法政大学学術機関リポジトリ
HOSEI UNIVERSITY REPOSITORY

PDF issue: 2024-05-29

敵対的性質を示す仮想類似データを利用した深層距離学習の汎化性能の向上

松岡, 佑磨 / MATSUOKA, Yuma

(出版者 / Publisher)
法政大学大学院理工学研究科
(雑誌名 / Journal or Publication Title)
法政大学大学院紀要. 理工学・工学研究科編
(巻 / Volume)
60
(開始ページ / Start Page)
1
(終了ページ / End Page)
6
(発行年 / Year)
2019-03-31
(URL)
https://doi.org/10.15002/00022055

敵対的性質を示す仮想類似データを利用した深層距離学習の汎化性能の向上

VIRTUAL ADVERSARIAL SIMILAR POINT
TO IMPROVE GENERALIZATION OF DEEP METRIC LEARNING

松岡佑磨
Yuma MATSUOKA
指導教員　彌冨仁

法政大学大学院理工学研究科応用情報工学専攻修士課程

Deep Metric Learning learns a small dimensional feature representation from input data points which has a geometry same as the input data points, in where the distance between similar data points are small and the distance between dissimilar datapoint are large. Therefore, it has been widely used in a variety of tasks like image retrieval and person re-identification. However, it requires to sample some kind of input data points to calculate similarity and dissimilarity to optimize itself but getting difficult to find efficient variations by hard example mining during its training. In this paper, we propose a novel deep metric learning method which is optimized by a loss function with generated virtual adversarial similar point and a metric loss and evaluate its performance in the Zero-shot learning benchmark with CUB-200-2011 and CARS-198 datasets.

Key Words *: deep metric learning, adversarial examples, fine-grained image recognition, zero-shot learning.*

１．背景

Convolutional Neural Network(CNN)に代表される深層学習の発展[1-3]により，深層距離学習は，画像検索やクラスタリング，顔認識，顔認証など様々な機械学習のタスクや分野に応用されるようになってきた．その理由として，深層距離学習は，入力データ空間において類似したデータはデータの距離が小さく，相違なデータはデータ間距離が遠いという関係性を保持したまま，入力データから特徴を抽出することを目的とする手法であることが挙げられる．

このような特徴空間を作成するために，深層距離学習ではペアの入力データを利用し，深層距離学習を行う contrastive loss[4]や 3 つのデータの組み合わせを利用する triplet loss[5-7]が提案されてきた．このように深層距離学習は学習に複数の入力データの組み合わせをサンプリングする必要であり，手法を適用するデータセットが大規模になればなるほど，膨大な組み合わせの中から学習に有効なデータの組み合わせを見つけることは難しくなる．そのため，比較的学習が容易なサンプルの組み合わせは学習が進み易く，距離学習の制約を満たしやすいため，学習が進むにつれ距離学習の誤差は値がほぼ 0 になり，誤差逆伝播法に基づき学習モデルのパラメータの更新を行う際にも，学習に貢献しないことが多い．以上のことから本当に学習すべき入力データの組み合わせを見つけることが重要であるが，深層距離学習はクラス間の分散を大きく，一方でクラス内のデータの分散を最小化する目的関数に基づいて特徴空間を学習する．そのため，学習データのマイニングは大きな計算コストを必要とし，学習が進むほど学習に貢献しない組み合わせが増える一方で，本当に学習すべき類似データ hard positive，相違なデータ hard negative の割合は少なくなる．このようなことから学習に有効な入力データの組み合わせをマイニングによって見つけるのは難しい．

本研究では，学習モデルの誤差の勾配情報を利用して擬似的に学習に有効な入力データを生成し，深層距離学習の学習に利用する手法を提案する．この手法では，敵対的な性質を示すデータを仮想的に生成し，敵対的な性質を持つ類似データ hard positive として学習データの組み合わせに利用する．学習が進むにつれ，学習に貢献しない大量の相違なデータが増え学習に有効な入力データの組み合わせが少なくなった場合にも，擬似的な positive data を学習に利用することで，データセットの大部分の不要なデータを再利用できる可能性がある．以上のことから，本研究では計算コストの高いデータのマイニング

を行う必要なく，敵対的性質を示す仮想データを利用して効率的な学習を行う深層距離学習を提案し，CUB-200-2011[8], Cars196[9]の 2 つの画像の検索のベンチマークにおいて提案手法の性能を検証する．

