Research Article

Study of Deep Learning Approach for Improving Semi-Supervised Algorithm

Hemlata P. Sahu[1], Dr. Ramgopal Kashyap[2]

[1]Department of Computer Science & Engineering
Amity University Chhattisgarh
Raipur, India
hema.sahu88@gmail.com
Department of Computer Science & Engineering
Amity University Chhattisgarh
Raipur, India
rkashyap@rpr.amity.edu

Article History: Received: 11 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 23 May 2021

Abstract: When trained on the extensive collections of labeled data, Deep neural networks have demonstrated their potential for a good variety of supervised learning activities to deliver impressive performance. However, it takes a significant amount of money, time, and energy to construct such large data sets. In many practical situations, such tools might not be available, restricting the adoption and implementing many deep learning methods. Deep neural network-based semi-supervised learning implementations increasingly engaged in research to scale down the amount of labeled data needed, either by creating new methods or introducing current semi-supervised deep learning frameworks. To beat the need for giant annotated datasets, for more data-efficient deep learning methods. This article concentrated on a detailed analysis of semi-supervised learning focused on deep learning, accompanied by its description.

Keywords: *S*emi-supervised learning(SSL), deep learning, neural networks, SSL methods, SSL approaches, SSL evaluation

1. Introduction

In the presence of both labeled and unlabeled information, semi-supervised learning maybe a learning paradigm for studying how machines and natural systems, such as humans, learn. One provided a data point set consisting of a particular input x and a corresponding output value in supervised learning. The objective is to build a classifier to predict the production of previously unseen inputs. At the same time, the exact value of output is not given in unsupervised learning. Instead, one tries to derive the fundamental function from the given inputs. For example, in unsupervised cluster analysis, the aim is to assume that the provided information is mapped into groups mapped to an analog group by specific identical inputs (Bishop 2006). A machine learning activity that aims to integrate these two activities may be semi-supervised learning. In one of these two tasks, semi-supervised learning usually aims to increase efficiency by using information commonly associated with the opposite.

Deep learning is an element of a broader family of machine learning techniques for machine learning focused on artificial neural networks with representation learning. In deep learning, the term "deep" refers to the use of several layers within the network. Early research showed that a linear perceptron can not be a universal classifier but can be the opposite of a network with a non-polynomial activation function with a hidden layer of unbounded duration. Artificial neural networks, specifically convolutionary neural networks (CNN) as shown in figure 1, are the foundation of recent deep learning models. Still, even deep generative models, along with deep convolutional nodes and deep Boltzmann networks, would have to include evaluative equations or implicit layer-wise ordered variables [1]. Multilayer feedforward ANNs are biologically inspired forms of CNNs. By stacking multiple convolutional (Conv) layers that measure convolution activity over sub-regions of the input image or the upper layer's output, visual characteristics are learned from low to high levels. The sub-regions correspond to cells in the visible area of the so-called receptive fields of cats[20].

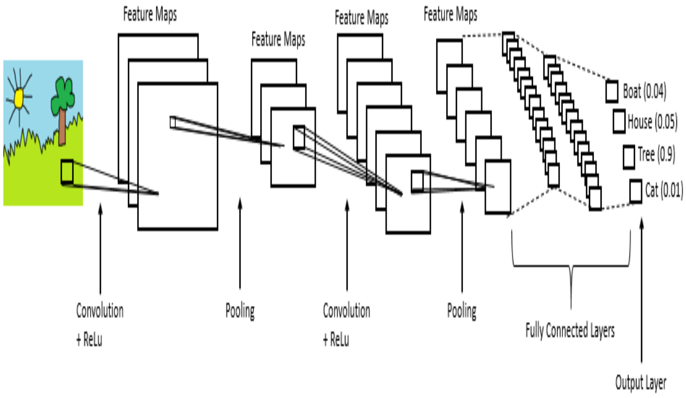

Figure 1. Complete CNN architecture

Semi-supervised learning refers to the process of machine learning, which combines a touch quantity of labeled data, including an out sized volume of unlabeled data during training. Among unsupervised learning and controlled learning, semi-supervised learning falls. The most simple drawback to A certain Supervised Learning Algorithm ensures that either a Machine Learning Engineer or a Knowledge Scientist would hand-label the information package. It is also a costly technique, especially when processing large volumes of data. The limited scope of execution is the main critical drawback to all unsupervised instruction.

The principle of semi-supervised learning became introduced during this learning process to address these drawbacks, and the algorithm was implemented on a confluence of classified and unclassified results.

2. Literature Review

This segment's primary focus is to study the various types of semi-supervised approaches supported by deep learning algorithms.

