DEIM Forum 2023 3a-3-3

インスタンスセグメンテーションおけるデータ類似性に基づく転移学習済みモデルの検索

日置　淳也† 　三林　亮太† 　山本　岳洋† 　窪内　将隆†† 　大島　裕明†

† 兵庫県立大学 大学院情報科学研究科　〒 651–2197 神戸市西区学園西町 8–2–1
†† 堺化学工業株式会社　〒 590–8502 大阪府堺市堺区戎島町 5–2
E-mail: †{junya.hioki.j,threeforest8}@gmail.com, ††t.yamamoto@sis.u-hyogo.ac.jp,
†††kubouchi-m@sakai-chem.co.jp, ††††ohshima@ai.u-hyogo.ac.jp

あらまし　本研究では，インスタンスセグメンテーションを行いたい画像に対して，良い性能を示す機械学習モデルを検索する手法について提案する．現在では，機械学習技術の発展によって，産業界においても機械学習技術が用いられている．その中でも本研究ではインスタンスセグメンテーションと呼ばれる物体検出技術に着目する．このような機械学習技術を画像に用いて，粒子の検査を行う企業も存在する．そのような企業にとって，粒子の検査を正確に行うために必要なことは，適した事前学習を行った機械学習モデルを選択することである．同一のアーキテクチャの機械学習モデルであっても，異なるデータで事前学習された機械学習モデルであれば，全く異なる結果を示す．そこで本研究では検査したい粒子画像に対して，インスタンスセグメンテーションを行う際に，適した機械学習モデルを検索する手法を提案する．具体的なアプローチとしては，画像の類似度に着目する．本研究では，予め全ての検索対象の機械学習モデルを用いて，インスタンスセグメンテーションを行った粒子画像の中から，最もクエリ画像と類似する画像を検索し，その画像に対して良い性能でインスタンスセグメンテーションを行うことができる機械学習モデルをランキング形式で出力する．

キーワード　深層学習，インスタンスセグメンテーション，物体認識，転移学習，機械学習モデル検索

1 はじめに

近年，機械学習技術の発展により様々な場面で機械学習技術が用いられるようになった．多岐にわたる分野の様々な課題において，深層学習ベースの技術を適用することで，従来より良い性能が得られることが報告されている．そこで重要な役割を果たしているのが，大量のデータを用いて様々な問題に汎用的に対応することができるように学習された事前学習済み機械学習モデルである．このような事前学習済み機械学習モデルはWeb 上で数多く提供されるようになっている。

自然言語処理や画像認識において，ある特定の解きたい問題があるユーザは，これら Web 上で提供されている事前学習済み機械学習モデルを入手し，解きたい問題に用いる．しかし，Web 上には事前学習済み機械学習モデルは数多く存在するため，ユーザの解きたい問題に適した事前学習済み機械学習モデルを選択することは大変困難になっている．

このような問題は，実際に機械学習を用いて業務を行う企業にも存在する．近年では，粒子の大きさや形状異常を検査するために，粒子の画像に対してインスタンスセグメンテーションと呼ばれる物体検出技術を用いる企業も現れるようになった．その結果，以前までは実際に人間が手を動かし，大きな人的コストと時間的コストをかけていた作業が効率良く行われるようになった．この業務において重要となるのが適した機械学習モデルの選択である．ここで間違ったモデルを選択してしまうと，うまく物体検出がされないため，間違った検査結果を示してしまう恐れがある．既存の機械学習モデルは大量に存在し，粒子の画像が与えられた際に，膨大な数の中から，その画像に適したモデルを選択することは大変困難な状況にある．最適な機械学習モデルを選択するためには，実際に一つ一つのモデルでインスタンスセグメンテーションを行い試していく必要があるが，時間的コストを考えると現実的ではない．

そこで本研究では，データの類似性に着目した機械学習モデルの検索を行う．具体的なアプローチとしては，類似する画像を用いたものである．予め全ての検索対象の機械学習モデルを用いて，インスタンスセグメンテーションを行った粒子画像の中から，最もクエリ画像と類似する画像を検索し，その画像に対して性能良くインスタンスセグメンテーションを行うことができる機械学習モデルをランキング形式で出力する．

これは図 1 に示すように，画像間の類似度が高ければ，同じ機械学習モデルを使ってインスタンスセグメンテーションを行った場合に，同程度の性能が得られるという仮説に基づいた手法である．

2 関連研究

本研究の目的は, 検査を行いたい粒子画像に対して, インスタンスセグメンテーションをする際に, 適した機械学習モデルを選択することである．関連研究として，物体検出を行った研究，画像から特徴量を抽出するバックボーンニューラルネットワー

