物流倉庫における新規物品の認識性能を向上させるFew-shot学習技術

Few-Shot Learning Technology to Improve Recognition of New Objects in Physical Distribution Warehouses

古茂田　和馬 KOMODA Kazuma

物流倉庫では，EC（電子商取引）市場の拡大と労働人口の減少を背景に，物品のピッキング作業自動化に向けて多種多様な物品を安定して認識する技術が求められている。しかし，従来の深層学習モデル（画像認識などに用いられるAIモデル）は，学習を行うときに新規物品ごとに多数のラベル付き教師学習画像（以下，学習画像と略記）が必要で学習への負担が大きく，また物品の重なりや遮蔽が生じる混載環境では認識性能が低下しやすかった。

東芝は，少数の学習画像から新規物品を効率的に追加学習できるFew-shot学習技術を開発した。学習画像の特徴のばらつきを考慮する多様性スコアとクラスタリングによる学習画像選択や，多様な条件の学習画像を得るデータ拡張，既存物品の認識性能を維持させる知識蒸留などを組み合わせることで，1～10枚の少数の学習画像による追加学習で，既存物品の認識性能を維持しながら新規物品を認識できることを確認した。

In the logistics and physical distribution field, there is a growing need for technologies that can stably recognize a wide variety of objects to automate picking operations in physical distribution warehouses, due to the expansion of e-commerce and workforce shortages. However, conventional deep learning-based recognition models not only require a large number of labeled training images for each new object, but their recognition performance also tends to degrade in cluttered environments where objects overlap or are partially occluded.

To overcome these issues, Toshiba Corporation has developed a few-shot learning technology that efficiently adds new objects to the recognition model by training on a very small number of labeled images. The technology comprises the following functions: (1) an image selection function based on a diversity score and clustering that takes feature variations in images into consideration, (2) a data augmentation function to obtain images under various conditions, and (3) a knowledge distillation function to maintain recognition performance for previously learned objects. Experimental results have confirmed that the few-shot learning technology can recognize new objects using only one to ten labeled training images without sacrificing recognition performance for previously learned objects.

1. まえがき

EC市場の拡大と労働人口の減少を背景として，物流倉庫では，物品のピッキング作業の自動化への要求が高まっている。東芝は，これまでにピッキングロボットを開発してきた（**図1**）[(1)]。ピッキングロボットの性能向上には，多種多様な物品を安定して認識できることが重要である。

従来の深層学習による認識モデル（以下，モデルと略記）では，ベースモデル（多数の学習画像で事前学習したモデル）が学習していない新規物品について，様々な向きや，配置，照明条件で撮影した多数の学習画像を用いて学習させる必要があった。このため，新規物品が頻繁に追加される物流倉庫では，新規物品の学習画像収集とアノテーション（学習画像の物品領域を指定してラベルを付け，学習用の正解データを作成する作業）の負担が大きく，モデルの更新に時間を要していた。

また，物流倉庫では複数の物品が同じコンテナ内に混在

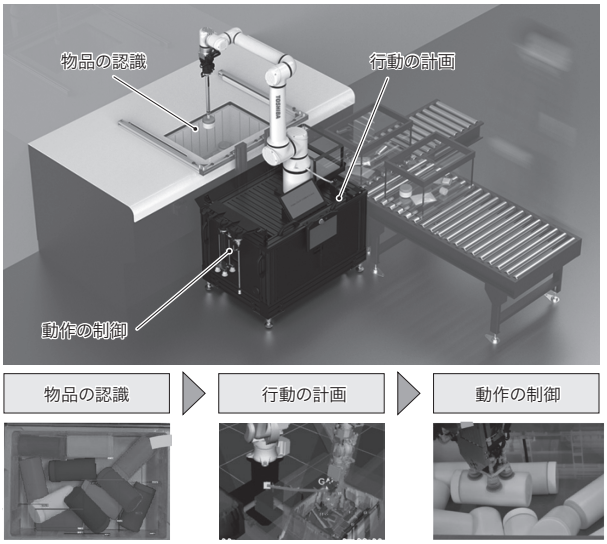


図1. ピッキングロボットシステムの構成

ピッキングロボットシステムは，物品の認識，行動の計画，及び動作の制御で構成され，物品認識の性能がピッキング成功率に大きく影響する。

Picking operation sequence of robot system

し，物品同士が重なったり，一部が隠れたりする遮蔽が頻繁に発生する。このような環境では，一つの物品を複数であると誤認識したり，物品全体ではなく一部分だけを認識したりすることがある。特に，新規物品の場合には，認識性能が低下しやすい。

