CH-014

新出に対応する深層学習を用いたメタ認知に基づく画像認識
Deep Meta-Recognition Networks for Open Set Recognition

竹木章人 † 伊神大貴 † 入江豪 ‡ 相澤清晴 †
Akito Takeki Daiki Ikami Go Irie Kiyoharu Aizawa

1. はじめに

深層学習の発展により，画像認識・音声認識・翻訳など多岐に渡るタスクにおいて目覚ましい進展があった．特に画像分類のタスクにおいては，一般画像認識の競技会である ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2012 [2] において従来の手法に大差をつけて優勝した AlexNet [7] の登場以来，深層学習，特に Deep Convolutional Neural Networks (DCNNs) を用いた認識モデルの研究が盛んに行われている．

一般に画像分類のタスクは，「テスト時に出現するクラスは訓練時に出現するクラスの中に含まれており，それ以外のクラス（新規クラス）は出現しない」という，閉じた世界での画像分類 (Closed set recognition) を仮定して行われている．しかし現実世界の画像分類においては，学習時の画像とは本質的に異なる特徴を持った新出の画像が入力されることも十分にあり得る．多くの場合において，たとえ ImageNet [2] のような大規模データセットを用いたとしても，出現しうる全クラスのデータを集めることはほとんど不可能である．このような未知の画像が入力された際に既知と検出することは本来避けるべきであるが，未知検出の仕組みを持たない認識システムでは，誤って既知と判断することは避けられない．新出画像が入力された場合に，正しく未知の画像と判定する画像分類のタスクは，開かれた世界での画像認識（Open set recognition) [9] (図 1) と呼ばれている．

従来の画像分類においても，明示的に識別を行う必要が無いクラスとして「その他クラス (Known unknown classes)」を定義することは可能である．しかし，全ての存在しうる「その他クラス」を利用して学習を行う事は不可能である．そこで，未知検出の仕組みを持つ Open set recognition は，「未知の『その他クラス』(Unknown unknown classes) [8]」についても考慮することができ，極めて有用であると言える．Open set recognition のための手法はいくつか存在しているが，Deep Neural Network (DNN) を用いて未知検出を行う手法は OpenMax [1] のみと少なく，更なる研究が必要とされている．

本論文では DNN を用いる Open set recognition の新たなアプローチとして，Deep Meta-Recognition Networks(DMRN) を提案する．通常の DCNN の最終層に信頼度推定を行うためのネットワークを接続することで，End-to-End の形で信頼度推定を行う．既知である訓練中のデータから大きく異なる特徴を持った画像が入力された場合，DCNN の最終層のスコアと信頼度を元にその画像が未知であると判断する．更に，訓練中に利用できる画像のカテゴリに対して，入力される画像のカテゴリが既知もしくは未知であるケースを厳密に分離した上で手法の性能評価を行った．それらの実験を通して，提案手法と従来の Open set recognition の手法との性能差についての考察を行った．

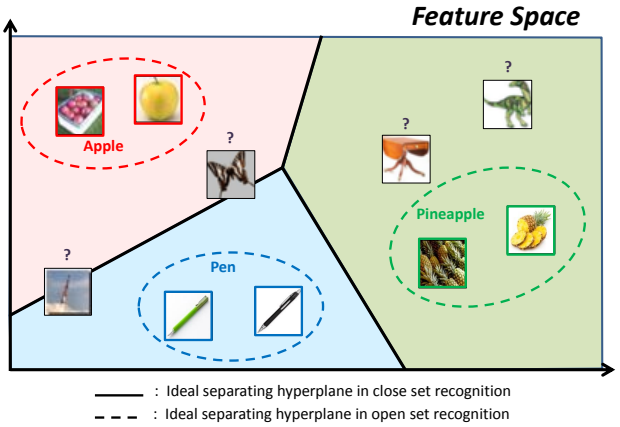

図 1: Open set recogntion の概要図．Open set recogntion では事前に登場しうるクラスを知ることはできない．既存の認識システムでは認識できるクラスの数は固定であるため，新たなクラスの画像にうまく対処できない．Open set recogntion では訓練中に登場したクラスはもちろん未知のクラスにも対応することを目指した，画像認識においてより現実的なタスクである．