２．方法

（１）提案手法の全体

提案手法は，Fine-grained Image Recognition のベンチマークにおいて一定の成果を出している深層距離学習法をベースとする．深層距離学習法の代表的な手法として Triplet Network[5]，さらに効率的な距離学習が可能な N-pair sampling[10]が提案されている．提案手法では N-pair sampling に基づく深層距離学習を，”敵対的な性質を示す”仮想データを生成し，学習すべき類似データとして学習に使用することで，上記の認識ベンチマークにおいて深層距離学習法の精度向上を行う．

（２）Triplet Network と N-pair sampling

Triplet Network のネットワークモデル図を Fig.1 に示す．Triplet Network はx_a，x_p，x_nの 3 つの画像データの組み合わせを入力データとして用いる．このときx_pはx_aと同じ教師ラベルを持つデータとする．x_nは異なるラベルを持つデータとする．これらの画像の組み合わせを(パラメータを共有する) CNN $f(\cdot)$に入力し，特徴ベクトル$f(x_a)$，$f(x_p)$，$f(x_n)$を得る．この CNN の学習はこれらの出力を用いて，式(1)で示す triplet hinge loss 関数を最小化するように誤差逆伝播法で最適化される．式(1)において，mは特徴空間におけるデータ間距離を決めるハイパーパラメータであり，式(2)は深層距離学習に使用する距離指標であり，triple hinge loss 関数では２つの画像データから抽出した特徴ベクトルのユークリッド距離の２乗を出力する関数を使用する．

$$L_{triplet}(x_a, x_p, x_n) = \max\{0, d(f(x_a), f(x_p)) - d(f(x_a) - f(x_n)) + m\} \quad (1)$$

$$d(x, y) = \|x - y\|_2^2 \quad (2)$$

この Triplet Network の距離学習と，学習によって形成される特徴ベクトル空間のモデル図を Fig.2 に示す．Triplet Network の距離学習が十分に収束した場合は，Fig.2 で示すように，教師ラベルごとに偏るような特徴空間を構築することが可能である．

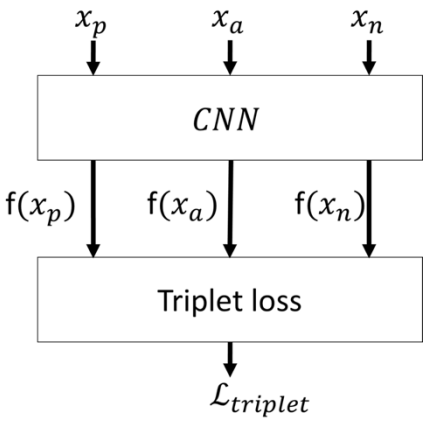

Fig. 1: Triplet Network のモデル図

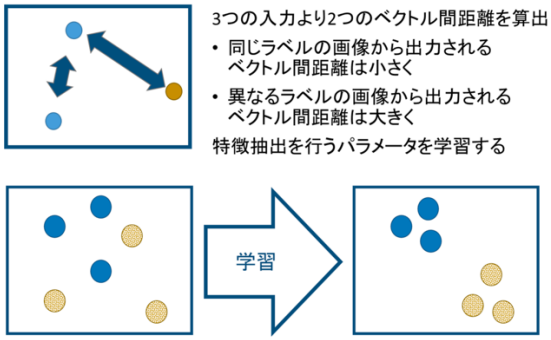

Fig.2: Triplet Network を用いた距離学習の図

Triplet Network は 3 つの画像の組み合わせを学習に使用する入力データとしてサンプリングしていたが，N-pair sampling は Fig.3 で示すように，1 つのクラスから 2 枚の画像を Pair(x_a，x_p)としてサンプリングし，ミニバッチのサイズと等しい Pair をランダムにサンプリングする．学習の誤差関数を計算する際には，Fig.3 で示すようにミニバッチ内の異なるクラスのデータをx_nとして用いることで，1 つのクラスからサンプリングした Pair ごとに多様な異なるクラスのデータと学習を行うことができるため，Triplet Network よりも効率的な深層距離学習を行うことができる．[10]で提案された N-pair loss 関数を式(3)として示す．

