

画像認識を目的とした複数の深層ニューラルネットワーク の索引付け手法

A Method for Indexing Deep Neural Networks for Image Recognition

Nailussa'ada¹ Fazlur Rahman Bintang¹ Tri Harsono¹ Ali Ridho Barakbah¹ 鷹野 孝典²
Nailussa'ada¹, Fazlur Rahman Bintang¹, Tri Harsono¹, Ali Ridho Barakbah¹, and Kosuke Takano²

¹ スラバヤ電子工学ポリテクニク

¹ Politeknik Elektronika Negeri Surabaya

² 神奈川工科大学 情報学部 情報工学科

² Department of Information and Computer Sciences, Faculty of Information Technology,
Kanagawa Institute of Technology

Abstract: The high capability of representation learning by deep neural networks (DNNs) is making a big impact on both academic and industry area in many services from face recognition to medical diagnosis. Basically, they are independently developed for recognizing some patterns in the specific domain that their systems or services require, and users cannot correctly decide which DNN is better to use for their purposes. Therefore, the technology of managing DNNs would be essential role for using them more efficiently. In this study, we propose a method for indexing deep neural networks for image recognition. Our method selects proper DNNs from a set of DNNs to answer the query image. The selection process will be based on what category the query is and which DNN that have the ability to answer it. In the experiments using multiple DNNs with different domains of image recognition, we evaluate the feasibility of our proposed method.

1. はじめに

画像認識分野において、深層学習が大きな成果を達成している[1][2]。深層学習は高い表現学習能力により、表情認識[7]から医療診断まで多岐に渡る分野の様々なサービスに適用されつつある[6][8]。

画像認識を行う深層ニューラルネットワーク（以下、DNN）が対象とするパターン認識領域は、適用されるシステムやサービスが対象とする領域を満たすように決定され、各 DNN はそのパターン認識領域に応じた画像集合を用いた学習により独立に構築される。また、大量の画像データを所有していなくても、事前学習済みの DNN[10][11][12]を用いた転移学習により、一般ユーザは自身の画像データを用いてパーソナルな DNN を構築することができる。今後、他の利用者が再利用できるような形式で、一般ユーザが作成した DNN もインターネット上で普及していくことが想定される。

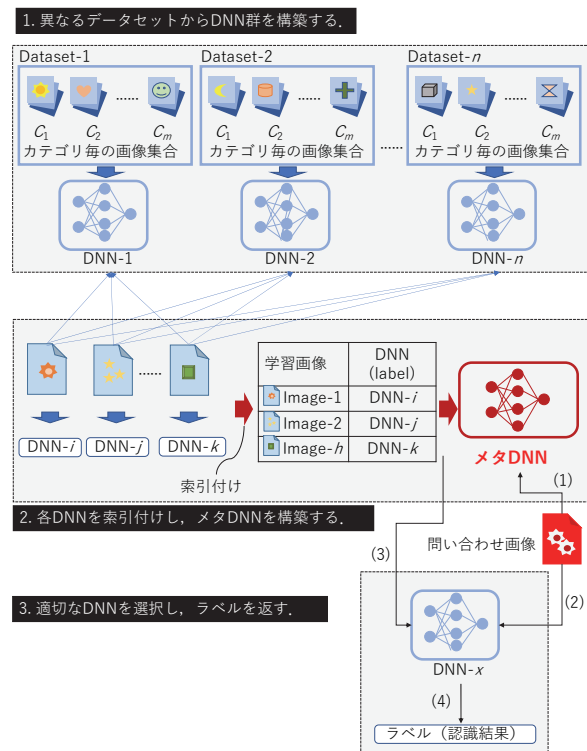
このような想定のもと、多様なパターン認識領域を持つ DNN がインターネット上で公開される場合、利用者は目的に沿ってどのパターン認識領域を持つ

