

## 複数階層においてカテゴリ推論を行うニューラルネットワークモデルの検討

A Study on Neural Network Models for Category Inference in Multiple Hierarchies

306215 渡邊 響輝 [加瀬澤研究室]

### 1. まえがき

近年、機械学習の技術向上により、画像中の物体認識技術は目覚ましく向上している。

しかしながら、そのほとんどは既に学習された物体を分類するための技術・手法であり、我々が日常的に行っている、未知の物体についての分類問題は未解決のままである。

そのため、このような問題の解決を目的として、最近の研究では、画像を用いて、複数の階層において同時に認識するための手法に関する検討が進められている。<sup>[1]</sup>この手法が確立できれば、未知の物体であっても少ない画像からその属性・カテゴリを認識することができる。

そこで、本研究では、複数の階層において同時に物体のカテゴリを推論するための手法について検討する。

### 2. 提案手法

物体は階層的なカテゴリ構造を持っている。猫を例に取って考えるとき、そのカテゴリの構造は、たとえば図1のように表すことができる。

本研究では、図中の最下層、すなわち「猫」にあたる階層を下位クラスあるいは小分類と呼ぶ。また、下位クラスよりも上位にある階層をすべて上位クラスと呼び、その中でも「動物」、「哺乳類」にあたる階層をそれぞれ大分類、中分類と呼ぶ。

提案手法は、このような階層的なカテゴリ構造の大分類・中分類・小分類の3つの階層において同時にカテゴリ推論を行うニューラルネットワークモデルである。

提案するネットワークモデルの構成を図2に示す。提案モデルはCNNをベースとしており、始めに小分類についてのカテゴリ推論結果を得る。その後、プーリングを介して得られるサイズの小さな抽象度の高い特徴マップに基づき、中分類、大分類の順にカテゴリ推論結果を得る。

### 3. 実験

#### 3.1 実験概要

実験では、まず、提案モデルの学習を行う。その後、学習には使用していない未知物体を集めたテスト用のデータセットを当該モデルに入力し、大分類・中分類階層における認識性能を検証する。

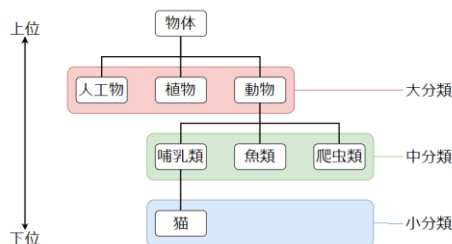


図1 階層的なカテゴリのイメージ

また、既存のネットワークモデルであるVGG16を転移学習したモデルを性能比較の対象として、同様の実験を行う。

#### 3.2 実験条件

実験で想定する階層構造を表1、表2に示す。図に示されるように、小・中・大分類に属するクラス数はそれぞれ70、14、3である。また、実験では、小分類クラスごとに

表1 実験で想定する階層構造

大	中	小分類		
①	①	ボーダーテリア, ベルシアン猫, バグ, 豚, シマウマ		
	②	コモドドラゴン, ナイルワニ, ニシキヘビ, アカウミガメ, エリマキトカゲ		
	③	金魚, シビレエイ, オニカマス, ホホジロザメ, クマノミ		
	④	クジャク, ワシ, ニワトリ, ダチョウ, コンドル		
	⑤	コガネグモ, テントウムシ, ハエ, バッタ, トンボ		
②	⑥	パイナップル, レモン, ザクロ, イチジク, バナナ		
	⑦	キャベツ, キュウリ, イチゴ, トウモロコシ, ピーマン		
	⑧	キノコ, マイタケ, ハラタケ, シロソウメンタケ, スッポンタケ		
	⑨	ヒナギク, たんぽぽ, バラ, ひまわり, チューリップ		
③	⑩	刑務所, 高架橋, レストラン, 灯台, 噴水		
	⑪	新幹線, マウンテンバイク, 旅客機, ミニバン, タクシー		
	⑫	本棚, 四柱式ベッド, 机, ロッキングチェア, 扇風機		
	⑬	シャベル, ハンマー, 手斧, 定規, ボールペン		
	⑭	Tシャツ, ジーンズ, ブラジャー, セーター, ジャケット・コート		
中分類		大分類		
①	哺乳類	⑩ 建築物	①	動物
②	爬虫類	⑦ 野菜	⑪	乗り物
③	魚類	⑧ キノコ	⑫	家具
④	鳥類	⑨ 花	⑬	道具
	昆虫		⑭	衣類