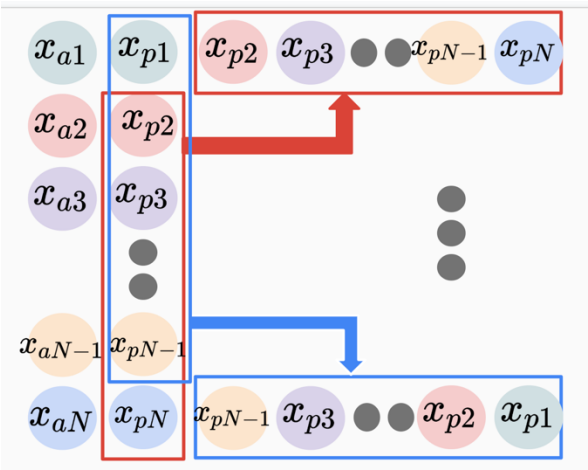

Fig.3: N-pair sampling によって作成する学習サンプルデータと N-pair loss の計算法のモデル図.

$$L_{N-pair}(x_a, x_p, x_n) = \log(1 + \exp(d(f(x_a) - f(x_p)) - d(f(x_a) - f(x_n)))) \quad (3)$$

式(3)において，誤差関数を最小化することで，Triplet Network と同じように, 同じクラスに属するデータ間距離を小さく，異なるクラスのデータ間距離を大きくする discriminative な特徴空間を学習することができる．本研究では，式(3)のベクトル間距離$d(\cdot)$は式(2)を用いる．

（３）敵対的性質を示す仮想データの生成と，この仮想データを類似画像として深層距離学習を行う提案手法

提案する手法は，深層距離学習法で使用する入力データx_aに対して敵対的性質を示す仮想データであるx_{vadv}を作成し，これを同じラベルの画像x_pとみなして深層距離学習を行う．本提案手法では式(4)の Triplet loss $L_{triplet}$に Npair-loss を使用し，敵対的な性質を示す仮想データを類似画像とみなして距離学習に利用し，学習を行う誤差関数 $L(x_a, x_{vadv}, x_n)$を算出する．そして式(4)で示すようにこの誤差と通常の入力データの組み合わせを用いた誤差$L(x_a, x_p, x_n)$の合計の誤差$\mathcal{L}$を最小化する深層距離学習を行う．敵対的性質を示す仮想データは VAT[11]で提案された学習誤差の勾配情報を使用する方法に基づき作成する．

$$\mathcal{L} = L(x_a, x_p, x_n) + \lambda L(x_a, x_{vadv}, x_n) \quad (4)$$

$$x_{vadv} = x + r_{vadv} \quad (5)$$

$$r_{vadv} = \arg\max_{r;\, \|r\|_2 \le \epsilon} \|f(x) - f(x + r)\|_2 \quad (6)$$

r_{vadv}はある入力xに対してユークリッド空間上のϵの範囲で, 最も出力に大きな影響を与える変動を示し, r_{vadv}方向のデータ変動を入力xに対して加算することで学習すべき敵対的な仮想データx_{vadv}を作成する．加算する際にx_{vadv}に任意の大きさϵを乗算し増幅するが，この時のϵは仮想データを作成するデータのドメインによって最適な値が異なるため，データセットごとに適切な値を決める必要がある．

３． 評価実験

提案手法の有効性を確かめるために使用したデータセット，学習に使用するハイパーパラメータ，そしてこれらの条件に基づいた環境で行なった実験の結果を示す．

（１）実験に使用するデータセット

検証実験には 2 種類のデータセットを用意した．それぞれのデータセットは多数の鳥の種や自動車の車種別の画像から構成される Fine-grained Image Recognition のデータセットである．それぞれのデータセットは極端に多いクラスで構成され，それぞれのクラスには極端に少ない画像データが用意されている．一部のクラスのデータで学習し，テストは学習に全く使用していない異なるクラスの画像を使用する Zero-shot Learning の枠組みで検証実験を行う．

1 つ目のデータセットは Fig. 3 で示すような画像から構成される Cars196 データセット[9] であり, それぞれ異なる 196 種類 の車の画像 16,000 枚から構成される．Zero shot learning で使われる実験プロトコルに従い最初の 98 種類のクラスの画像を学習用データとして, 後半の 98 種類のクラスに属する画像をテスト用データとしてラベルを予測する．2 つ目は Fig. 4 で示すような画像から構成される CUB-200-2011 データセット[8]であり，このデータセットは 200 種類の鳥の画像 12,000 枚から構成される．最初の 100 クラスの画像 5,864 枚を学習に使用し，残りのデータをテストに用いる．