DNN を利用するのが適切であるかを決定する必要がある。その際、DNN を管理することにより、利用者が効率的に DNN を選択し、利用できる仕組みが重要となる。

本研究では、画像認識を目的とした複数の深層ニューラルネットワークの索引付け手法を提案する。提案方式は、複数の DNN のうち、問い合わせとして入力した画像に対して認識結果を応答するものを選択する。選択された DNN は、問い合わせ画像に対して、正解となるラベルを認識結果として返す。提案方式の特徴は、このような機能を実現するために、各 DNN が対象とする領域の画像集合を用いて、高次の DNN（以下、メタ DNN）が深層学習することにより各 DNN の索引付けを行う点にある。深層学習を適用した索引付け（学習）を行うため、大量の画像データを用いて学習することにより、目的の DNN を高精度で選択する機能を実現することが可能となる。

本研究では、異なる領域の画像認識をする複数の深層ニューラルネットワークを用いた実験により、提案方式の実現可能性を検証する。

2. 提案方式



提案方式の概要を図 1 に示す。提案方式は、異なる認識領域を持った DNN 群を、メタ DNN により索引付けを行うことで、問い合わせ画像に対して適切な認識結果（ラベル）を返す DNN を選択する点を特徴とする。提案方式は、次の手順で実行される。

1. 異なる画像データ集合からの DNN 群の構築
2. DNN 群のメタ DNN による索引付けと学習
3. 問い合わせ画像に応じた DNN 選択
4. 選択された DNN による認識結果の提示

以下の節で各手順の詳細について説明する

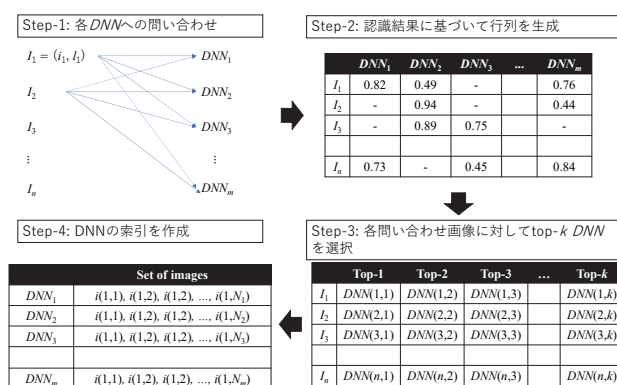
2. 1 DNN 群の構築

目的とするパターン認識領域の画像データ $Dataset_n$ を用いて、DNN 群 $DNN_1, DNN_2, \dots, DNN_n$ を独立に構築する。

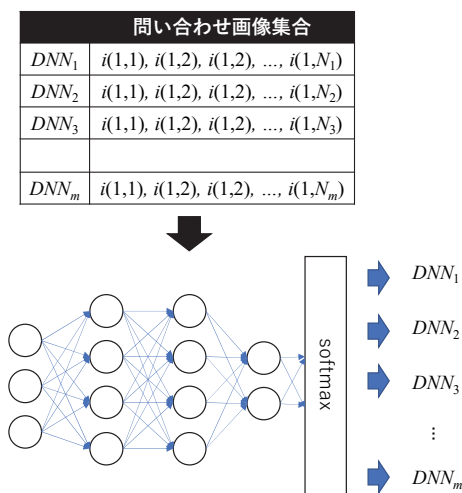
2. 2 DNN 群のメタ DNN 上での索引付け

このプロセスでは、まず、各 DNN を索引付けする画像集合を下記の手順で抽出する（図 2）。

- Step-1: 問い合わせ画像 $I_x = (i_x, l_x)$ を DNN_i に入力し、認識結果とスコアを得る。
- Step-2: 認識結果に基づいて、問い合わせ画像 I_x と DNN_i の関係を表する行列を生成する。ここで、行列の各値は DNN_i が返す最も高い値である。
- Step-3: 問い合わせ画像 I_x に対して、高スコアの認識結果を返した top- k DNN が選択される。
- Step-4: 画像 i_x が選択された k 個の DNN_i に関連付けられる。
- Step-1 から Step-4 の手順を全ての問い合わせ画像について実施することにより、DNN の索引を作成する。