2. 関連研究

はじめて Open set recognition のタスクについて論じた [9] では，訓練中に現れる既知クラスを含んだ半径 r_o の超球 S_o に対する学習後の認識器が既知のクラスと認識する特徴量空間 O の比率を”Open space risk”と定義し，Open space risk を最小化する事によって未知クラスに対応することが出来るとした．しかし Open space risk は単なる定義でしかなく，特徴量空間における距離の定義の仕方や特徴量空間 O をどのように定義するかについては定義されていない．そのため，距離関数の定義の仕方や特徴量空間 O の決定方法が問題となっている．

[9] では Support Vector Machine(SVM) の学習時において，正例と負例を分離する超平面と、その超平面と平行なもう 1 つの超平面で正例を挟み込むことにより、特徴量空間 O の範囲の制限を行っている．[10] では，Compact Abating Probability(CAP) と呼ばれる，特徴量空間において既知クラスの分布から離れるにつれ，その既知クラスへの所属確率が低下す

†東京大学大学院 情報理工学系研究科, Graduate School of Information Science and Technology, The University of Tokyo
‡NTT コミュニケーション科学基礎研究所, NTT Communications Science Laboratories

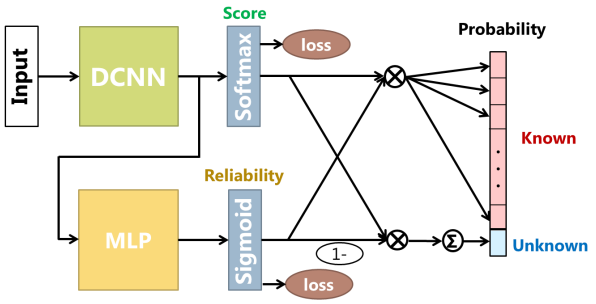

図 2: Deep Meta-Recognition Network (DMRN) の概要図

るモデルを提案し，OC-SVM [12] と 1-vs-all SVM の学習結果に対して統計的処理を行い，未知クラスであるかを判定している．

初めて Open set recognition に対して Deep Neural Network (DNN) を適用した OpenMax[1] は，入力に対して認識システムの出力したスコアから認識システムの予測が成功しているかを判定する Meta-Recognition [11] の概念を導入し，AlexNet の最終層である fc8 の出力としての Activation Vector(AV) を認識スコアとして扱い，そのスコアの信頼度を Weibull 分布を用いた統計学的処理を行うことで求めることによって入力された画像が未知クラスかどうかを判定している．

3. 手法

本章では，DNN を用いた Open set recognition の新たなアプローチとして，Deep Meta-Recognition (DMRN) を提案する．更に，今までの Open set recognition の枠組みにおいては議論されてこなかった既知・未知カテゴリ間の関係性について考慮した実験の枠組みを提案する．

3.1.Deep Meta-Recognition Network

提案手法の流れを図 2 に示す．DMRN は，入力画像から認識したい各クラスごとの認識スコアを DCNN を用いて計算する認識部と，抽出されたスコアの値が既知クラスからやってきた画像として取り得る値であるかを多層パーセプトロン (Multi Layer Perceptron , MLP) を用いて判定する信頼度推定部の 2 つの部分から構成される．DCNN から出力される認識スコアにソフトマックス正規化を行った値と MLP から出力される信頼度との要素積を最終的な既知クラスの確信度とする．また，認識スコアをソフトマックス正規化した値と MLP の不信頼度との要素積の総和を未知クラスの確信度とする．

DMRN の学習は通常の DCNN の学習と同様に確率的勾配降下法を用いて End-to-End の形式で行う事ができる．認識部の誤差関数としてはソフトマックス交差エントロピー関数を，信頼度推定部の誤差関数としてはシグモイドクロスエントロピー関数を用いる．ソフトマックスクロスエントロピー関数を誤差関数とした学習により，通常の DCNN を用いた学習と同等の認識特徴量空間を得る．また，信頼度推定部の Sigmoid クロスエントロピー関数を用いた学習により，認識部の出力が取り得る認識スコアの値であるかどうかをクラスごとに学習する．これら 2 つの学習を End-to-End の形で同時に行う．なお、[1] においても、DCNN の出力値から信頼度を求めている

