

言語的な階層構造を用いた一般物体認識

下田 誠† 藤江 真也†

†千葉工業大学工学研究科未来ロボティクス専攻

1 はじめに

一般物体認識は、写真のような実世界のシーン画像に含まれる物体を、普段我々が使用する一般的な名称で答えるタスクであり、近年盛んに研究されている。従来の一般物体認識では識別したいカテゴリ群に対して、単一の識別器のみを用いる方法が主である。そのため、例えば「識別結果のニンジンが根菜であり、野菜の一種」のような、カテゴリの持つ上位概念を識別過程で獲得することはできない。また、対象とするカテゴリ数が大規模となった場合、単一の識別器のみを用いる方法では学習やチューニングに多大な計算的、時間的コストが必要であり、拡張性が高いとは言えない。

本研究ではカテゴリの持つ上位概念を取得しつつ、大規模識別への拡張性を得るため、言語的階層構造に沿った一般物体認識手法を提案する。提案手法は、異なる識別タスクを持つ複数の識別器を階層的に組み合わせることで実現させる。

2 提案システム概要

本研究で行う識別の例を図1に示す。本手法で利用する言語的階層は木構造で表される。このうち、根ノード、中間ノードに子ノードを出力とする識別器を割り当てる。各識別器が出力する子ノードの確率値を利用して、根ノードから葉ノードまでのスコアを計算する。このスコアが最も高い経路上のノードを、最終的な識別結果として出力する。

3 階層的な識別器の構築

3.1 WordNet を用いた階層構造の構築

WordNet は名詞、動詞などに関する英語の概念辞書型データベースである [1]。各単語は類義語のセットとしてグループ化されており、この類義語に基づいて単語や概念同士の関係性が記述されている。

そこで、階層構造の構築者が設定したカテゴリ名に対して下位語を取得し、得られた語に対しても同様な操作を再帰的に行うことで、設定したカテゴリを根ノードとした言語階層を取得する。本研究では階層の根ノードと

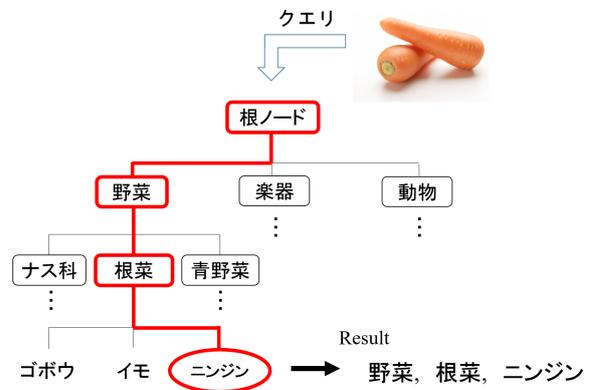


図1: 提案システムにおける識別の流れ。各階層の識別を統合した結果を見ることで、ニンジンという名称だけでなく、野菜、根菜といった上位概念を得ることができる。

して「vegetable」を設定した。その結果、合計 176 個のノードで構成された言語階層構造が取得できた。

3.2 ImageNet からの画像取得

ImageNet は、WordNet の概念構造に沿ったカテゴリ階層で画像データを保持している大規模画像データベースである [2]。ここから本研究で識別対象カテゴリに対応する画像を収集したところ、合計で 134,518 枚の画像が得られた。

しかしながら、カテゴリの中には取得した画像データ数が 10 枚未満のものや、識別対象とするカテゴリ間で極端なデータ数の偏りが見られた。そのため、カテゴリの持つ画像数に基づき、階層構造に利用するカテゴリを選定した。選定の基準として、本研究では 100 枚以上、かつ、識別対象カテゴリ間の平均の半数以上の画像を持つカテゴリを階層構造のノードとして利用する。

カテゴリ選定の結果、最終的に全ノード数が 92、葉ノード数が 62、階層の最大深さが 6 層の言語階層構造を取得した。このうち、識別器が必要なノードは図2の赤色で示す 28 個となった。

3.3 識別器学習

識別器の学習には Convolutional Neural Network (CNN) を利用する。CNN は主に画像データを扱う際に用いられること多い順伝播型ネットワークで、一般物体認識