次に、作成した DNN 索引に基づいて、DNN と関連付けられた画像集合を深層学習により学習することで、メタ DNN を生成する（図 3）。メタ DNN は、問い合わせ画像が入力されたときに、適切な認識結果を返す DNN を選択する機能を持つ。



2. 3 問い合わせ画像に応じた DNN 選択

このプロセスでは、メタ DNN は入力された問い合わせ画像に対して、適切な認識結果を返す DNN を、下記の手順で選択する (図 4)。

- Step-1: 問い合わせ画像 q_{i_x} をメタ DNN に入力する。
 Step-2: メタ DNN は、各 DNN に対して問い合わせ画像 q_{i_x} へのスコアを算出する。
 Step-3: 問い合わせ画像 q_{i_x} に対して高スコアであった top- p DNN $DNN(x, 1), DNN(x, 2), \dots, DNN(x, p)$ が選択される。

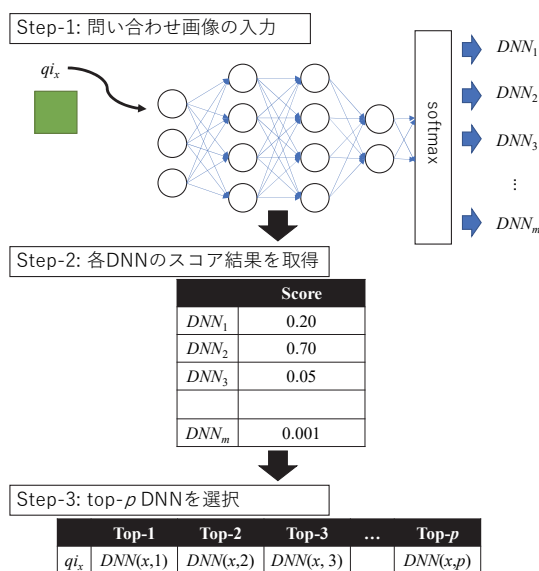


図 4 DNN の選択プロセス

2. 4 選択された DNN からの認識結果取得

問い合わせ画像 q_{i_x} に対して、認識結果を返す最終プロセスについて説明する。このプロセスでは、選択された p 個の DNN に問い合わせ画像 q_{i_x} を入力し、各 DNN から認識結果として得られたラベルのスコアを算出し、ランキングして提示する (図 5)。

- Step-1: 問い合わせ画像 q_{i_x} が選択された p 個の DNN $DNN(x, 1), DNN(x, 2), \dots, DNN(x, p)$ に入力される。
 Step-2: $DNN(x, 1), DNN(x, 2), \dots, DNN(x, p)$ は問い合わせ画像 q_{i_x} に対する認識結果としてラベル $object_j$ を返す。
 Step-3: ラベル $object_j$ の最終スコア $Final_score(object_j)$ を、各 DNN が算出したスコアの重み付き総和として下記の式で算出する。

$$Final_score(object_j) = \sum DNN_score \times Recognition_score(object_j)$$

ここで、 DNN_score はメタ DNN が問い合わせ画像 q_{i_x} に対する DNN スコアとして返す値である。また、 $Recognition_score(object_j)$ はラベル $object_j$ について算出されたスコアである。

- Step-4: 問い合わせ画像 q_{i_x} に対して、Step-3 で算出された最終スコアに基づいてラベル $object_j$ がランキングされて提示される。