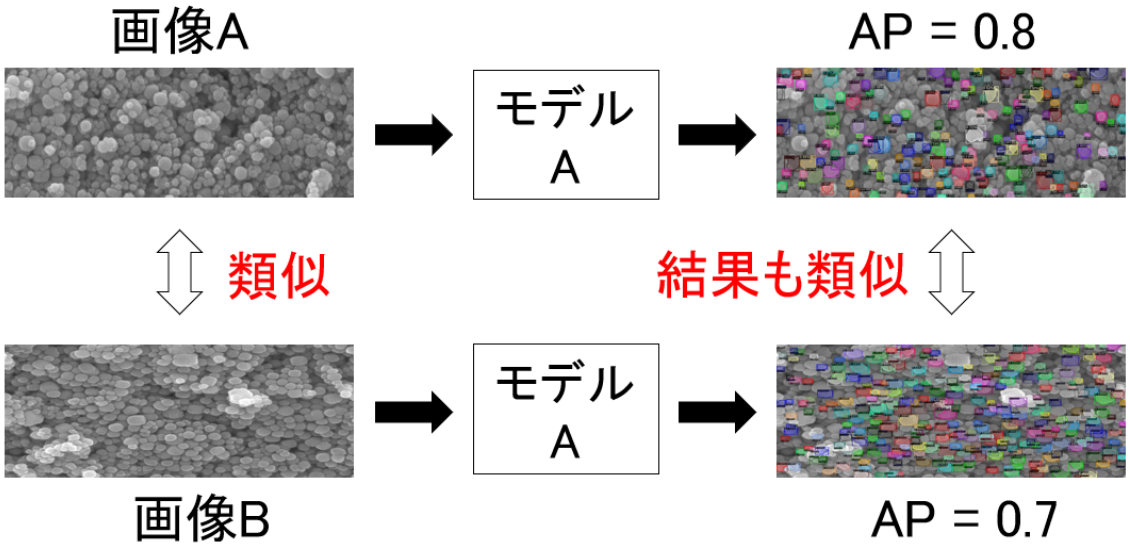

図 1　仮　　説

ク構造の研究，類似度を用いた研究，機械学習モデルの検索を行った研究が挙げられる．

2.1　セグメンテーション

画像に写る物体に対して，物体の位置とその物体のクラスを予測する手法にセグメンテーションがある．セグメンテーションは，セマンティックセグメンテーション（semantic segmentation），インスタンスセグメンテーション（instance segmentaion），パノプティックセグメンテーション（panoptic segmentation）[9] の大きく 3 つに分類される．図 2 を用いてそれぞれのセグメンテーションについて詳細を述べる．図 2 の（a）は，元画像である．この画像に対して，セマンティックセグメンテーションを行った画像が（b），インスタンスセグメンテーションを行った画像が（c），パノプティックセグメンテーションを行った画像が（d）である．(b）に示すセマンティックセグメンテーションは，画像中のすべての画素に対して，クラスラベルを予測することを目的とする．(c）に示すインスタンスセグメンテーションは，画像中のすべての物体に対して，クラスラベルを予測し，一意の ID を付与することを目的とする．セマンティックセグメンテーションとの主な違いは，重なりのある物体を別々に検出する点や，空や道路などの定まった形を持たない物体などはクラスラベルの予測を行わない点などがある．また各物体に対して一意の ID を付与するため，例えば，一つの画像に複数の車が写っている場合，それぞれの車を別々の物体と認識することが可能である．(d）で示すパノプティックセグメンテーションは，セマンティックセグメンテーションとインスタンスセグメンテーションを組み合わせた手法であり，画像のすべての画素に対して，クラスラベルを予測し，一意の ID を付与することを目的とする．

セグメンテーションの手法については，数々の手法が過去にも提案されてきた．

その中でも，CNN によって得られた特徴マップを用いた Johnson ら [8] の手法が主な手法として提案されている．

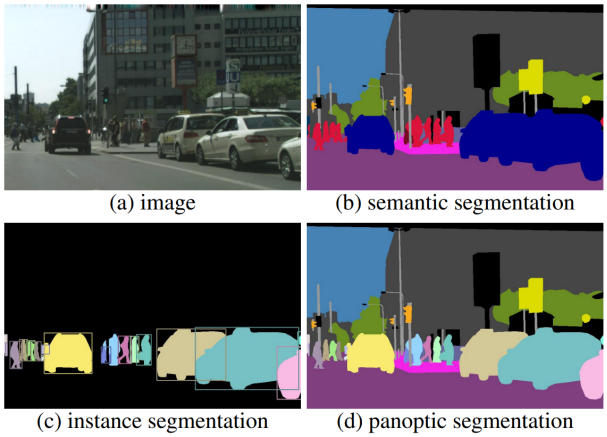

図 2　セグメンテーションの種類（参考文献 [9] より引用）

セグメンテーションを行う代表的なアーキテクチャとして，R-CNN [5] や Fast R-CNN [4]，Faster R-CNN [17]，Mask R-CNN [6]，YOLO [15]，SSD [11] などが挙げられる．本研究で対象とするセグメンテーション及びアーキテクチャは，インスタンスセグメンテーションと第 4.2.1 節で後述する Mask R-CNN を用いる．

2.2　画像処理における特徴量抽出

画像処理における，個別のタスクとしては，「画像分類」，「物体検出」，「セグメンテーション」などが存在する．それらのすべてのタスクにおいて汎用的に必要となるのは，画像の特徴量を抽出するニューラルネットワーク構造である．この部分はバックボーンと呼ばれ，代表的には，ResNet [7]，Inception [19]，Vision Transformer（ViT）[2] などが挙げられる．本研究では，画像の特徴量を抽出するバックボーンとして ResNet50 を用いる．