X. Bi, C. Zhang, X. Zhao, D. Li, Y. Sun, and Y. Ma addresses the most challenges of social stream classification by proposing an efficient incremental semi-supervised classification method named CODES (Classification Over Drifting and Evolving Stream). The proposed CODES approach consists of a useful semi-supervised total learning module and a module for the dynamic novelty threshold update[2].

L. Wang, N. Guan, D. Shi, Z. Fan, and L. Su, has proposed to beat the deficiency above, an excellent semi-supervised NMF model (RSS-NMF). To promote approximation, RSS-NMF uses the L2/L1-norm and makes the model insensitive to outliers by preventing them from dominating the value function[3].

Dupre, J. Fajtl, V. Argyriou, and P. Remagnino proposed a unique semi-supervised learning technique that, in conjunction with learned threading techniques and an ensemble decision network, is based on a simple iterative learning loop. In combination with learning threading methods and an ensemble decision network, the main goal was to implement a straightforward iterative learning loop. State-of-the-art model efficiency and increased volume training are shown by using unlabeled data in the training of deeply learned classification models and improving the trained models' accuracy [4].

J. Wan and Y. Wang develops a unified cost-sensitive semi-supervised face recognition architecture, which will jointly iteratively refine the inferred label information and hence the classifier. Here experiments on face benchmark data sets show that the proposed approach can significantly improve the overall system efficiency, especially in terms of classification errors related to high costs, compared to the state-of-the-art techniques for label propagation and cost-sensitive learning. This paper's most important feature is designing cost-sensitive latent semantic regression to integrate. It allows for a cost-sensitive label propagation process that iteratively updates the estimated labels with the learned classifier data[5].

R. Fierimonte, S. Scardapane, A. Uncini, and M. Panella proposed within the context of MR a fully decentralized algorithm for SSL. The essence of this proposal is a distributed protocol designed to compute the Laplacian matrix. A fully distributed calculation of the training patterns' adjacency matrix is the critical component of the proposed algorithm. At the present end of the proposal, the diffusion adaptation system [6] was assisted by a unique algorithm for low-rank distributed matrix completion.

J. Zhang and Y. Peng proposed to perform simplified hash function learning by simultaneously retaining semantic similarity and underlying data structures, the semi-supervised deep hashing technique. To jointly reduce the empirical error on labeled data, the authors suggested a semi-supervised loss, also because the embedding error on both labeled and unlabeled data can maintain the semantic similarity and catch the meaningful neighbors for efficient hashing on the underlying data structures[7].

L. Zhang, D. Zhang, X. Yin, and Y. Liu, the concerns of this paper are threefold: several distinct modalities often defining each function; it is difficult and hardly impossible to manually mark sensory data in a real application, resulting in a problem of inadequate labeled data; and classifier learning is generally independent of feature engineering[8].

A. H. Akbarnejad and M. S. Baghshah, proposed an embedding-based process that integrates non-linear label vectors using a stochastic approach, thus more accurately predicting the tail labels[9].

Yanbei Chen, Xiatian Zhu, Shaogang Gong described typical pipeline classification utilizing the function of cross-entropy loss, trained with label supervision, adding a memory module and amnesia as shown in figure 1. Memorizes the representation of class-level processes and ambiguity of model inference (key-value). SSL, performed in every training iteration, is achieved through assimilation-accommodation steps. Using a memory module with a multi-user key-value[10].

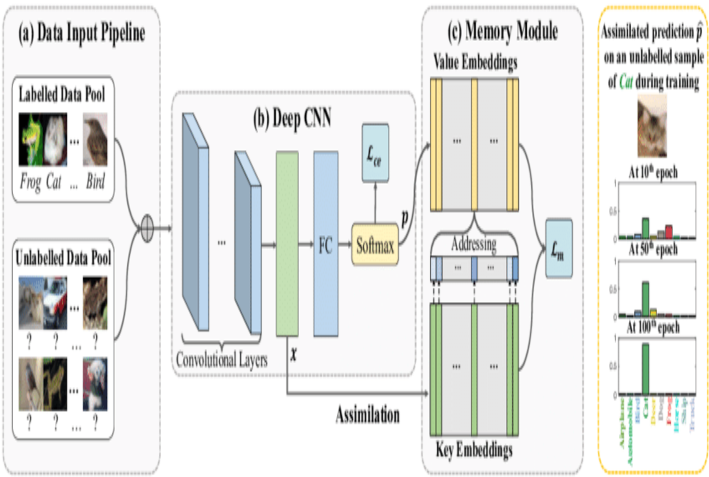

Figure 2. The memory-based deep learning network for Semi-Supervised Learning[10]

Classification of social streams was researched thoroughly. In Some scholars carried out supervised learning experiments earlier, techniques for classifying social documents. Support Vector help[23] has been used as a training set to think about the socially evaluated social stream features[22]. Model Naive Bayes is requested recognition of incidents from media platforms[21]. MuENL[4] uses normalized SVMs and two approaches for upgrading. However the tree-structured helps to assess the cost of processing to be high. In addition, in the social stream, evolving documents may not have mark data for more supervised learning, drifting and changing principles and detection.