（２）実験環境の詳細

先行研究[12] に従いベンチマークを行なった．深層学習に使用する CNN は VGG16-BN[2, 13]最適化手法 optimizer は momentum SGD とし，momentum の重みは 0.9 に，weight decay の値は 5e-4 を選択した．学習に使用するデータはランダムスケール，クロッピング，垂直方向のフリッピングを前処理として行い，data augmentation を行い，いっぽうでテストデータはセンタークロップのみ前処理として行った．学習はバッチサイズ 32 のミニバッチ学習を行い，Cars-196 データセットと CUB-200-2011 は 20,000 イテレーション，Online Product は 200,000 イテレーションの学習を行った．optimizer の初期学習率 leraning rate は CUB-200-2011 で 1e-3, cars196 データセットで 1e-2, Online Product で 2e-3 とした．GDML[12]の実験設定に基づき，VGG Net の最終層の fc6 layer だけでなく，1 つ下の Pool5.3 layer から出力される特徴ベクトルを用いて Recall@1 の精度を算出した．また，提案手法の仮想データを生成する際のパラメータϵ

は実験によって最適な値を決定した.

（３）実験結果

ベースラインとなる先行研究[12] の再現実験を行った後，提案手法でチューニングする必 要のあるハイパーパラメータϵ の値を決定する．CUB-200-2011 データセットにおいてϵ の 値を変えて提案手法を学習させ，テストデータに対する Recall@1 の値を算出した. この結果を Fig.6 に示す.

Fig. 6 において，ϵ の値が 2 から 10 の範囲で高い精度が出ていることがわかる．Car196 データセットにおいてϵ の値を変えて提案手法を学習させ，テストデータに対する Recall@1 の値を算出した．この結果を Fig. 7 として示す． Fig. 7 において，fc6 の結果からϵ の値が 7 から 32 の範囲で高い精度が出ていることがわ かる．一方で pool5.3 の結果からは特定のϵ の値によらず一定の結果が得られていることがわ かる．提案手法とこれまでの先行研究の画像検索の精度 Recall@1 の結果を Table. 4.1 に示す． Table. 4.1 より，提案手法は Recall@1 の評価指標において, layer pool5.3 の出力特徴ベクトルでも, layer fc6 の出力特徴ベクトルでもベースラインとなる先行研究[31] の精度向上を果たしていることがわかる．特に layer fc6 において高い精度向上が確認できた.

Fig.6 : CUB-200-2011 データセットにおいて提案手法のハイパーパラメータεの値を変えた時の実験結果

Fig.7: Car196 データセットにおいてハイパーパラメータεを変えた時の実験結果

４． 考察

提案手法において人手で調節する必要があるハイパーパラメータによる性能の違い，既存の先行研究との比較を行う.

Table.1: 提案手法と比較手法の画像検索の Recall@1 の結果

Method	Network	Dim	Cars-196	CUB-200-2011
Lifted structure	GoogLeNet	64/64/512	53.0	47.2
Facility	Inception v1 BN	64	58.1	48.2
Angular Loss	GoogLeNet	512	71.4	54.7
Proxy-NCA	GoogLeNet	64	73.2	49.2
ABE	8 Heads Ensemble	512	85.2	60.6
Baseline(rebuild GDML)	VGG16-BN(pool5.3)	512	87.9	67.2
Baseline(rebuild GDML)	VGG16-BN(fc6)	512	80.7	63.4
Our method	VGG16-BN(pool5.3)	512	**88.17**	**68.8**
Our method	VGG16-BN(fc6)	512	**82.92**	**65.6**

（1）ハイパーパラメータと Deep Metric Learning の汎化性能の関係

Cub200-2011 データセットにおける提案手法の Recall@1 の精度は Fig. 4.4 より，敵対的仮想類似データを作成する際に付与する敵対的なノイズの大きさϵは影響を与えていることがわかる．ϵの値が小さいと元データと敵対的仮想類似データの特徴ベクトル間の距離がほぼ同じになってしまい，過学習が生じている可能性がある．一方で，ϵの値が大きすぎる場合は，汎化性能が上がるような有益な仮想データが作れていない可能性があり，そのために精度が上がらない，または下がっている可能性がある． Fig. 4.5 は pool5.3 の Recall@1 の結果がϵに影響を受けず，ほどんど同じ結果となった．一方で，fc6 の Recall@1 の値より $7 < \epsilon < 32$ の範囲で精度の向上が見られるため，提案手法の敵対的仮想類似データを用いた学習は全結合層の fc6 の過学習を防いでいると考えられる．[12] の提案した pool5.3 で Recall を計算する手法は，fc6 と比較して過学習しづらい特性を持つために fc6 よりも精度が高く，提案手法の過学習抑制効果があまり得られないのではないかと考えられる．