2.3　類似度を用いた研究

類似度を用いた研究は過去にもいくつか存在する．Reimers

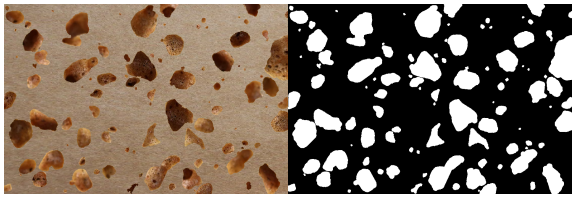

図 3　食品データセット

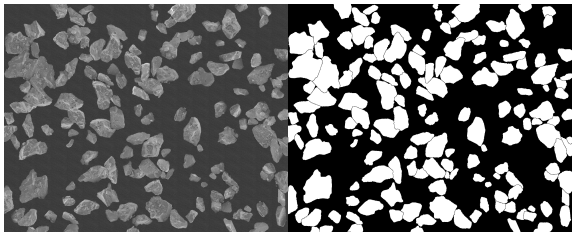

図 4　鉱物データセット

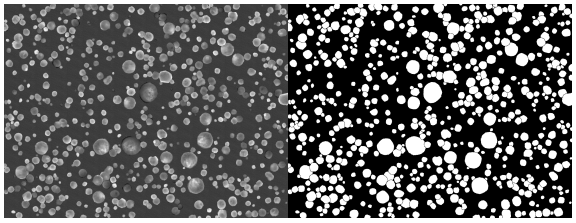

図 5　円形粒子データセット

ら [16] が提案した Sentence-BERT では，類似する文章のペアを探す自然言語処理タスクにおいて，高速化を実現した．自然言語処理タスクにおいて BERT [1] や RoBERTa [12] などの事前学習済みモデルは汎用的に用いられている．しかし，二つの文章を比較するタスクにおいては二つの文章を同時にネットワークに入力する必要があるため，全組み合わせを試行する必要があるため計算時間が膨大になっていた．そこで提案された Sentence-BERT は文章の特徴表現を得ることで特徴量同士の類似度を比較することを可能とした．本研究は画像処理タスクではあるが，類似するペアを探すという点で関連が深い．また，本研究では高速に画像間の類似度を計測する必要がある．高速で類似度計測を行うライブラリには，FLANN [14] や，Annoy, Faiss [3]，NMSLIB [13] など様々なものが存在する．本研究では，MetaAI によって開発された Faiss を用いる．

2.4　機械学習モデルの検索を行った研究

事前学習済み機械学習モデルの検索に関する研究は少ない．Ueno ら [20] は，画像分類タスクにおいて，複数の事前学習された機械学習モデルの中から実際にファインチューニングを行うことなく，適したモデルを選択する手法を提案した．ここで用いられた手法は，畳み込みニューラルネットワークの最後の畳み込み層が出力する特徴マップに対して，指標をつけて評価するというものである．しかし，物体検出やセグメンテーションといった他の画像処理のタスクにおいては，検証されていない．

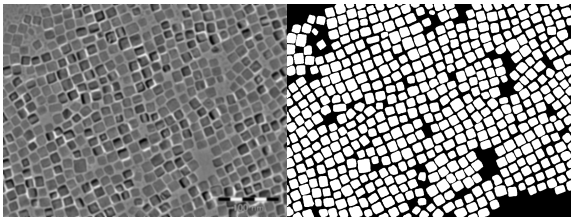

図 6　四角形粒子データセット

3　画像の類似度を用いた機械学習モデルの検索

3.1　概　　要

本研究では，画像の類似度を用いた機械学習モデルの検索を行う．これは図 1 に示すように，画像間の類似度が高ければ，同じ機械学習モデルを使ってインスタンスセグメンテーションを行った場合に，同程度の性能が得られるという仮説に基づいた手法である．

問題定義は以下の通りである．

- **入力：検査を行いたい粒子画像**
- **出力：機械学習モデルのランキング**

本研究では，ある粒子画像を与えた際に，その画像に対してインスタンスセグメンテーションを性能良く行える機械学習モデルをランキング形式で出力する．

具体的なアプローチは，図 7 のようにクエリ画像に最も類似する代表画像に対して，性能良くインスタンスセグメンテーションを行えた機械学習モデルをランキング形式で出力するというものである．まず，事前に準備した複数枚の画像に対して，全ての検索対象の機械学習モデルを用いてインスタンスセグメンテーションを行う．次に，それぞれの画像で各モデルのインスタンスセグメンテーションの性能の良さをもとに機械学習モデルのランキングを作成する．そして，それらの画像の中から最もクエリ画像と類似する画像を検索し，その画像が持つ機械学習モデルのランキングを予測ラベルとして出力する．以下で詳細を述べる．

前準備として図 8 に示すように，代表画像になり得る粒子画像に対して，検索対象の全ての機械学習モデルを用いてインスタンスセグメンテーションを行う．ここで得られた各機械学習モデルの性能を評価し，その画像特有の機械学習モデルのランキングを得る．

ここから代表画像の決定について述べる．概要を図 9 に示す．まず最初に，検索対象のモデルとは全く異なる畳み込みニューラルネットワークを用意する．そのモデルに対して，粒子画像のクラス分類タスクで fine-tuning を行い，特徴量抽出器 (CNN Particle，以下 CNN-P) を得る．そこで得られた CNN-P を用いて事前に準備したランキングを保有する粒子画像の特徴量を抽出し，抽出した特徴量に対して K-means でクラスタリングを行う．ここで得られた各クラスの重心に最も近い画像を代表画像と定義する．