In recent years, detailed studies have said the efficiency of clustering could be vastly enhanced if a large volume of prior information about the data is incorporated[24,25]. Semi-supervised classification is applied in order to fairly use the contextual information data samples. In general, there seem to be two methodologies of semi-supervised clustering: Constraint-based these approaches apply knowledge about constraint constraints to the clustering process. Semi-supervised cluster - based processes depending on distance exploit a basic distance parameter to fulfill previous comprehension and awareness. A current clustering algorithm is being used to understand the correlation between points of data.

Weston et al[26] is inspired by the early progress of deep learning to discuss it in a functionally clear way for semi-supervised learning. Weston et al. picks an unattended individual Or even a semi-supervised algorithm of learning, including the Mark Zhu and Ghahramani propagation[27] and LapSVM mutation by Belkin et al.[28], then link it to an internal deep profile also as regularized a single or several layer model frameworks. This model, with labelled and unlabelled components, they are then equally trained. The authors have shown that semi-supervised encoding, testing with just a deep, multi-layer framework tracked on every In complex activities, network layers can offer real benefits.

3. Semi Supervised Learning Approaches

Many SSL methods and techniques have been implemented over the years. These algorithms are also commonly divided into the following categories:

Consistency Regularization - It supports the idea that if a wise disturbance is an outlook, it will not adjust significantly as extended to unlabeled data points. On a given unlabelled example and its disrupted version, the model can trained.

Generative models - Almost the same as the supervised environment, where on one task, the learned characteristics are often passed onto another downstream task. Generative models that can perform other downstream duties. For a given task with targets y, generative models capable of generating Images from the p(x) data distribution must gain transferable features for a p(y|x) supervised task.

Proxy-label Approaches - To provide further training examples by labeling instances of those heuristics supported by the unlabeled set, these techniques use a trained model on the labeled set. These methods can also be cited as bootstrapping[11] algorithms. We are watching and asking them for proxy-label techniques from Ruder et al.[12].

Graph-Based Approach - Graph nodes are used to represent both labeled and unlabeled data points. Thus, they are intended to use the associations between xi and xj nodes, specified by the sting frequency between the two nodes, to distribute the labels of the named nodes to the unlabeled nodes. There is also some SSL work on entropy minimization, and these main groups push the model by decreasing the projections' entropy to make confident predictions.

Consistent planning may also, with a small modification, instead of accepting the projections as ground-truths and calculating cross-entropy loss with a slight change, are used as a proxy-label tool. It imposes consistency by minimizing the difference between the outputs, forecasts. Two dominant learning paradigms, transductive and inductive learning, are used to promote SSL methods. During this case, In the unlabeled instances found at training time, transductive learning tends to use the qualified classifier; it does not generalize to unnoticed models. On maps, such as embedding random nodes[13, 14] In specific, this type of algorithm is used where the goal is to mark the unlabeled graph nodes present at the training time.

4. Semi Supervised Learning Assumptions

Any reference to the underlying distribution of the information must exist to allow some use of unlabeled data. Among the corresponding assumptions, semi-supervised learning algorithms use a minimum of one:

Continuity presumption - A label is more likely to be shared by points on the edge of one another[15]. In supervised learning, it is commonly believed and gives priority to symmetrically basic decision making. The smoothness assumption also provides an option for low-density decision limits in semi-supervised learning in low-density areas. The smoothness assumption also provides a preference for low-density regions in semi-supervised learning for decision limits. Few points are on the edge of each other, but in many groups.

Manifold hypothesis - The data lies roughly on a multiple of a much smaller dimension than the input space[15]. The curse of dimensionality can be avoided by learning the manifold, which uses both the data, labeled and unlabeled using distances and densities defined on the manifold; learning can then proceed as shown in figure 3.

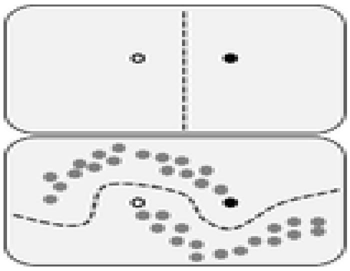

Figure 3. An example of the effect of unlabeled data in semi-supervised learning

When a mechanism creates high-dimensional data may be challenging to model, it just has a few flexibilities; the multiple assumptions are practical. For example, a few vocal folds control the human voice[15], and A few muscles regulate pictures of varying facial expressions. Distances and smoothness in these instances within the natural area of the issue of generation is superior to that consideration, respectively, of the space of all potential acoustic waves or images.