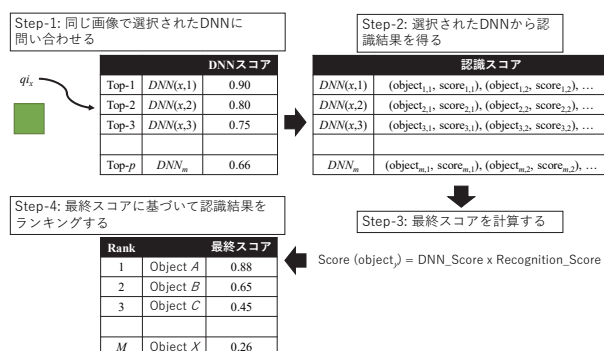


図 5 DNN からの認識結果取得プロセス

3. 実験

実験では、プロトタイプとして構築した 10 個の DNN 群およびそれらを索引付け学習して生成したメタ DNN を用いて、提案方式の実現可能性を検証する。

表 1 DNN のパターン認識領域

DNN	認識領域
1	Fruits (apple, cherries, mango, orange, pineapple, strawberry)
2	Animals (crocodile, dolphin, elephant, fox, kangaroo, leopard, lion, lobster, mouse, rabbit)
3	Trees (maple, oak, palm, pine, willow)
4	Transportations (bicycle, bus, motorcycle, pickup truck, streetcar, tank, train)
5	Flowers (orchid, bluebell, buttercup, rose, sunflower, tulip)
6	Fish (aquarium fish, dolphin, shark, trout, flatfish)
7	Stationary (Pen, book, eraser, paper clip, sticky notes)
8	Fruits, Animals & Flowers
9	Transportations, fruits, flowers
10	Transportation, trees

3. 1 実験環境

表 1 に 10 個の DNN のパターン認識領域を示す。学習用の画像データには、CIFAR-100 データセット [13] および検索エンジンで収集した画像を用いた。各カテゴリは、200 から 500 枚の画像からなる。これらの画像データを用いて、25 層からなる DNN を構築した。

また、DNN 索引付け用の画像を別途収集した。索引付け用画像と正解ラベルの例を表 2 に示す。これらの索引付け用画像を問い合わせとして各 DNN に入力した際に、スコアが高かった上位 3 件の DNN を抽出した。各索引付け用画像により抽出した DNN の例を表 3 に示す。さらにこの結果に基づいて、表 4 および図 6 の例に示すように、各 DNN の索引付けに用いる画像集合を抽出した。

表 2 索引付け用画像と正解ラベルの例

画像 ID	画像名	正解ラベル
1	00000000.png	apple
2	00000001.jpg	apple
3	00000002.jpg	apple
4	00000003.jpg	apple
5	00000004.jpg	apple
6	00000001.jpg	Aquarium fish
7	00000002.jpg	Aquarium fish
8	00000003.jpg	Aquarium fish
9	00000004.jpg	Aquarium fish
10	00000005.jpg	Aquarium fish
11	00000000.png	bicycle
12	00000001.png	bicycle
13	00000002.jpg	bicycle
14	00000003.png	bicycle
15	00000004.png	bicycle

3. 2 実験方法

20 個の問い合わせ画像を用いて、各問い合わせ画像に対して、(1)メタ DNN が適切な DNN を選択可能であるか、および(2)選択した DNN が返す認識結果に基づいて適切なラベルを提示可能であるか、について評価を行う。

3. 3 実験結果

表 5 に 20 個の問い合わせ画像に対して、メタ DNN による DNN の選択結果を示す。また、表 6 に各問い合わせ画像の認識結果を示す。

表 3 索引付け用画像により抽出した DNN の例

画像 ID	DNN		
	1st DNN ID (スコア)	2nd DNN ID (スコア)	3rd DNN ID (スコア)
1	DNN8 (0.901)	DNN9 (0.837)	DNN3 (0.820)
2	DNN8 (0.995)	DNN9 (0.991)	DNN2 (0.692)
3	DNN9 (0.998)	DNN8 (0.998)	DNN3 (0.925)
4	DNN8 (0.991)	DNN9 (0.989)	DNN3 (0.691)
5	DNN9 (0.932)	DNN5 (0.852)	DNN1 (0.743)
6	DNN9 (0.88)	DNN2 (0.821)	DNN10 (0.70)
7	DNN1 (0.911)	DNN2 (0.670)	DNN5 (0.60)
8	DNN2 (0.990)	DNN9 (0.794)	DNN3 (0.780)
9	DNN1 (0.900)	DNN6 (0.816)	DNN8 (0.673)
10	DNN9 (0.860)	DNN6 (0.822)	DNN3 (0.584)
11	DNN1 (0.774)	DNN6 (0.738)	DNN7 (0.724)
12	DNN10 (0.916)	DNN7 (0.890)	DNN1 (0.831)
13	DNN10 (0.995)	DNN3 (0.934)	DNN6 (0.869)
14	DNN2 (0.896)	DNN10 (0.864)	DNN7 (0.847)
15	DNN10 (0.899)	DNN7 (0.834)	DNN1 (0.739)