[1] では，DCNN から出力される認識スコアのことを Activation Vector (AV) と呼び，訓練データに含まれる画像の AV の平均 (Mean ActivationVector, MAV) を各クラスごとに事前に計算を行う．そして，その MAV と訓練データに含まれる画像の AV との距離を Weibull 分布でフィッティングし，新たに DCNN に入力された画像の AV と MAV との距離がフィッティングした Weibull 分布関数に従う場合，入力画像が既知クラスである確率が高いと判定して高い信頼度を出力する．

3.2. カテゴリの差異を意識した評価

Open set recognition のタスクに対する手法はいくつか提案されているが、現在 Open set recognition のタスク用に作成されたデータセットは存在しておらず，自然画像分類用に作成されたデータセットをランダムに分割することで評価を行っている．[9, 1] しかし，訓練時の学習データに含まれるクラスと同じカテゴリの未知クラスの画像と訓練時の学習データに含まれない未知のカテゴリに含まれる未知クラスの画像が分類器に入力された場合，未知画像としての分類の容易さに差異があることが予想される．例えば，イヌ・ネコ・シカなどの動物カテゴリに含まれるクラスの画像について学習した場合，テスト時に登場する新規クラスの画像が訓練画像と同じカテゴリに含まれるキツネより，食べ物のカテゴリに含まれるリンゴの方が，直感的には未知クラスの画像であると判断しやすい．

同一カテゴリに含まれるクラスの画像は，視覚的にも共通の特徴を持つ事が多い．大規模画像データセットで画像分類の訓練を行った DCNN から得られる特徴量空間においても，同一のカテゴリに属する画像は近傍に存在しやすいことが知られている．そのため，未知のカテゴリに含まれる未知画像を未知と判定するよりも，既知のカテゴリに含まれる未知画像を未知と判定するタスクの方がより難易度が高いと予想される．従来の既知・未知クラスをランダムに選択するような枠組みでは，このような既知・未知カテゴリ間の関係性について議論する事は難しい．

そこで，本稿では学習データセットに含まれるクラスをカテゴリごとに分類した，2 種類の実験を行う．1 つ目の実験は，単一カテゴリの中で既知クラスと未知クラスの分割を行い，訓練時は既知クラスの画像のみで学習する．テスト時は既知・未知両方のクラスの画像を用いて評価を行う．未知画像が既知のカテゴリであった場合の Open set recognition の性能を比較する．2 つ目は，2 つカテゴリの中で一方のカテゴリを既知クラス，もう一方のカテゴリを未知クラスとした分割を行う．1 つめの実験と同様，訓練時は既知クラスの画像のみで学習する．テスト時は既知・未知両方のクラスの画像を用いて評価を行う．これにより，未知画像が既知のカテゴリには含まれない場合の性能を比較する．

表 1: 既知クラス（Bird）画像と未知クラス画像の枚数の比率 r に対する F 値

訓練	Bird								
テスト	Bird			Bird, Dog			Bird, Food		
手法	Softmax	OpenMax	Ours	Softmax	OpenMax	Ours	Softmax	OpenMax	Ours
$r = 0.1$	**0.946**	0.868	0.919	**0.946**	0.867	0.922	**0.946**	0.871	0.923
$r = 0.5$	**0.870**	0.854	0.833	**0.861**	0.850	0.837	**0.888**	0.867	0.854
$r = 1.0$	0.831	**0.835**	0.784	0.820	**0.835**	0.802	**0.858**	0.855	0.822
$r = 5.0$	**0.731**	0.729	0.703	0.699	0.705	**0.715**	0.721	0.708	**0.733**
$r = 10.0$	0.647	**0.666**	0.659	0.628	0.647	**0.669**	0.643	0.632	**0.707**

表 2: 既知クラス（Dog）画像と未知クラス画像の枚数の比率 r に対する F 値

訓練	Dog								
テスト	Dog			Bird, Dog			Dog, Food		
手法	Softmax	OpenMax	Ours	Softmax	OpenMax	Ours	Softmax	OpenMax	Ours
$r = 0.1$	**0.938**	0.832	0.876	**0.937**	0.831	0.893	**0.940**	0.840	0.891
$r = 0.5$	**0.824**	0.788	0.769	0.820	0.776	**0.825**	**0.826**	0.820	0.820
$r = 1.0$	**0.755**	0.745	0.684	**0.762**	0.721	0.758	0.775	**0.794**	0.760
$r = 5.0$	**0.565**	0.564	0.460	**0.619**	0.574	0.611	0.657	**0.674**	0.616
$r = 10.0$	0.463	**0.493**	0.375	**0.569**	0.529	0.547	0.568	**0.604**	0.554