そこへクエリ画像が与えられた際には，クエリ画像に対しても CNN-P を用いて特徴量を抽出する．ここで得られたクエリ

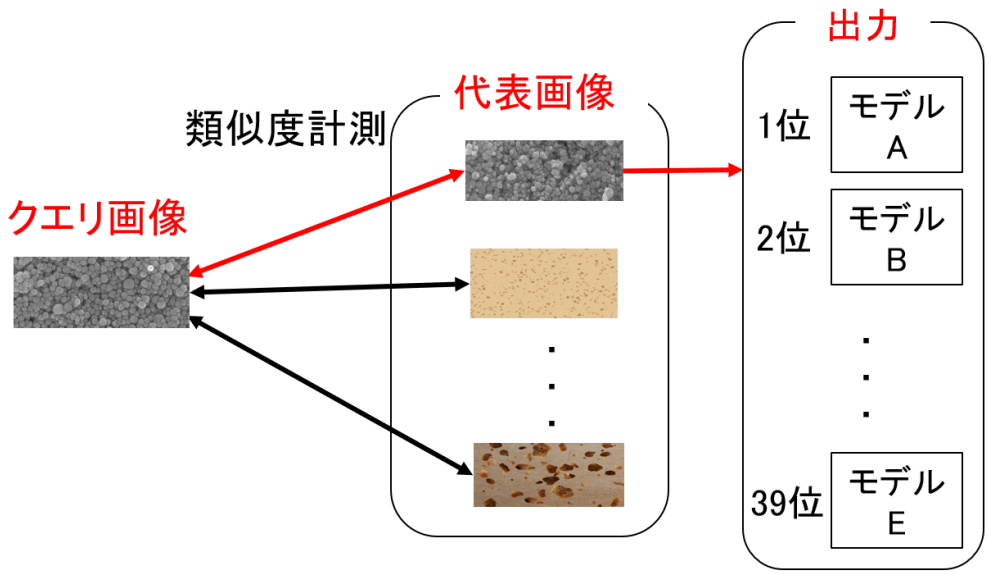

図 7　アプローチ

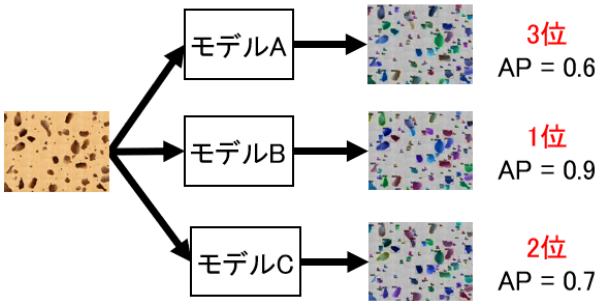

図 8　ランキングの作成

画像の特徴量と先ほど定義した全ての代表画像の特徴量の類似度を計測する．

ここでクエリ画像に，最も類似する代表画像が保有する機械学習モデルのランキングを出力とする．

3.2　クラス分類学習による特徴量抽出器の作成

本節では特徴量抽出器の作成について述べる．まず ImageNet [18] で事前学習された畳み込みニューラルネットワークを用意する．その重みを初期の重みとして，粒子画像を用いてクラス分類学習を行う．学習後，ネットワークの最終層の全結合層直前のプーリング層の出力を抽出特徴量とすることで CNN-P を得る．

本研究では，ネットワークに ResNet50 を用いた．分類学習後には，最終層の全結合層を取り除き，特徴量抽出器とした．得られる特徴量の次元数は 2048 次元である．

3.3　クラスタリングを用いた代表画像の決定

本節では，クラスタリングを用いた代表画像の決定について述べる．第 3.2 節で得られた特徴量は 2048 次元であるため，主成分分析を用いて 2 次元に圧縮した．

ここで得られた 2 次元の特徴量空間に k-means でクラスタリングを行い，各クラスの重心に最も近い画像を代表画像と定義する．

本研究ではクラス数を徐々に変化していき，代表画像の枚数を変化させて実験を行う．

表 1　データセット枚数

分類学習用	449
代表画像候補用	250
クエリ画像用	275

3.4　代表画像との類似度計測と出力

本節では，代表画像との類似度計測と出力について述べる．代表画像が決定した後に，クエリ画像を与える．クエリ画像も代表画像と同様に，CNN-P を用いて 2048 次元に特徴量化する．ここで得られたクエリ画像の特徴量と，全ての代表画像の 2048 次元の特徴量のコサイン類似度を計測する．ここでは，高速で最近傍探索を行うために，近傍探索ライブラリ Faiss を用いた．全ての代表画像のうち，最もクエリ画像と類似度の高い画像の保有するランキングを出力とする．

4　実　　験

4.1　データセット

データセットとして共同研究を行う堺化学より提供された粒子画像を使用した．本研究において必要なデータセットは CNN-P を作成するための分類学習用データと、インスタンスセグメンテーションタスクに用いる代表画像候補用データとクエリ画像用データの 3 つである．各用途のデータセットの枚数を表 1 に示す．

また，インスタンスセグメンテーションタスクに用いる代表画像候補用データとクエリ画像用データについて述べる．本研究におけるモデル性能の評価は，第 4.2.2 節で後述する AP である．ある画像に対して，インスタンスセグメンテーションを行った際の AP に基づいて機械学習モデルのランキングを出力