（2）提案手法と既存研究との Recall@1 の精度の比較

Table. 1 より，提案手法は既存の手法でありベースラインとなる，高い汎化性能を達成している[13] の手法の精度向上を達成した．この結果より，提案手法は敵対的な性質を示す擬似的なデータ敵対的仮想データを同じクラスのデータとみなして学習を行うことが Deep Metric Learning の学習に有益な学習をもたらすとわかる．提案手法は Deep Metric Learning の positive データを擬似的に生成した．既存研究にはより洗練された Metric Loss や，学習に効率的な Negative データを作成する手法があるので，これらの手法を組み合わせて，より精度の高い画像検索手法を実現したい．

５．結論

本研究では，多数のクラスラベルが少数の画像に付与された同じ粒度の画像から構成される Fine-grained Image recognition と Zero-shot learning の枠組みの中で検索システムの性能向上を達成した．Deep Metric Learning というデータ群から特徴を抽出する際に，よりクラスラベルごとに偏るようなクラスタを形成する手法に，敵対的な性質を示す擬似的なデータ Virtual point を生成し，あるデータに対して学習すべき hard positive データとして用いることで汎化性能を向上させた．本報告では提案手法がこのベースラインとして用いた State of The Art の先行研究に対してその精度を向上させたことを確認した．今後は提案手法でチューニングの必要なハイパーパラメータの自動推定や，より高い汎化性能を達成するための手法に取り組んでいく．

謝辞：本研究にあたり，全般にわたるご指導をしてくださった彌冨仁准教授，および彌冨研究室の皆様に深く御礼申し上げます．また，修士課程２年間のうち１年間を Visiting Graduate Scholar として過ごした Johns Hopkins University の CCVL(Computational Cognition, Vision, and Learning) research group の Bloomberg Distinguished Professor の Alan Yuille 教授，そして研究留学中に多くのディスカッションをした PhD. Candidates に深く御礼申し上げます．

参考文献

1) Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classication with deep convolutional neural networks. In Advances in neural information processing systems, pp. 1097-1105, 2012.
2) Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large scale image recognition. CoRR abs/1409.1556, 2014.
3) Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pat- tern recognition, pp. 1-9, 2015.
4) Sergey Zagoruyko, and Nikos Komodakis. Learning to compare image patches via convokutional neural networks. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2015.
5) Jiang Wang, Yang Song, Thomas Leung, Chuck Rosenberg, Jingbin Wang, James Philbin, Bo Chen and Ying Wu. Learning fine-grained image similarity with deep ranking. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 1386-1393, 2014.
6) Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. CoRR abs /1503.03832, 2015.
7) Eland Hoffer and Nir Ailon. DEEP METRIC LEARNING USING TRIPLET NETWORK. International Workshop on Similarity-Based Pattern Recognition, 2015.
8) C. Wah, S. Branson, P. Welinder, P. Perona, and S. Belongie. The Caltech-UCSDBirds-200-2011 Dataset. Technical report, 2011
9) Jonathan Krause, Michael Stark, Jia Deng, and Li Fei-Fei. 3d object representations for ne-grained categorization. In Proceedings of the IEEE International Conference on Computer Vision Workshops, pp. 554-561, 2013.
10) Kihyuk Sohn. Improved deep metric learning with multi-class n-pair loss objective. In Advances in Neural Information Processing Systems, pp. 1857{1865, 2016.
11) Takeru Miyato, Shin-ichi Maeda, Shin Ishii, and Masanori Koyama. Virtual adversarial training: a regularization

method for supervised and semi-supervised learning. IEEE transactions on pattern analysis and machine intelligence, 2018.

12) Nam Vo and James Hays. Generalization in metric learning: Should the embedding layer be the embedding layer? CoRR abs/ 1803.03310.

13) Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. CoRR abs/1502.03167, 2015.