そこで当社は，少数の学習画像によって追加学習を行うことで，重なりや遮蔽がある環境でも物品を認識させることができるFew-shot学習技術を開発した。Few-shot学習技術は，ベースモデルを基に，新規物品の少数学習画像だけで追加学習済みモデルを構築する技術である[2]。これにより，モデルをゼロから学習し直すことなく，既存の認識機能を活用しながら新規物品の認識に対応できるモデルを作ることができる。

ここでは，Few-shot学習技術について述べる。

2. Few-shot学習技術の概要

図2にFew-shot学習技術の構成を示す。ベースモデルは，学習画像全体から特徴を抽出する機能を持つViT-CoMer Mask R-CNN[3]を採用した。

Few-shot学習技術の開発において，まず，既存物品の学習画像を用いてベースモデルを学習させる。次に，追加学習の対象となる新規物品の学習画像を選択した後，アノテーション，データ拡張，知識蒸留，及びファインチューニング（事前学習済みモデルのパラメーターを新規物品向けに調整する処理）を行い，追加学習モデルを構築する。この結果，少数の新規物品の学習画像による追加学習だけで，ベースモデルの既存物品の認識性能を維持しながら新規物品の認識性能を持つ追加学習済みモデルを獲得できた。

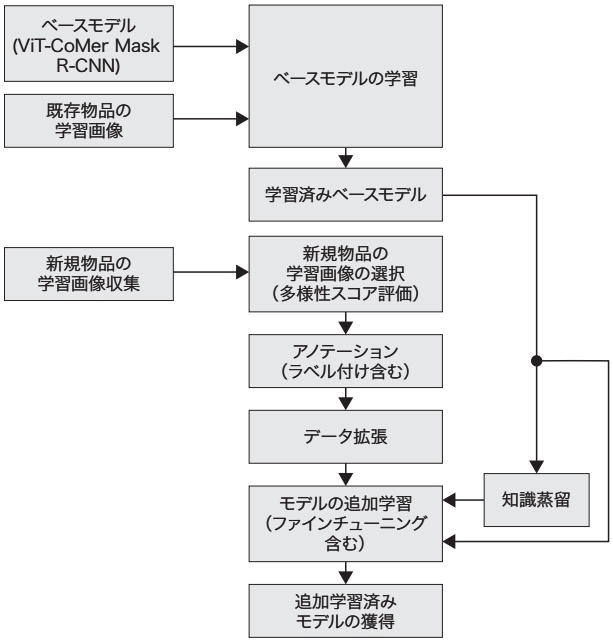


図2. Few-shot学習技術の構成

Few-shot学習技術は，新規物品の学習画像選択，データ拡張，及び知識蒸留などを組み合わせ，少数の新規物品の学習画像によるモデルの追加学習を実現する。

Few-shot learning architecture

3. Few-shot学習技術の重要機能

3.1 遮蔽状態での物品特徴抽出機能

Few-shot学習技術では，物品ごとの輪郭を画素単位で識別するインスタンスセグメンテーションの精度と，学習していない新規物品の学習画像に対する認識性能（汎化性能）を向上させるため，マルチスケール特徴抽出（物品の大きさの違いに対応する抽出機能）とアテンション機能（学習画像中の重要な部分に着目する機能）を用いて，遮蔽状態でも物品の特徴を効果的に抽出しやすい。

3.2 知識蒸留

新規物品だけを追加学習すると，以前に学習した既存物品の認識性能が低下する（破滅忘却）場合がある。そこで，**図3**に示すように，学習済みベースモデルを参照しながらモデルの追加学習を行う知識蒸留を適用した。この結果，既存物品の認識性能を維持しつつ，新規物品の認識性能を獲得できる。

