Implementation of Unsupervised Domain Adaptation by Backpropagation

Chih-Hui Ho Yuan Qi Xingyu Gu
Department of Computer and Science Engineering
University of California San Diego
chh279@eng.ucsd.edu yuq083@eng.ucsd.edu xig090@eng.ucsd.edu

Abstract

Deep learning models have been shown to be impressively powerful in tasks such as image classification and object recognition given massive amounts of labeled data. However, it is still an open challenge on the domain where only few labeled data are available. Inspired by the previous work Unsupervised Domain Adaptation by Backpropagation [5], aiming at minimizing the gap between deep features from target domain and source domain, we decide to implement the content mentioned in this paper. The project will contains experiments evaluating the performance of the classifiers trained with the proposed architecture in [5]. Several domain adaptation tasks will be conducted using dataset including MNIST [13], MNIST-M, Synthetic numbers, SVHN [18], Synthetic Signs, GTSRB [24], Office. Moreover, we will extend the paper in the follow 4 directions (i) applying the proposed architecture to different kinds of pretrained models (ii) conducting experiments on dataset that are not listed in [5] (iii) examining and comparing the adaptation difficulties between different datasets.(iv) investigating the possibility of solving the domain adaptation problem with generative adversarial network [8, 1, 2]

1. Introduction

The last few years have shown that large convolutional neural networks (CNNs), such as Alexnet [12], VGG [23], GoogleNet [25], ResNet [9], etc. achieve previously unmatched image classification performance. The successfulness of these deep models are established on the large scale labeled image dataset ImageNet [4].

For most of the applications which uses deep learning models as basic architecture, the models are pretrained on ImageNet and transferred to the specific task by fine-tuning. However, fine-tunning is only useful when the labeled data for the specific task is available to the developer. For the cases where only limited amount of labeled data or no labeled data are available, insufficiently-tunned deep model will fail in solving the interested task. Such scenarios are quite common in the real world application. For example, medical images are really rare and hard to collect, which makes fine-tunning techniques no longer useful in medical images classification task.

Although lacking of large-scale labeled dataset for the targeted task, there are lots of relevant datasets that might be useful for the targeted task. Take multiview image classification task for example. There is no existing large-scale labeled dataset containing multiview of the real images, but the synthetic graphical 3D image dataset, such as ModelNet [29] and ShapeNet [3], are available to the public. As a result, the question becomes whether it is possible to utilize those labeled data from different domain and apply the trained classifier to the targeted task.

In this situation, domain adaptation often provides an attractive solution, given that labeled data of similar nature but from a different domain (e.g. synthetic images) are available. Among all the domain adaptation relative works, the paper Unsupervised Domain Adaptation by Backpropagation [5] is one of the earliest papers in this field, which is cited by lots of papers. In this project, we want to implement the proposed solution in this paper, and test it by the experiments talked in this paper. Moreover, if time is enough, we also want to apply this approach in other fields, which are not talked about in this paper.

2. Related work

There has been a great amount of research in the field of domain adaptation and transfer learning in the past few decades. Most of the previous work are established on the shallow learning methods, including [22, 19, 6]. The previous shallow learning domain adaptation methods aim to solve the domain shift between the source and target domains and those methods can be mainly separated into 2 categories. The first category learns the share feature space between target and source domain [19, 6], while the other category trains the classifier on the weighted source data [22].

As the emergence of deep learning, many deep features are extracted using deep learning methods. The deep features are the high level abstract representation of the input and is shown to effectively encode the important information of the input. An emerging problem in deep learning is that it is prone to overfitting to the training data or source domain. Such problem restricts the generalization of the trained classifier. As a result, recent works have intensively addressed this problem by combining the domain adaptation task with deep learning.

The literature of domain adaptation can be separated into 2 categories. The first category uses part of the target data, either unsupervised or semi-supervised, to bridge the discrepancy between source and target domain. Previous works include [10, 16, 15, 20, 28]. The second category uses a discriminator to distinguish between domains, so that the feature extractor can learn a domain invariant feature regardless the input domain to fool the domain discriminator. [14, 11, 26, 27] and the paper *Unsupervised Domain Adaptation by Backpropagation* that we are implementing falls into this category.