表2 テスト用画像詳細

大	中	小分類
①	①	キットギツネ, レッサーパンダ, ゴリラ
	②	イモリ, キスイガメ, ウィップテール・リザード
	③	ウナギ, ダツ, アカエイ
	④	フクロウ, メキシコカワガラス, 鴨
	⑤	ハンミョウ, オサマシ, 蜂
②	⑥	オレンジ, グラニースミス, パラミツ
	⑦	ブロッコリー, スズキニ, バターナッツかぼちゃ
	⑧	シャグマアミガサタケ, ボルチーニ, テングタケ
	⑨	アジサイ, ハイビスカス, ユリ
③	⑩	納屋, ダム, 城
	⑪	飛行船, モーターボート, リムジン
	⑫	食器棚, 折り畳みいす, 洗濯機
	⑬	傘, モップ, カップ
	⑭	毛皮のコート, トレーナー, ポンチョ

訓練用画像約 1300 枚, テスト用画像 50 枚を使用する。  
ここで, サイズは  $112 \times 112$  であり, カラー画像を用いる。

提案モデルを図 2 に, VGG16 を転移学習したモデルを図 3 に示す。また, 学習パラメータは, エポック数 32, Validation Split 0.2, バッチサイズが提案モデルで 64, VGG16 ベースモデルで 32 である。

ここで, VGG16 を転移学習したモデルは, 全結合層を GAP と Dense, softmax による出力に差し替えたものであり, 各分類の出力はすべて最後に行っている。

3.3 実験結果

未知物体の画像を入力したときの, 中, 大分類における正答率, 混同行列を図 4, 5 に示す。図において, 縦・横軸がそれぞれ正解カテゴリと得られた推定カテゴリである。また, 背景が緑のセルは推定が正解している箇所, 青のセルは正解の 60% 以上の回数推定された不正解の推定である。

4. 考察

畳み込み・プーリングが行われるにつれて上位クラスの持つ特徴を抽出しやすくなると考えていたが, 図 4, 5 に示されるように, 順次推定結果を得ていく提案モデルは, 最後に全ての推定結果を得る VGG16 ベースモデルより正答率が低いという結果となった。この原因として, 提案

モデルは, VGG16 ベースモデルと比較して層が浅いことが考えられる。しかしながら, 今回使用した両モデルは層の深さが異なるため, ネットワークモデルの構成上の優位性を示すまでには至っていない。

一方, 図 5 において, 再現率の低いクラスが複数ある。例えば, 両モデルにおいて, 果物を野菜であると数多く推定している。一般に, 我々が用いる物体の階層的なカテゴリ構造は, 必ずしも物体の外見的特徴に基づいているわけではない。そのため, 前述したような果物と野菜を外見だけで正確に分類するためには, 階層的な構造の見直しや, 大量のデータを学習させる必要があると考えられる。

6. むすび

本研究では, 複数階層において同時に物体のカテゴリ推論を行うための手法についての検討を行った。

今回行った実験によれば, 畳み込み・プーリングが行われるにつれて上位クラスの持つ特徴が抽出しやすくなるわけではなく, 我々が日常的に使用している物体の階層的な構造をそのままカテゴリ分類に利用して良いとは限らないという知見が得られた。今後の課題としては, 各階層における分類の出力・推定を得る位置の検討など, ネットワークモデルの見直し構成の見直しが挙げられる。

参考文献

[1] H. Liu, et al., “What is a Tabby? Interpretable Model Decisions by Learning Attribute-Based Classification Criteria”, IEEE Trans on PAMI, vol.43, No.5, May 2021.