表 4 各 DNN への索引付けに用いた画像集合の例

DNN	画像集合
1	apple_00000.jpg, apple_00001.jpg, cherries_0000.jpg, ...
2	lion_00000.jpg, lobster_00001.jpg, crocodile_0000.jpg, ...
3	oak_00000.jpg, maple_00001.jpg, palm_0000.jpg, ...
4	bus_00000.jpg, pickup_00001.jpg, truck_0000.jpg, ...
5	orchid_00000.jpg, rose_00001.jpg, tulip_0000.jpg, ...
6	dolphin_00000.jpg, shark_00001.jpg, trout_0000.jpg, ...
7	pen_00000.jpg, book_00001.jpg, eraser_0000.jpg, ...
8	orange_00000.jpg, apple_00006.jpg, leopard_0000.jpg, ...
9	bus_00005.jpg, streetcar_00004.jpg, bus_0010.jpg, ...
10	bicycle_00002.jpg, oak_00018.jpg, bicycle_00041.jpg, ...



図 6 画像集合の例

表 5 DNN の選択結果

問い合わせ画像 ID	正解 DNN	選択 DNN	DNN スコア	判定
1	4 or 9	9	0.875	Correct
2	5 or 9	9	0.23	Correct
3	1	8	0.40	Incorrect
4	7	7	0.55	Correct
5	4 or 10	10	0.962	Correct
6	1	1	0.873	Correct
7	4	4	0.747	Correct
8	3 or 10	10	0.645	Correct
9	6	6	0.931	Correct
10	5	5	0.998	Correct
11	5	8	0.25	Incorrect
12	4 or 9	4	0.254	Correct
13	2 or 8	2	0.99	Correct
14	6	6	0.681	Correct
15	3	3	0.51	Correct
16	3 or 10	9	0.29	Incorrect
17	7	7	0.86	Correct
18	1	1	0.59	Correct
19	4 or 10	4	0.633	Correct
20	2	2	0.68	Correct

表 6 画像の認識結果

問い合わせ画像 ID	正解ラベル	結果ラベル	スコア	判定
1	Streetcar	Streetcar	0.82	Correct
2	Sunflower	Sunflower	0.615	Correct
3	Mango	Orange	0.933	Incorrect
4	Paper clip/pen/book	Paper clip	0.386	Correct
5	Bicycle	Bicycle	0.993	Correct
6	Strawberry	Strawberry	0.994	Correct
7	Motorcycle	Motorcycle	0.96	Correct
8	Oak tree	Oak tree	0.62	Correct
9	Dolphin	Dolphin	0.93	Correct
10	Bluebell	Bluebell	0.998	Correct
11	Orchid	Rose	0.49	Incorrect
12	Tank	Tank	0.25	Correct
13	Fox	Fox	0.902	Correct
14	Trout	Trout	0.439	Correct
15	Pine tree	Pine tree	0.78	Correct
16	Maple tree	Sunflower	0.85	Incorrect
17	Book or pen	Book	0.57	Correct
18	Mango	Mango	0.98	Correct
19	Train	Train	0.59	Correct
20	Elephant	Elephant	0.96	Correct

表 5 の結果から, 20 個の問い合わせ画像に対して, 85% の正答率で適切な DNN を選択できたことが確認できる. 問い合わせ画像 ID が 3, 11, 16 の問い合わせについては正しい DNN を選択することができなかった. ただし, 不正解であった DNN の DNN スコアはいずれも 0.5 以下であった. このため, DNN スコアが最上位であったため選択はされているものの, 問い合わせ画像との関連性は低いと考えられる.