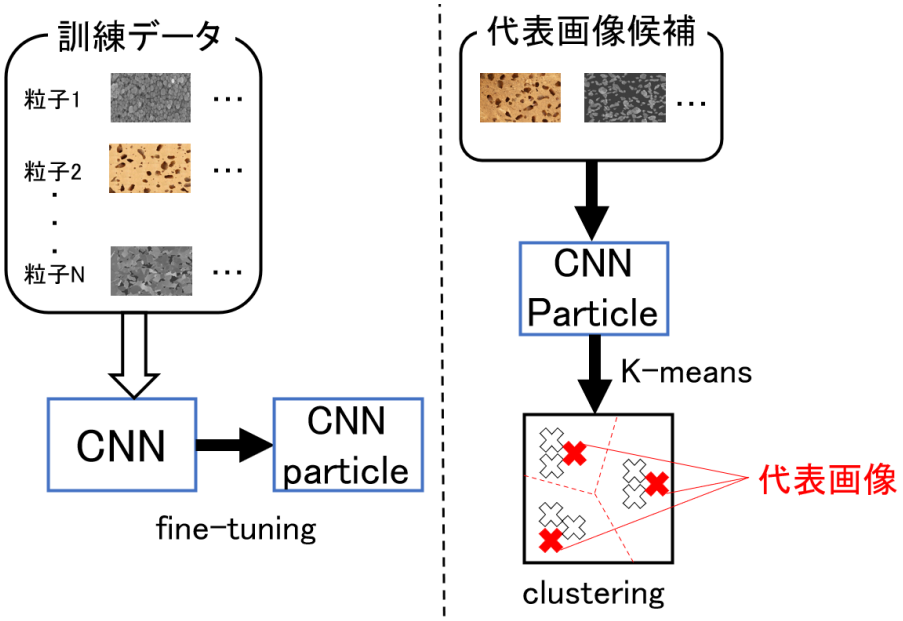

図 9　代表画像の決定方法

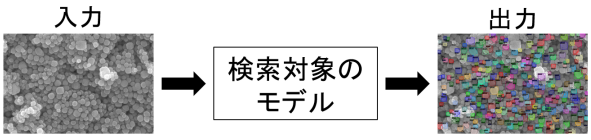

図 10　検索対象のモデルの入出力

する．

粒子画像のセグメンテーションタスクの評価に用いるデータセットは，2 枚をペアとする画像である必要がある．一枚は粒子のオリジナル画像，もう一枚はその画像の正解データとなるマスク画像である．マスク画像は，粒子部分を白色，背景を黒色で塗りつぶした画像である．データセットには，図 3 のようなパンの気泡を拡大した画像や，髪の毛を拡大した画像，図 4 のような鉱物の画像や，図 5 のような円形の粒子画像，図 6 のような四角形の粒子の画像等が含まれる．

4.2　検索対象となる機械学習モデル

本節では，検索対象となる機械学習モデルについて述べる．本研究で，検索対象とする機械学習モデルは全部で 39 種類で，共同研究を行う堺化学より提供された機械学習モデルである．これらの機械学習モデルは，図 10 のように粒子画像を入力として与えると，粒子部分がインスタンスセグメンテーションされた画像が出力される．

本節で検索対象とする機械学習モデルは，全て同じアーキテクチャを持つ機械学習モデルである．これらは図 12 のように同様の事前学習を行った機械学習モデルに対して，それぞれ別のデータを用いて転移学習を行い作成された機械学習モデルである．

4.2.1　Mask R-CNN

本研究で，検索対象となる機械学習モデルはインスタンスセグメンテーションを行うモデルである．取り扱う機械学習モデルのアーキテクチャは，Meta の He らが開発した Mask R-CNN である．本研究で検索対象となる 39 種類の機械学習モデルは，いずれも Microsoft で作成された COCO データセット [10] で事前に訓練したモデルを，それぞれ別の粒子画像のデータセットで転移学習を行ったモデルである．Mask R-CNN のモデル構造は以下に示すように 4 つの段階で構成されている．

- バックボーンニューラルネットワークで特徴量抽出
- 特徴マップから関心領域（ROI）の抽出
- ROI の形状の同一化
- 分類，バウンディングボックスの推定，マスクの予測

本研究では，Data Augumentation に関して以下の処理を行った．

- 画像を 0.5〜1 倍にリサイズ
- 画像の輝度を 0.5〜1.5 倍に調整
- 画像のコントラストを 0.5〜1.5 倍に調整
- 水平方向で反転

まず，画像を読み込みバックボーンニューラルネットワークに入力される．本研究では ResNet50 をバックボーンニューラルネットワークとして使用する．バックボーンニューラルネットワークは，入力画像の特徴マップを出力する．

次に，RPN という関心領域（ROI）を抽出するネットワークを使用する．

そして，最後の出力層に入力するためには，関心領域（ROI）の形状は同じである必要である．そのため，RPN で抽出された特徴は，ROI Align という方法で，関心領域（RPN）の同一化を行う．