3.3 学習画像選択のための多様性スコア

新規物品の学習画像からFew-shot学習に用いる学習画像を選択する際の評価指標として，多様性スコアを導入した。多様性スコアは，学習済みモデルによる認識に対し，同一学習画像に対する複数回の認識結果のばらつきを，認

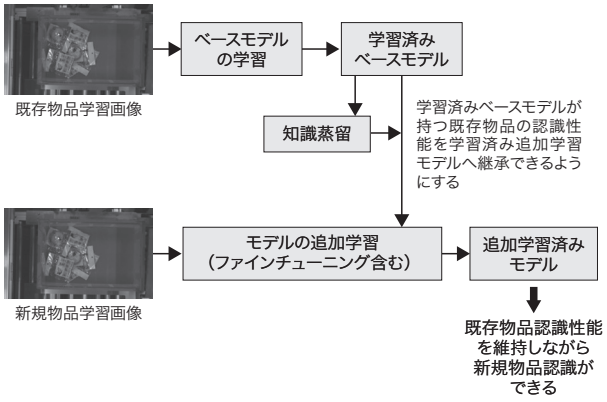


図3. 知識蒸留による既存物品認識性能の維持

追加学習済みモデルがベースモデルの認識性能を継承することで，新規物品を学習した後も既存物品の認識性能低下を抑制できる。

Recognition performance for previously learned objects maintained by knowledge distillation function

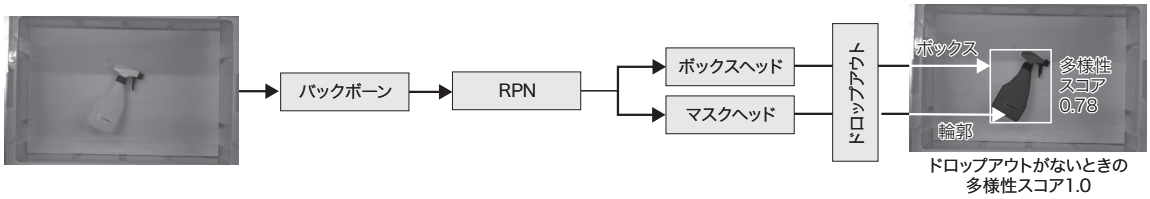


図4. 多様性スコアによる画像評価

学習画像認識フローの後段にドロップアウトを加え，学習画像に対する認識結果のばらつきから，多様性スコアを算出する。このスコアを用いて，追加学習に有効な学習画像を選択する。

Flowchart of image evaluation through diversity score

識された物品領域の重なり具合（IoU：Intersection over Union）を用いて数値化した指標である[4]。

この指標の評価では，**図4**に示す学習画像認識フローを用いた。まず，特徴抽出機能（バックボーン）で学習画像の特徴を抽出し，物品の候補領域抽出機能（RPN）で候補領域を抽出する。その後，物品の種類と位置を推定する機能（ボックスヘッド）で物品の種類と位置を，物品領域（輪郭）を推定する機能（マスクヘッド）で物品の輪郭を，推定する。認識結果のばらつきを評価するため，学習画像認識フローの後段にドロップアウト（学習時に一部の機能をランダムに停止する手法）を追加して意図的にノイズを発生させ，同一入力学習画像に対する認識結果の変化を評価する。出力の変動が小さい学習画像は，安定した学習画像であると考えられる。一方，認識結果のばらつきが大きい学習画像は，物品の多様な特徴を含むため，追加学習に有効であり，多様な見え方にも対応できるようにする学習画像として選択する。多様性スコアは，このような学習画像選択の判断材料として用いた。

3.4 多様性のある学習画像の選択

少数の学習画像による学習では，限られた枚数の中から学習効果の高い学習画像を選択することが重要である。類似した学習画像ばかりを選ぶと，追加学習で得られる情報が偏る。一方，多様な姿勢や，配置，遮蔽状態，背景条件などを含む学習画像を選択すると，少数の学習画像でもモデルの汎化性能を高めることができる。