The nature of Deep Domain Adaptation (DDA) [5] is a minimax game, very like that of a Generative Adversarial Network (GAN). However, traditional GAN [7] is very hard to train because the loss cannot indicate the network's performance. Furthermore, the training of generator and discriminator has to be balanced well, which is just like what we do to the label classifier and domain classifier in DDA and is very tricky. However, Wasserstein GAN [1] fixed the problem by using Wasserstein distance as the optimization goal instead of KL divergence or JS divergence. This makes the loss able to indicate training process and makes the discriminator can be trained to converge. Taken into account Wasserstein distance's success in both theory and practice, and its similar nature to DDA, it is highly possible that it can be applied to DDA to improve the performance.

3. Schedule

We plan to divide our project into four phases.

(1) We are going to read the paper detailedly about how the approach works, how the experiments in this paper are conducted, etc. These work will be finished before our presentation in class, which is planned on April 30th.

(2) Then we will implement the approach in pytorch, as we plan at present, and test the implementation in the situation: source dataset: MNIST [13], target dataset: MNIST-M. Phase two is to be finished before May 10th.

(3) For phase three, we will try to complete all the experiments talked about in this paper, and then compare the final project with the paper and see whether we have achieved our goal, this work is expected to be accomplished before Jun 1st.

(4)For phase four, this is an optional one, based on the implementation we accomplished in the first three phase, we will test it in other fields which is not talked about in the paper. Moreover, since the proposed method does not have well adaptation from MNIST [13] to SVHN [18], we would like to investigate the possibility to use generative adversarial network [8, 1] to solve the problem.

4. Resources

Training resource: Provided by this course.

Software: We plan to finish our work in Pytorch, and there is no realized Pytorch code for this paper on the Internet.

Datasets: MNIST [13], MNIST-M [5], Synthetic numbers [5], SVHN [18], Synthetic Signs, GTSRB [24], Office Dataset [21]. Synthetic Signs dataset is used in [5], and is provided by [17].

Based on the experiments of the work Unsupervised Domain Adaptation by Backpropagation [5], the source domain and the target domain will be the followings in Table 1.

Table 1. Source and target domain for each experiment

Experiment	Source Domain	Target Domain
1	MNIST	MNIST-M
2	SYN NUMBERS	SVHN
3	SVHN	MNIST
4	SYN SIGNS	GTSRB
5	AMAZON	WEBCAM
5	DSLR	WEBCAM
5	WEBCAM	DSLR

Like the work in Unsupervised Domain Adaptation by Backpropagation [5], excluding the experiments for domain adaptation, we will also conduct the ”source-only” experiments(i.e. if no adaptation is performed), which is the lower performance bound, and the ”target-only” experiments (training on the target domain data with known class labels), which is the upper bound on the DA performance. Especially, for the target-only experiments, we will divide the target domain data with known class labels by 8:2 for training data and testing data.

5. Question need to be answered

(1) Hyper-parameters. In [5], the hyper-parameter λ is critical to the model's performance. It controls the trade-off between the two objectives that shape the features during learning. How to search a good λ efficiently remains as a question.

(2) The proposed solution in [5] fails in adapting data from MNIST to SVHN. What could be the reason leading to this failure? If we can figure out this, we may find a way to improve it. Another idea would be introducing methods that performs well in other tasks (e.g. Wasserstein GAN [1]) to our work.

References

[1] M. Arjovsky, S. Chintala, and L. Bottou. Wasserstein gan. *arXiv preprint arXiv:1701.07875*, 2017.

[2] K. Bousmalis, G. Trigeorgis, N. Silberman, D. Krishnan, and D. Erhan. Domain separation networks. *CoRR*, abs/1608.06019, 2016.