最後に，同一化された ROI の分類，バウンディングボックスの推定，マスクの予測を行う．

4.2.2　検索対象の機械学習モデルの性能の評価

本研究では Faster R-CNN，Mask R-CNN などのオブジェクト検出モデルを評価するために使用される最も一般的な評価

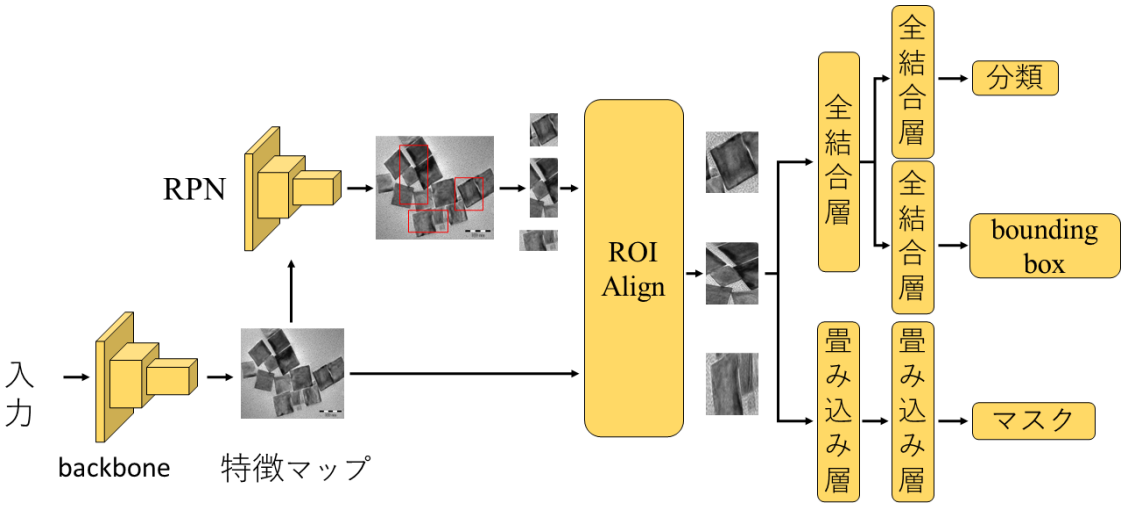

図 11　Mask R-CNN モデル構造

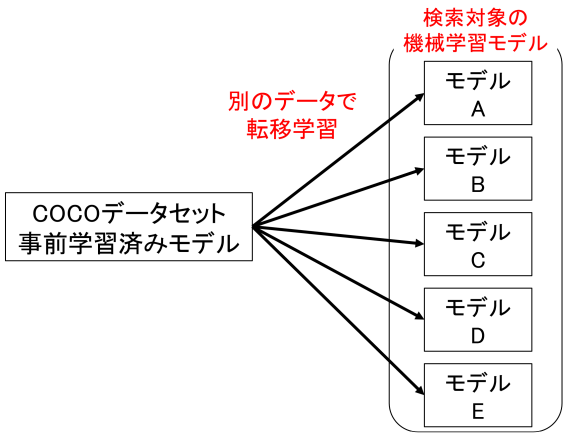

図 12　転移学習による検索対象の機械学習モデルの作成

指標 AP を使用する．r を再現率、p を適合率とすると以下の式で表せる．

$$AP = \int_0^1 p(r)dr$$

4.2.3　検索対象の機械学習モデルの汎用性

本節では，検索対象の機械学習モデルの汎用性について述べる．本研究では 39 個の機械学習モデルを検索対象としている．その中には，幅広い画像に対して，性能良くセグメンテーションを行える機械学習モデルもあれば，どの画像に対してもセグメンテーションの性能が低い機械学習モデルも存在する．ここでは，250 枚のクエリ画像に対して，最も高い AP を示した機械学習モデルをベストモデルとして，ベストモデルの出現頻度を図 14 のヒストグラムで示す．図 14 から，ベストモデルの出現頻度にかなり偏りがあることが分かる．しかし，このような問題は本研究特有の問題ではない．インスタンスセグメンテーションタスクにかかわらず，他の画像処理タスクや自然言語処理タスクにおいても，汎用性の高さは機械学習モデルによって異なる．よって本研究では，このような条件で実験を行う．

4.3　クラス分類学習の実装の詳細

特徴量抽出器を作成する際のクラス分類学習時には，5 クラスの粒子画像を分類するタスクを設定した．学習率は 2×10^{-6} とし，過学習が起こる前に学習がストップする early stopping を実装し，97epoch で学習を止めた．ミニバッチのサイズは 64，入力画像は 256 × 256 にリサイズし，224 × 224 に中央部をくり抜いた．分類学習の学習曲線を図 13 に示す．

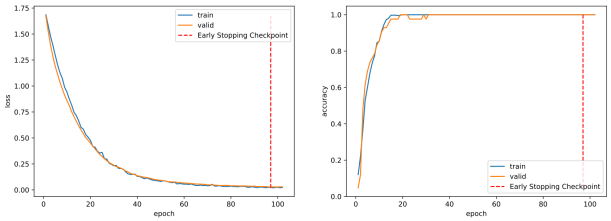

図 13　分類学習の学習曲線

4.4　k-means を用いた代表画像の決定

代表画像を決めるにあたって，本研究では k-means を用いた．特徴量抽出器から抽出した 2048 次元の全ての代表画像候補の特徴量を主成分分析を用いて 2 次元に圧縮し，2 次元の特徴量に対して，k-means を用いてクラスタリングを行った．各クラスの中心に最も近い画像を代表画像と定義した．ここではクラス数を 250，125，50，25 と徐々に変化させて結果を確認する．