4. 実験結果

4.1. 訓練

今回の実験は、1000 クラスの画像を持つ ILSVRC2012 のデータセットの内、bird, dog, food のカテゴリに属する画像 225 クラスを用いて行った．これら 3 つのカテゴリは比較的多くのクラスを持つカテゴリであり、各カテゴリが互いに異なる視覚的特性を持っている．このデータセットは 1 クラスに対して約 1000〜1300 枚の訓練用画像，50 枚の検証用画像と 150 枚のテスト用画像を含んでいる．150 枚のテスト用画像については正解ラベルが公開されていないため，今回の実験の評価は検証用画像を用いて行った．

カテゴリごとの訓練については，まずカテゴリ内のクラスの中からランダムに 30 クラスを既知クラスとして選択し，残りのクラス全てを未知クラスとして評価に用いた．今回の実験に用いた bird・food・dog カテゴリの未知クラス数はそれぞれ 29，90，16 クラスであった．テスト時に用いる未知クラスの画像は，検証用画像を用いて行う既知クラスとは異なり，出来る限り多くの画像を利用するために訓練用画像を用いて行った．

提案手法における認識部のネットワークモデルには，DenseNet-121(k=32) [5] を用いた．信頼度推定部については 3 層の多層パーセプトロン（MLP）を用い，活性化関数としては Rectified Linear Units（ReLU）[3] を用いた．各層への入力の前にバッチ正規化処理 [6] を行い，MLP の中間層のチャンネル数は全て 100 とした．

確率的勾配降下法を用いて全体の学習を行い，ミニバッチサイズは 32 とした．最初の学習率を 0.1，最終的な学習イテレーション数を 10 万回に設定し，5 万イテレーションと 7.5 万イテレーションに達した時に学習率を 0.1 倍した．[5] に従い，重み減衰は 10^{-4} に設定し，パラメタが 0.9 である減衰なしの Nesterov momentum [13] を用いた．各層の重みの初期化は [4] の方法を用いた．

4.2. 評価

Open set recognition の性能評価においては，複数のクラスに対する既知画像分類の誤りと入力画像が既知であるか未知であるかの判定の両方について考慮する必要がある．本稿では，性能評価の指標には既知画像と未知画像の比率に対してロバストな指標である F 値を用いた．また，テストに使用する既知画像と未知画像の枚数の比率を変化させた場合の Open set recognition の性能についても評価した．

提案手法との差分を示す比較手法として，Softmax と OpenMax [1] を用いた．Softmax では，DMRN の DCNN 部分の最終層である全結合層（Activation Vector(AV)）の出力を利用した．ある画像の入力に対する AV の最大値がある一定の閾値以下の場合，その画像は未知クラスに属する画像であると判定した．各学習の結果に対して最も高い F 値を出す閾値を選択し，それを最終結果として評価した．

4.3. 結果

学習時に登場するカテゴリが 1 つであり，テスト時に入力されるカテゴリが既知もしくは未知である場合の結果を表 1–3 に示す．既知画像の枚数については 1 クラス当たり 1,500 枚と固定し，未知クラス画像は 1 クラスあたり 150–15,000 枚で変化させて実験を行った．

表 1 の結果より，既知カテゴリが Bird の場合は未知画像の枚数の比率を大きくしても全体的に F 値

表 3: 既知クラス（Food）画像と未知クラス画像の枚数の比率 r に対する F 値

訓練	Food								
テスト	Food			Bird, Food			Dog, Food		
手法	Softmax	OpenMax	Ours	Softmax	OpenMax	Ours	Softmax	OpenMax	Ours
$r = 0.1$	**0.941**	0.806	0.876	**0.943**	0.823	0.885	**0.943**	0.817	0.868
$r = 0.5$	**0.795**	0.745	0.786	**0.849**	0.811	0.812	**0.833**	0.782	0.767
$r = 1.0$	**0.725**	0.693	0.709	**0.800**	0.786	0.753	**0.795**	0.747	0.683
$r = 5.0$	0.554	**0.559**	**0.559**	**0.644**	0.626	0.612	**0.710**	0.605	0.554
$r = 10.0$	0.456	**0.463**	0.454	**0.562**	0.528	0.512	**0.674**	0.573	0.507

