の画像を使用して、DNNを再学習した場合の検出精度を図5(b)に示す。再学習により、登録だけの場合と比べると精度は更に改善した。特徴ベクトル登録



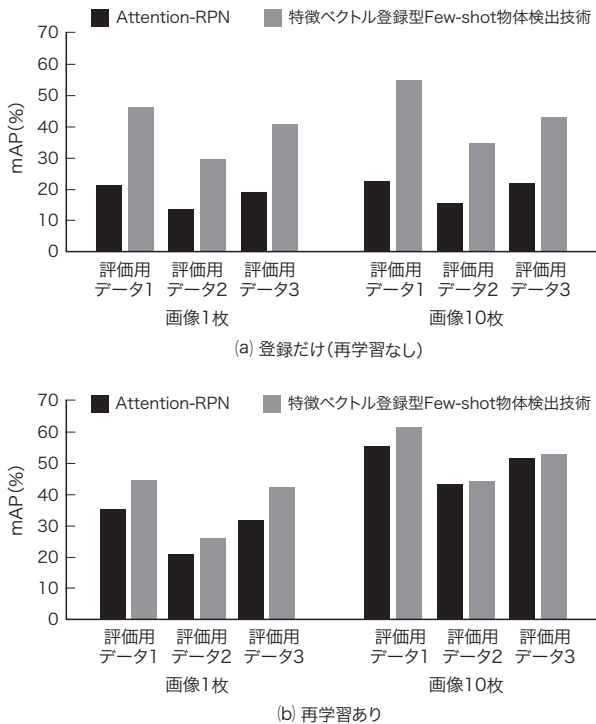


図5. 新規物体の検出精度の評価結果

特徴ベクトル登録型Few-shot物体検出技術では、Attention-RPNに比べて、どの条件でも検出精度が向上し、特に登録だけの場合に、大幅な検出精度改善を確認できた。

Results of evaluation of new object detection accuracy

型Few-shot物体検出技術は、いずれの評価用データにおいてもAttention-RPNを上回った。

最後に、新規物体の画像を1枚だけ登録した場合の検出結果の画像を、図6に示す。Attention-RPNでは推定したカテゴリーや矩形位置に誤りがあるが、特徴ベクトル登録型Few-shot物体検出技術では適切に検出できた。

## 5. あとがき

新規物体が映った数枚の画像と正解情報から抽出した特徴ベクトルを登録するだけで、DNNの再学習なしで新規物体を検出できる特徴ベクトル登録型Few-shot物体検出技術を開発した。自己教師あり学習方式を導入することで、新規物体の検出精度を高められることを確認した。

今後も、継続的に改良を加えて検出精度を向上させるとともに、早期の実用化を図る。

## 文 献

(1) Wang, X. et al. "Frustratingly Simple Few-Shot Object Detection". Proceedings of the 37th International Conference on Machine Learning. Montreal, 2020-07, PMLR. 2020, p.9919-9928.

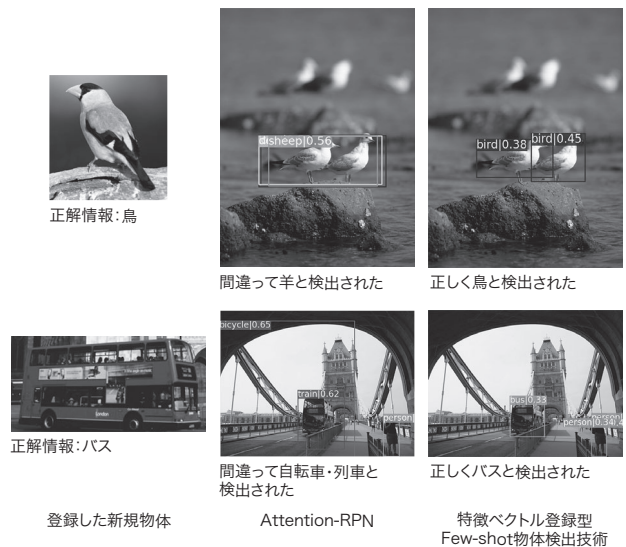


図6. 新規物体の検出結果の比較

検出対象と見た目の異なる画像を登録した場合も、正しい位置に正しいカテゴリーで検出できることを確認した。

Comparison of new object detection results

(2) Fan, Q. et al. "Few-Shot Object Detection with Attention-RPN and Multi-Relation Detector". Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR2020). Seattle, WA, 2020-06, IEEE. 2020, p.4013-4022.

(3) Kobayashi, D. "Self-supervised Prototype Conditional Few-Shot Object Detection". Image Analysis and Processing - ICIAP 2022: 21st International Conference. Lecce, Italy, 2022-05, Italian Association for Computer Vision, Pattern Recognition and Machine Learning (CVPL- ex-GIRPR). Springer, 2022, p.681-692.

(4) Ren, S. et al. "Faster R-CNN: Towards Real-time Object Detection with Region Proposal Networks". Proceedings of the 28th Conference on Neural Information Processing Systems (NeurIPS 2015). Montreal, Canada, 2015-12, NeurIPS. 2015, p.91-99.

(5) Uijlings, J. R. et al. Selective Search for Object Recognition. International Journal of Computer Vision. 2013, **104**, 2, p.154-171.

(6) Everingham, M. et al. The Pascal Visual Object Classes (VOC) Challenge. International Journal of Computer Vision. 2010, **88**, 2, p.303-338.



小林 大祐 KOBAYASHI Daisuke  
 研究開発センター  
 知能化システム研究所 メディア AI ラボラトリー  
 映像情報メディア学会会員  
 Media AI Lab.