ここで，多様性スコアに基づいて学習画像を評価する手法を**図5**に示す。学習画像ごとの特徴を抽出した後，k-meansクラスタリング（類似した学習画像を自動的にグループ化する手法）によって学習画像をグループ化する（図5(a)）。多様性スコアがしきい値より大きい場合，各グループの中から多様性スコアが一番大きな学習画像を選択する（図5(b)）。一方，多様性スコアがしきい値より小さい場合，各グループを更にクラスタリングで細分化し，各小グループの中から，グループ全体の平均的な特徴に最も近い学習画像を代表学習画像として選択する（図5(c)）。これにより，多様な特徴を持つ学習画像を効率的に選択する。

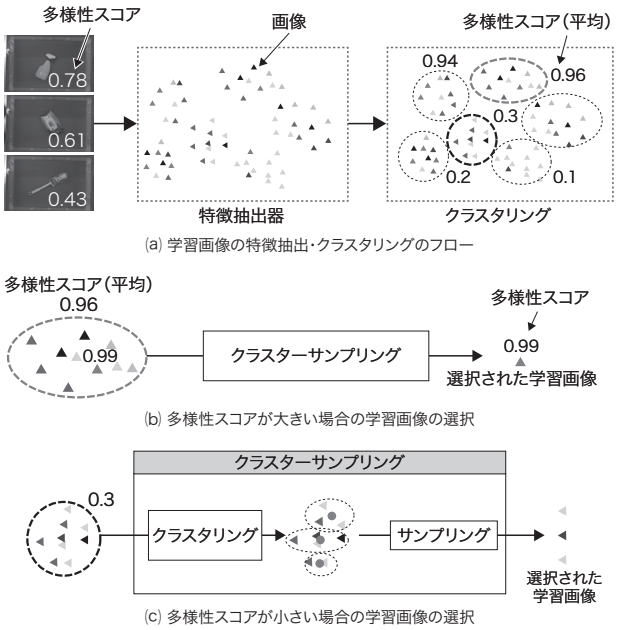


図5. 多様性を考慮した画像選択

多様性スコアを用いた学習画像選択手法を示す。学習画像を特徴に基づいてクラスタリングし，各クラスターから多様性スコアに基づいて代表画像を選択する。

Image selection clustered by diversity scores

3.5 データ拡張

データ拡張では，まず新規物品の学習画像とは別に，背景として用いる画像（以下，背景の学習画像と呼ぶ）を準備する。次に，背景の学習画像のコントラストをランダムに変更し，明るさや見え方の異なる背景の学習画像を生成する。続いて，学習対象となる新規物品の学習画像の中から，マスクされた学習画像をランダムに抽出する。抽出した物品学習画像に対して，水平反転，垂直反転，及びサイズ変更の変換を行い，背景の学習画像上のランダムな位置に配置する。これにより，物品の姿勢や，大きさ，配置，背景条件などが異なる複数の学習画像を生成できる。以上の処理により，少数の学習画像から多様な学習画像を作成する。

4. Few-shot学習技術の評価結果

4.1　新規物品の認識性能の向上

Few-shot学習技術の認識性能を評価するため，ベースモデルとしてYOLOv7を用いた従来技術と，Few-shot学習技術とを比較した。Few-shot学習技術では，新規物品の学習画像として，54種類の日用品の学習画像1,848枚から，多様性スコアやクラスタリングなどにより各新規物品について選択した10枚の学習画像を用いて，新規物品の学習画像の追加学習を行い，認識性能を評価した。

物品を漏れなく正確に認識できるかを評価するために，物品検出とインスタンスセグメンテーションの平均適合率（mean Average Precision：mAP，物品認識結果の総合的精度指標）を用いて，IoUしきい値を0.75以上の場合を正解として認識性能を評価した。結果を**図6**に示す。

従来手法では，物品検出の認識性能が68.2 %，インスタンスセグメンテーションの認識性能が58.8 %であった。認識結果は，一つの物品を複数の物品として検出する過剰分割が見られた。