また, 表 6 の結果から, 85% の正答率で正しい認識ラベルが抽出されたことが確認できる. ただし, 表 5 の結果で正しい DNN が選択できなかった問い合わせ画像 ID が 3, 11, 16 について, 同様に正しい認識ラベルを得ることができなかった. これは, 選択された DNN が, 問い合わせ画像が属するカテゴリを含んでいなかったためである. 例えば, 問い合わせ画像 ID が 3 の画像はマンゴーが正しいラベルであるが, 認識結果はオレンジであった. これは選択された DNN-8 が「マンゴー」というラベルが存在してないために認識不能であった結果である.

以上の結果より, 提案方式により構築したメタ DNN が適切な DNN を選択することが可能であり, 選択した DNN が返す認識結果に基づいて適切なラベルを提示可能であることを確認することができた.

4. まとめ

本研究では, 画像認識を目的とした複数の深層ニューラルネットワークの索引付け手法を示した. また実験では, 異なる領域の画像認識を行う複数の深層ニューラルネットワークを索引付け学習することによりメタ DNN を構築し, 提案方式の実現可能性を検証した. 実験結果から, 問い合わせ画像に対して 85% の精度で適切な DNN を選択できることが確認できた. また, 適切な DNN を選択できた場合は, 正しい認識ラベルを提示することも確認できた.

今後の課題として, 索引付けをする DNN の規模を拡大し, 問い合わせ画像の認識精度, 実行時間, およびスケーラビリティに関する評価を行い, 提案方式の有効性を検証していく予定である.

参考文献

- [1] D. C. Cireş, U. Meier, J. Masci, L. M. Gambardella and J. Schmidhuber: High Performance Convolutional Neural Networks for Image Classification, Proceedings of 22nd International Joint Conference on Artificial Intelligence, pp. 1237-1242, AAAI Press, (2011)
- [2] R. K. Srivastava, K. Greff and J. Schmidhuber: Training Very Deep Networks, Proceedings of 28th International

- Conference on Neural Information Processing Systems, pp. 2377-2385, (2015)
- [3] J. Schmidhuber: Deep learning in neural networks: An overview", Neural Networks, vol. 61, pp. 85-117, (2015)
 - [4] D. C. Cireş, L. M. Gambardella, A. Giusti and J. Schmidhuber: Deep Neural Networks Segment Neuronal Membranes in Electron Microscopy Images, Proceedings of the 25th International Conference on Neural Information Processing Systems, vol. 2, pp. 2843-2851, (2012)
 - [5] L. Deng: A Tutorial Survey of Architectures, Algorithms, and Applications for Deep Learning, APSIPA Transactions on Signal and Information, vol. 3, (2014)
 - [6] E. Nishani and B. Çiço: Computer Vision Approaches based on Deep Learning and Neural Networks, Proceedings of 6th Mediterranean Conference on Embedded Computing, pp.1-4, (2017)
 - [7] R. K. Kumar, R. G. A. Kumar and G. Sanyal: Facial Emotion Analysis using Deep Convolution Neural Network, Proceedings of International Conference on Signal Processing and Communication (ICSPC'17), pp.369-374, (2017).
 - [8] A. Chakma, B. Vizena, T. Cao, J. Lin and J. Zhang: Image-based Air Quality Analysis using Deep Convolutional Neural Network, Proceedings of International Conference on Image Processing, pp.3949-3952, (2017)
 - [9] BVLC GoogLeNet Model,
https://github.com/BVLC/caffe/tree/master/models/bvlc_googlenet, (2017)
 - [1 0] A. Krizhevsky, I. Sutskever, G. E. Hinton: ImageNet Classification with Deep Convolutional Neural Networks, Proceedings of the 25th International Conference on Neural Information Processing Systems, pp.1097-1105, (2012)
 - [1 1] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich: Going Deeper with Convolutions, Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), vol. 1, pp.1-9, (2014)
 - [1 2] K. He, X. Zhang, S. Ren, J. Sun: Deep Residual Learning for Image Recognition, Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.770-778, (2016)
 - [1 3] Alex Krizhevsky: CIFAR-100 dataset,
<https://www.cs.toronto.edu/~kriz/cifar.html>, (2009)