General object recognition with semantic hierarchical structure

†Makoto SHIMODA †Shinya FUJIE

†Dept. Advanced Robotics, Chiba Institute of Technology

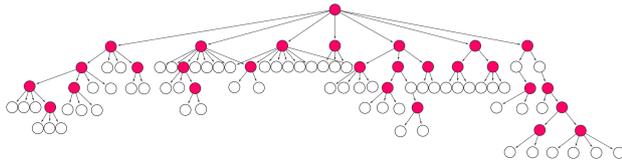


図 2: 最終的に取得した階層構造

タスクにおいても高い性能が示されている．ネットワークのモデルは，5 層の畳み込み層と 3 層の全結合層で構成される AlexNet を採用した [3]．出力層は Softmax 関数であり，この出力値を確率値として解釈し，経路決定時のスコア計算に利用する．

また，本研究では下位カテゴリ間の識別に用いる識別器を学習する際，既に学習済みの上位識別器から重みのパラメータをコピーし，初期値として利用する転移学習を利用する．転移学習は，特に識別対象の特徴が似ているタスク間で行うことで学習の収束，識別性能の面で大きな効果があるとされている [4]．予備実験として，重みの初期値をランダムに決定したものと，別の識別器からコピーしたものとで学習の比較を行った．その結果，転移学習を行った方が学習の収束性，テストデータへの正解率といった観点で優れていることが分かった．

3.4 経路決定法

最終的な識別結果は，各識別器から出力された確率値に基づいたスコアによって決定する．一般的には，根ノードからある葉ノードへ到達する確率は，経路内に存在する枝ノードがその経路を選択する確率値の積によって計算される．しかし，本手法によって構築した階層構造では葉カテゴリによって階層の深さが異なる．そのため，そのまま確率値の積をスコアとして適用すると，階層の深いカテゴリであるほど不利な結果となってしまう．

そこで，本研究では確率値の積を相乗平均した値をスコアとして扱う． k 番目の葉ノードに到達するまでのノードの集合を C_k ，その一つの要素 i が識別器で選択される確率を p_i とすると，最終的な葉ノード k_{pred} は以下の式で選択される．

$$k_{\text{pred}} = \arg \max \left\{ \left(\prod_{i \in C_k} p_i(x) \right)^{\frac{1}{|C_k|}} \right\} \quad (1)$$

4 性能評価実験

提案手法によるカテゴリの正解率を評価する実験を行った．テストデータは，葉ノードのカテゴリから 50 枚ずつ，合計 3,100 枚の画像データを使用した．比較対象として，単一の識別器に関しても評価を行った．二つ

表 1: テストデータの識別結果

識別手法	Top-1 [%]	Top-3 [%]	Top-5 [%]
単一の識別器	46.3	69.7	78.2
提案手法	33.2	49.1	55.9

表 2: 深さ別の正解率

	Depth					
	1	2	3	4	5	6
Top-1 [%]	64.5	44.6	37.6	33.0	38.0	40.0
Top-3 [%]	77.1	58.0	52.3	52.1	56.7	67.0
Top-5 [%]	82.9	64.1	58.7	58.7	64.9	83.3

の手法で，葉ノードのカテゴリに対する正解率を表 1 に示す．Top-1 は最も確率の高いカテゴリ，Top-5 は上位 5 カテゴリ内で正解カテゴリが存在した場合に正しく認識されたものとする．

単一の識別器の正解率と比較すると，提案手法は Top-1 の識別では 13.1 ポイント，Top-5 では 22.3 ポイント低い結果となった．ここで，カテゴリがどの深さの識別で誤りを起こしているか調べるため，深さ毎の正解率を計算した．正解カテゴリが持つ経路のうち，識別結果がどの経路まで正しく識別できたかを割合で表すと，表 2 のようになった．最も深い第 6 層の正解率よりも 3 層から 5 層の正解率が低いことから，これらの層に存在する葉ノードに対する識別器の性能が低いことが予想される．

5 まとめ

本稿では，言語的な階層構造を用いることで識別時にカテゴリの持つ上位概念を取得可能な一般物体認識手法を提案した．野菜に関するカテゴリの識別タスクでは，Top-1 の識別で 33.2%，Top-5 の識別で 55.9% の識別性能を示した．今後の課題として，各識別器のチューニング，経路探索時のスコア計算法の改良などが挙げられる．また，大規模な識別タスクにおいて，提案手法がどの程度有効であるかを評価したい．

参考文献

- [1] Miller, George A. "WordNet: a lexical database for English." *Communications of the ACM* 38.11 (1995): 39-41.
- [2] Deng, Jia, et al. "Imagenet: A large-scale hierarchical image database." *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on.* IEEE, 2009.
- [3] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems*. 2012.
- [4] Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks?. In *Advances in neural information processing systems* (pp. 3320-3328).