4.5　評　　価

本節では，ランキングの評価について述べる．本研究では nDCG(normalize Discounted Cumulative Gain) を用いて評価を行う．

ここでは本研究の提案手法以外に，比較手法についても評価を行う．

4.5.1　機械学習モデルの汎用性を考慮したランキングを出力とする手法

本研究の検索対象となる機械学習モデルは，第 4.2.3 節で前述したようにそれぞれ汎用性が異なる．複数の画像に対して，性能良くインスタンスセグメンテーションを行える機械学習モ

デルもあれば，どの画像に対してもあまり性能良くインスタンスセグメンテーションを行えない機械学習モデルも存在する．どの画像に対しても性能良くインスタンスセグメンテーションを行える機械学習モデルを上位に，その他のモデルを下位とするランキングを，すべてのクエリ画像に対して出力とする手法を比較のために評価を行う．

ランキングを作成するに当たって、ボルダ得点を用いて，それぞれの機械学習モデルに得点を与えて，得点を降順に並べることによってランキングを作成した．

4.5.2 特徴量抽出器による性能の違い

本研究では，ImageNet で事前学習された ResNet50 の重みを初期の重みとして，粒子画像を用いて分類学習を行うことによって．特徴量抽出器を作成した．

ここでは分類学習を行う前の，ImageNet で事前学習されただけの ResNet50 を特徴量抽出器に用いた場合（CNN ImageNet，以下 CNN-I）と，CNN-P を特徴量抽出器として用いた場合を比較する．

4.5.3 nDCG

nDCG(normalized Discounted Cumulative Gain) を用いてランキングの評価を行う．nDCG は以下に示す式で表される．

$$nDCG@k = \frac{DCG@k}{DCG^{*}@k} = \frac{\sum_{i=1}^{k} \frac{g(i)}{\log_2(1+i)}}{\sum_{i=1}^{k} \frac{g^{*}(i)}{\log_2(1+i)}}$$

$$g\,(i) = rel_i$$

$nDCG@k$: 理想的ランキングに対する $DCG@k$

rel_i: i 位のアイテムの適合度

本研究では，以下の 2 つの設定で評価を行う．

4.5.4 評価方法 1

実際にクエリ画像に対して，全ての検索対象の機械学習モデルを用いて，インスタンスセグメンテーションを行い作成した機械学習モデルのランキングのうち，1 位の機械学習モデルが示す AP の 8 割以上の AP を示す機械学習モデルの適合度を 1 とする．それ以外を 0 とする．

例を挙げると，クエリ画像が保有するランキングの 1 位の機械学習モデルの AP が 0.9 の場合，0.72 以上の AP を示す機械学習モデルの適合度を 1 として評価する．

4.5.5 評価方法 2

実際にクエリ画像に対して，全ての検索対象の機械学習モデルを用いて，インスタンスセグメンテーションを行い作成した機械学習モデルのランキングのうち，1 位の機械学習モデルの適合度を 1 とする．それ以外を 0 とする．

5 結果と考察

まず，k-means を用いてクラス数を変化させて nDCG を比較することで，適した特徴量を抽出することが出来れば，代表画像の枚数を減らすことが出来ることを示す．

表 2 は，CNN-I を特徴量抽出器として用いてクラス数を変化させた場合の nDCG を比較したものである．k-means を用いて代表画像の枚数を 250 枚から 25 枚まで変化させた結果，代表画像の枚数を減少させると nDCG の値は低下しているものの，評価方法 1 で 0.932 という極めて高い値を示した．

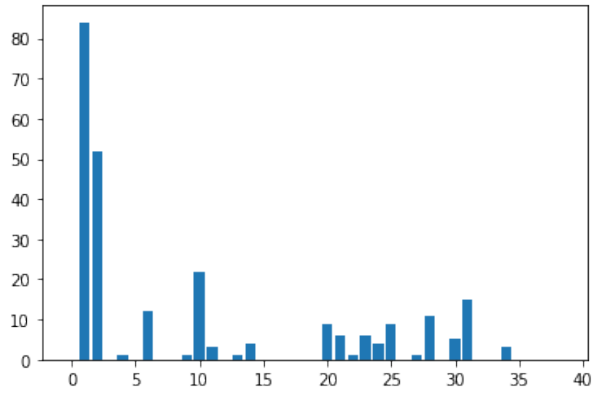

図 14 ベストモデル出現頻度

表 2 CNN-I を特徴量抽出器として用いた場合のクラス数と nDCG の比較

class	評価方法 1	評価方法 2
250	0.983	0.855
125	0.974	0.838
50	0.939	0.756
25	0.932	0.679

表 3 は，CNN-P を特徴量抽出器として用いてクラス数を変化させた場合の nDCG を比較したものである．CNN-I を特徴量抽出器として用いた場合と比較して，評価方法 1 と評価方法 2 のどちらにおいても nDCG の値は低下した．これより，本実験においては Imagenet で事前学習された ResNet50 から作成した特徴量抽出器から抽出した特徴量の方が，粒子画像の類似度を計測する上では，優れていたと言える．