これに対し，Few-shot学習技術では，物品検出の認識性能が78.5 %，インスタンスセグメンテーションの認識性能が67.2 %となり，過剰分割が抑制され，認識性能が向上した。

4.2　知識蒸留による既存物品の認識性能の低下抑制

新規物品の学習に伴う既存物品に対する認識性能の劣化を検証した。Few-shot学習技術に対して，知識蒸留を用いなかった場合と用いた場合とで認識性能を比較した。知識蒸留を用いなかった場合，既存24物品の物品検出の認識性能は87.9 %，インスタンスセグメンテーションの認識性能は81.1 %であった。これに対し，知識蒸留を用いた場合，物品検出の認識性能は93.3 %，インスタンスセグメンテーションの認識性能は88.5 %であった。この結果から，知識蒸留によって既存物品の認識性能の低下を抑制できることを確認した。

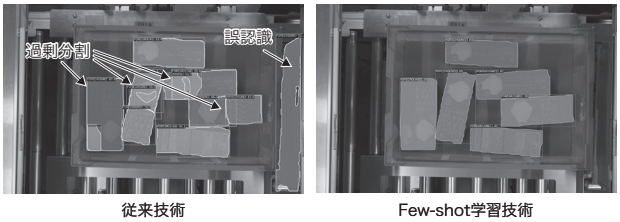


図6. 従来技術とFew-shot学習技術の認識結果比較

従来技術では新規物品の認識結果に過剰分割や誤認識が見られる一方，Few-shot学習技術では物品領域をより適切に認識できる。

Comparison of image recognition results obtained by conventional and few-shot learning technologies

4.3　One-shot条件におけるFew-shot学習技術の有効性検証

少数の学習画像による追加学習の有効性を検証するため，各新規物品から1枚の学習画像だけを追加して認識性能を評価した。

その結果，新規物品検出の認識性能は76.3 %，インスタンスセグメンテーションの認識性能は65.0 %となり，4.1節で示した10枚とほぼ同等の認識性能が得られることを確認した。

5. あとがき

当社は，物流倉庫において少数の学習画像で既存物品と新規物品の両方を認識できるFew-shot学習技術を開発した。この技術で，多様性スコアとクラスタリングによる学習画像選択や，データ拡張，知識蒸留などを組み合わせることで，新規物品と既存物品の認識性能を両立させた。評価実験では，54種類の日用品の学習画像を対象に，1〜10枚の少数の追加学習画像で新規物品を認識できることを確認した。

今後は，学習コスト削減に向けた実証を進め，現場への早期導入を目指す。

文　献

(1) 岩崎利夫，ほか．人との協働が可能な可搬性と可用性を備えたピッキングロボットシステム．東芝レビュー．2024，**79**，3，p.9–12．<https://www.global.toshiba/content/dam/toshiba/jp/technology/corporate/review/2024/03/a04.pdf>，（参照 2026-06-19）．

(2) 古茂田和馬，ほか．“物流倉庫内の未知物品に対する認識性能を向上する多様性スコアに基づいたFew-shot学習”．2025年度人工知能学会全国大会（第39回）論文集．大阪，2025-05，人工知能学会．2025，3M1-GS-10-04．

(3) Xia, C. et al. "Vit-CoMer: Vision transformer with convolutional multi-scale feature interaction for dense predictions". Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Seattle, WA, 2024-06, IEEE. 2024, p.5493–5502.

(4) Blok, P. M. et al. Active learning with MaskAL reduces annotation effort for training Mask R-CNN on a broccoli dataset with visually similar classes. Computers and Electronics in Agriculture. 2022, **197**, 106917, p.1–16. <https://www.sciencedirect.com/science/article/pii/S0168169922002344>, (accessed 2026-07-07).



古茂田　和馬　KOMODA Kazuma, Ph.D.
総合研究所 生産技術センター
ロボット・メカトロニクスソリューション技術研究部
博士（工学）　IEEE・日本ロボット学会・日本機械学会・人工知能学会会員
Robotics and Mechatronics Solutions Technology Research Dept.