[3] A. X. Chang, T. A. Funkhouser, L. J. Guibas, P. Hanrahan, Q. Huang, Z. Li, S. Savarese, M. Savva, S. Song, H. Su, J. Xiao, L. Yi, and F. Yu. Shapenet: An information-rich 3d model repository. *CoRR*, abs/1512.03012, 2015.

[4] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09*, 2009.

[5] Y. Ganin and V. Lempitsky. Unsupervised domain adaptation by backpropagation. In *Proceedings of the 32Nd International Conference on International Conference on Machine Learning - Volume 37*, ICML'15, pages 1180–1189. JMLR.org, 2015.

[6] M. Gheisari and M. S. Baghshah. Unsupervised domain adaptation via representation learning and adaptive classifier learning. *Neurocomput.*, 165(C):300–311, Oct. 2015.

[7] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[8] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'14, pages 2672–2680, Cambridge, MA, USA, 2014. MIT Press.

[9] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition.

[10] J. Hu, J. Lu, Y.-P. Tan, and J. Zhou. Deep transfer metric learning. *Trans. Img. Proc.*, 25(12):5576–5588, Dec. 2016.

[11] P. Isola, J. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. *CoRR*, abs/1611.07004, 2016.

[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[13] Y. LeCun and C. Cortes. MNIST handwritten digit database. 2010.

[14] M. Liu and O. Tuzel. Coupled generative adversarial networks. *CoRR*, abs/1606.07536, 2016.

[15] M. Long, J. Wang, and M. I. Jordan. Deep transfer learning with joint adaptation networks. *CoRR*, abs/1605.06636, 2016.

[16] M. Long, J. Wang, and M. I. Jordan. Unsupervised domain adaptation with residual transfer networks. *CoRR*, abs/1602.04433, 2016.

[17] B. Moiseev, A. Konev, A. Chigorin, and A. Konushin. Evaluation of traffic sign recognition methods trained on synthetically generated data. In J. Blanc-Talon, A. Kasinski, W. Philips, D. Popescu, and P. Scheunders, editors, *Advanced Concepts for Intelligent Vision Systems*, pages 576–583, Cham, 2013. Springer International Publishing.

[18] Y. Netzer, T. Wang, A. Coates, A. Bissacco, B. Wu, and A. Y. Ng. Reading digits in natural images with unsupervised feature learning. 2011.

[19] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang. Domain adaptation via transfer component analysis. In *Proceedings of the 21st International Jont Conference on Artifical Intelligence*, IJCAI'09, pages 1187–1192, San Francisco, CA, USA, 2009. Morgan Kaufmann Publishers Inc.

[20] X. Peng and K. Saenko. Synthetic to real adaptation with deep generative correlation alignment networks. *CoRR*, abs/1701.05524, 2017.

[21] K. Saenko, B. Kulis, M. Fritz, and T. Darrell. Adapting visual category models to new domains. In *European conference on computer vision*, pages 213–226. Springer, 2010.

[22] W. sheng Chu Fern. Selective transfer machine for personalized facial action unit detection.

[23] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[24] J. Stallkamp, M. Schlipsing, J. Salmen, and C. Igel. Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition. *Neural Networks*, (0):–, 2012.

[25] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, et al. Going deeper with convolutions.

[26] E. Tzeng, J. Hoffman, T. Darrell, and K. Saenko. Simultaneous deep transfer across domains and tasks. *CoRR*, abs/1510.02192, 2015.

[27] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell. Adversarial discriminative domain adaptation. *CoRR*, abs/1702.05464, 2017.

[28] E. Tzeng, J. Hoffman, N. Zhang, K. Saenko, and T. Darrell. Deep domain confusion: Maximizing for domain invariance. *CoRR*, abs/1412.3474, 2014.

[29] Z. Wu, S. Song, A. Khosla, X. Tang, and J. Xiao. 3d shapenets for 2.5d object recognition and next-best-view prediction. *CoRR*, abs/1406.5670, 2014.