が比較的高めであることが分かる．これは通常の DCNN の学習においても高い精度で学習に成功していた（87.0%）ことから，画像認識に適した特徴量空間を得ることができたためと考えられる．また，未知画像の枚数の比率が小さい場合は DMRN よりも Softmax と OpenMax の F 値が高いが，比率が大きくなるにつれて順位が逆転している．これは未知画像の枚数の比率が大きくなると単純な閾値では取り除くことが困難であるような，特定のクラスに対して高い認識スコアを出す画像の総量が増加するためであると考えられる．

表 2 の結果より，既知カテゴリが Dog の場合で未知カテゴリも同じ Dog である場合，全体的に F 値が低くなることが分かる．これは，同一のカテゴリに属する未知クラスは検出が難しいという当初の予想に一致する．表 3 の結果においても同様に既知カテゴリに属する Dog の未知クラス画像は検出が難しいという結果を示している．また既知カテゴリが Bird の場合の結果と異なり，未知画像の枚数の比率に関わらず常に DMRN が低い F 値を示している．これは通常の DCNN の学習における精度があまり高くない（74.6%）ことから，十分によい画像の特徴量空間を得られていない，もしくは学習データ内に存在する画像の特徴量の分布の偏りのため，正しい既知クラスの画像までも未知と判定しがちであることが考えられる．これは表 3 の結果においても同様のことが予想される．

更に表 3 の結果より，既知カテゴリが Dog である場合に未知と判定しやすいカテゴリは Food であり，既知カテゴリが Food である場合に未知と判定しやすいカテゴリは Dog であることから，Dog の学習で得られる特徴量と Food の学習で得られる特徴量の差が大きいことが示唆される．

5. まとめ

本稿では，Deep Neural Networks を用いた Open set recognition の新たなアプローチとして，Deep Meta-Recognition Network (DMRN) を提案した．Deep Convolutional Neural Networks の最終層に信頼度推定を行うためのネットワーク として多層パーセプトロン (MLP) を接続することで End-to-End の形で信頼度推定を行い，既知のデータから大きく異なる特徴を持った画像が入力された場合，DNN の最終層のスコアと MLP の信頼度を元にその画像が未知であると判断するような学習を行った．加えて，入力される画像のカテゴリが既知もしくは未知であるかを区別した性能評価を行った．それらの実験を通して，Open set recognition のための各手法の性能について考察を行った．

謝辞

本研究の一部は，JST CREST JPMJCR1686 の支援を受けた．

参考文献

[1] A. Bendale and T. Boult. Towards open set deep networks. In *CVPR*, pages 1563–1572, 2016.

[2] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A large-scale hierarchical image database. In *CVPR*, pages 248–255, 2009.

[3] X. Glorot, A. Bordes, and Y. Bengio. Deep sparse rectifier neural networks.

[4] K. He, X. Zhang, S. Ren, and J. Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *ICCV*, pages 1026–1034, 2015.

[5] G. Huang, Z. Liu, K. Q. Weinberger, and L. van der Maaten. Densely connected convolutional networks. *arXiv preprint arXiv:1608.06993*, 2016.

[6] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. 2015.

[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *NIPS*, pages 1097–1105, 2012.

[8] D. Rumsfeld. Known and unknown: a memoir. *Penguin*, 2011.

[9] W. J. Scheirer, A. de Rezende Rocha, A. Sapkota, and T. E. Boult. Toward open set recognition. *TPAMI*, 35(7):1757–1772, 2013.

[10] W. J. Scheirer, L. P. Jain, and T. E. Boult. Probability models for open set recognition. *TPAMI*, 36(11):2317–2324, 2014.

[11] W. J. Scheirer, A. Rocha, R. J. Micheals, and T. E. Boult. Meta-recognition: The theory and practice of recognition score analysis. *TPAMI*, 33(8):1689–1695, 2011.

[12] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson. Estimating the support of a high-dimensional distribution. *Neural computation*, 13(7):1443–1471, 2001.

[13] I. Sutskever, J. Martens, G. E. Dahl, and G. E. Hinton. On the importance of initialization and momentum in deep learning. *ICML (3)*, 28:1139–1147, 2013.