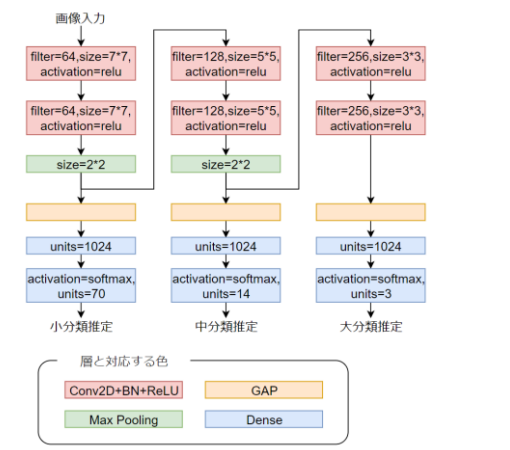


図 2 提案ネットワークモデルの構成

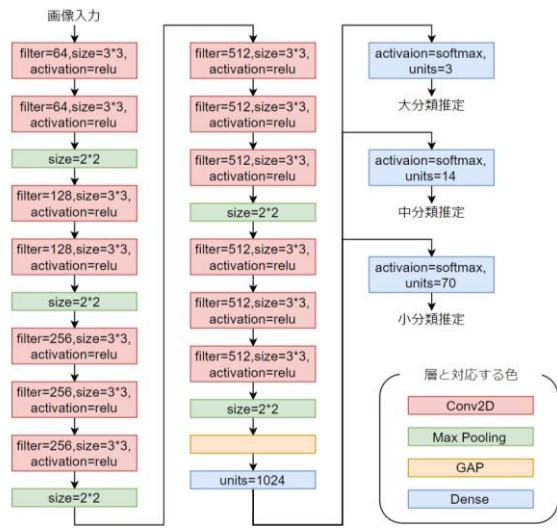


図 3 VGG16 を転移学習したモデルの構成

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
(1)	25	8	15	32	3	1	1	53	3	5	0	0	4	0
(2)	2	16	11	2	35	11	7	36	8	5	1	3	13	0
(3)	1	24	63	4	8	6	6	7	4	10	3	2	12	0
(4)	5	12	21	26	13	3	3	27	2	13	12	3	5	0
(5)	0	5	3	14	47	17	8	13	17	4	1	2	18	1
(6)	0	5	14	9	6	57	27	8	17	1	0	0	6	0
(7)	0	2	2	2	4	35	57	11	5	5	2	5	16	4
(8)	1	3	1	1	5	6	3	124	3	0	0	1	2	0
(9)	0	0	1	11	2	7	17	6	97	1	2	0	0	6
(10)	4	3	3	8	3	2	2	2	1	100	14	3	5	0
(11)	0	1	18	4	7	2	1	0	1	10	80	1	25	0
(12)	2	0	1	3	4	0	1	2	2	25	12	60	30	8
(13)	4	4	5	7	11	4	12	2	8	11	3	13	51	15
(14)	2	1	3	13	7	3	8	10	3	14	1	12	46	27

(a)提案モデル(正答率:39.5%)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
(1)	69	33	6	23	5	2	0	4	1	1	1	1	2	2
(2)	2	51	29	6	16	4	6	21	1	2	4	2	6	0
(3)	0	51	54	2	8	5	9	6	3	2	2	1	5	2
(4)	6	42	15	33	18	6	4	11	2	3	5	0	1	4
(5)	4	23	7	9	77	4	5	9	3	2	3	3	1	0
(6)	0	1	1	3	0	97	43	4	0	0	0	0	0	1
(7)	1	3	2	6	2	52	51	9	2	1	1	1	11	8
(8)	5	4	6	8	2	4	2	113	0	0	1	0	3	2
(9)	0	0	2	0	12	13	31	18	74	0	0	0	0	0
(10)	7	10	2	11	2	0	1	1	2	84	10	7	2	11
(11)	1	11	33	3	3	4	0	0	0	1	70	3	21	0
(12)	0	0	0	0	2	0	0	0	0	20	1	98	16	13
(13)	1	2	14	0	11	15	6	2	4	5	1	25	42	22
(14)	2	5	4	1	2	2	8	5	1	5	0	8	4	103

(b)VGG16(正答率: 48.4%)

図 4 中分類の混同行列

	(1)	(2)	(3)
(1)	597	234	119
(2)	58	493	49
(3)	134	81	583

(a)提案モデル  
(正答率 67.9%)

	(1)	(2)	(3)
(1)	585	109	56
(2)	56	513	31
(3)	112	51	587

(b)VGG16  
(正答率: 80.2%)

図 5 大分類の混同行列