CNN-P を特徴量抽出器として用いた際に，nDCG の値が低下した原因として，学習方法に問題があったのではないかと考えられる．CNN-P は ImageNet で事前学習された ResNet50 を粒子の画像で分類学習を行うことで作成された．このようなクラス分類の学習では，各クラス内と各クラス間のサンプル間の距離を考慮せずに全結合層で分離可能な特徴量の抽出を目的として学習される．クラス分類の学習では，訓練データに各クラスのサンプル数が十分含まれている場合は，高い性能を示すが，訓練サンプルの少ないクラスがある場合には適していない．本実験では，サンプル数が十分でなかったこと，各クラスのサンプル数に偏りがあったため，適した学習が行えなかったのではないかと考えられる．

次に，機械学習モデルの汎用性を考慮したランキングを出力とした場合と，CNN-I と CNN-P を特徴量抽出器として用いた場合の代表画像の枚数を 25 枚とした時の nDCG を比較したものを表 4 に示す．その結果，CNN-I を特徴量抽出器として用いた場合の，画像の類似度を用いた検索手法が最も優れた結果となった．

表 3　CNN-P を特徴量抽出器として用いた場合のクラス数と nDCG の比較

class	評価方法 1	評価方法 2
250	0.881	0.635
125	0.877	0.598
50	0.849	0.558
25	0.818	0.532

表 4　3 つの手法の比較

	評価方法 1	評価方法 2
CNN-I(k = 25)	0.932	0.679
CNN-P(k= 25)	0.818	0.532
boulder score	0.829	0.547

6　まとめと今後の課題

本研究では，予め結果を保有する代表画像とクエリ画像の類似度を計測することで，クエリ画像に適した機械学習モデルのランキングを得た．検索対象の機械学習モデルの汎用性を考慮したランキングを出力とした場合と比較して，機械学習モデルの検索の性能は向上した．

本研究の提案手法は，いかに優れた特徴量を画像から抽出できるかが重要である．そのため，機械学習モデルの検索の性能は特徴量抽出器に依存する．今後は，検索対象の機械学習モデルに画像を入力した際の，インスタンスセグメンテーションの性能という観点から，画像の特徴量を抽出できる特徴量抽出器の作成を発展として考えている．

謝　　辞

本研究は JSPS 科研費 JP21H03775，JP21H03774，JP21H03554，JP18H03244，JP22H03905，2022 年度国立情報学研究所公募型共同研究 21S1002 の助成を受けたものです．ここに記して謝意を表します．

文　　献

[1] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL'19*, pp. 4171–4186, 2019.

[2] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. In *Proceedings of the 2021 International Conference on Learning Representations*, pp. 1–22, 2021.

[3] Clement Farabet, Camille Couprie, Laurent Najman, and Yann LeCun. Billion-scale similarity search with GPUs. *IEEE Transactions on Big Data*, pp. 1915–1929, 2013.

[4] Ross Girshick. Fast R-CNN. In *Proceedings of the 2015 International Conference on Computer Vision*, pp. 1440–1448, 2015.

[5] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In *Proceedings of the 2014 Conference on Computer Vision and Pattern Recognition*, pp. 580–587, 2014.

[6] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask R-CNN. In *Proceedings of the 2017 International Conference on Computer Vision*, pp. 2961–2969, 2017.

[7] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. In *Proceedings of the 2016 Conference on Computer Vision and Pattern Recognition*, pp. 770–778, 2016.

[8] Jeff Johnson, Matthijs Douze, and Hervé Jégou. Learning Hierarchical Features for Scene Labeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 535–547, 2019.

[9] Alexander Kirillov, Kaiming He, Ross Girshick, Carsten Rother, and Piotr Dollar. Panoptic Segmentation. In *Proceedings of the 2019 Conference on Computer Vision and Pattern Recognition*, pp. 9404–9413, 2019.

[10] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft COCO: Common Objects in Context. In *Proceedings of the 2014 European Conference on Computer Vision*, pp. 740–755, 2014.

[11] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg. SSD: Single Shot MultiBox Detector. In *Proceedings of the 2016 European Conference on Computer Vision*, pp. 21–37, 2016.

[12] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. RoBERTa: A Robustly Optimized BERT Pretraining Approach.

[13] Yury A. Malkov and Dmitry A. Yashunin. Efficient and robust approximate nearest neighbor search using Hierarchical Navigable Small World graphs. *CoRR*, pp. 1–13, 2016.

[14] Marius Muja and David G. Lowe. Fast Approximate Nearest Neighbors with Automatic Algorithm Configuration. In *Proceedings of the 2009 International Conference on Computer Vision Theory and Applications*, pp. 331–340, 2009.

[15] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You Only Look Once: Unified, Real-Time Object Detection. In *Proceedings of the 2016 Conference on Computer Vision and Pattern Recognition*, pp. 779–788, 2016.

[16] Nils Reimers and Iryna Gurevych. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks . In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, pp. 3982–3992, 2019.

[17] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1137–1149, 2017.

[18] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. Imagenet large scale visual recognition challenge. *Int. J. Comput. Vision*, Vol. 115, pp. 211—-252, 2015.

[19] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the Inception Architecture for Computer Vision. In *Proceedings of the 2016 Conference on Computer Vision and Pattern Recognition*, pp. 2818–2826, 2016.

[20] Yosuke Ueno and Masaaki Kondo. A Base Model Selection Methodology for Efficient Fine-Tuning. In *Proceedings of the 2020 International Conference on Learning Representations*, pp. 1–12, 2020.