5. Related Issues

In active learning[32, 33], a wide collection of unlabeled sample points is supplied mostly with learning algorithm, also with ability to interactively advise that every illustration given from the unlabeled range be labelled. Unlike in classical distance learning, where the samples to be labelled are immediately picked again with the unlabeled collection, active learning attempts to use the samples on being labeled explicitly again from unlabeled collection, active learning attempts to specifically select the samples to be labeled in order to gain greater specificity while using these kind of requests as available, thus reducing the expense of acquiring labeled data. This is of special importance in issues where there might be ample knowledge, but marks are sparse or costly to procure.

Although a universally successful active learning strategy[34] cannot be accomplished, there are various heuristics[33] that are said to be successful. Information quality and relative importance[35] are the two commonly known selection standards. Based on the activities tests whether an unlabeled instance helps to minimize a mathematical model's ambiguity, while overconfidence measures whether an instances helps to describe the new input structure.

As all seek to use a finite number of data to maximize a learner, active learning and SSL are inherently related. Several reports have suggested integrating SSL and Active learning for various activities. [36] shows a substantial reduction of errors with minimal labelled voice comprehension results, [37] suggests an sucessful semi-supervised pedestrian recognition learning system, [38] incorporates AL and SSL utilizing synthetic dataset binomial fields and[39] leverage SSL to refine knowledge by unlabeled data using both labeled and unlabeled data, which enhances learning algorithm and sample collection.

Transfer learning[40] is substance utilized by moving the information acquired from a similar domain, known to as the target domain, to boost a learner in one field, labeled the goal domain. For example, we will want to validate the algorithm on a generic, inexpensive data to produce, on real results, with the purpose of doing this. Within that case, the target domain always seemed to construct the computer is the target domain, but it differs mostly from given target domain required to fully test the model. When this source and purpose vary but are associated,, but are associated, then it is possible to apply transfer learning to achieve greater precision mostly on target data.

Domain adaptation [41] is one common method in transfer learning. A form of multi - label transfer learning is domain adaptation, in which the main goal stays as much as a history, but the meaning varies. The purpose of transfer learning have to train an efficient learner in the generalization of various distributions throughout multiple domains in which data from the named source domain is available. As with the target domain, consider the example of the absence of secret data as unmonitored transfer learning on target, while semi-supervised and supervised transfer learning refers to circumstances in which we receptively have a target domain that is small or completely labeled[29].

SSL and the unsupervised transformation of features are closely linked; we are supplied with lable, unlable in both cases. A feature capable of making assumptions to that same unlable data sets and unseen examples with the purpose of understanding. However, even that labelled and unlabeled are within the same source in SSL, while the target as well as source variables vary from one source to another unsupervised transfer learning. Methods can be interchangeably leveraged with all subjects. In SSL[42], it was indicated that exclusionary propagation alignment could be used for semi-supervised object recognition using just a limited number of samples labelled for semi-supervised processes, such as consisting of unmonitored domain adaptation. Semi-supervised approaches, such as uniformity regularization[43] have demonstrated their efficacy in transfer learning with regard to unsupervised transfer learning.

Most large deep learning models use some sort of poor monitoring to address any need for big hand-labeled and costly training sets: lower-quality, but large number of training sets designed by techniques or to use cheap annotators[44]. The target would be the same in weakly supervised learning as in deep classification, except with a ground-truth identified test data, one or even more similar base examples are given that may emerge by crowd workers, if it is the development of Bayesian rules, the production of separate oversight, or the usefulness of several other classifiers. For example, incorrect annotations are replaced by pixel-level labels that are harder and harder to obtain. Another example, pixel-level labels that are tougher and more difficult to procure are substituted for incorrect annotations in weakly controlled semantic segmentation[45] and bounding boxes [46]. Somehow in a situation, if a small number of highly labelled data are present while still looking to take advantage of the loosely labelled data, SSL approaches may be used to further increase performance.

Training from disruptive labels[47] can be problematic because noise can have a detrimental effect on the output of deep learning models if the noise is large. Many existing approaches for based on deep learning with noisy labels aim to overcome the loss function in order to solve this. The treatment of most of the examples as equivalent and redefining the disruptive examples becomes the form of correction, although proxy labeling approaches would be used for the redefining procedure. A re-weighing including its training set to discriminate between clear and disruptive samples is another form of correction.

6. Evaluating Semi Supervised Learning Approaches

The conventional method of evaluating SSL methods is to pick a dataset typically used for supervised learning[16,17,18] and then disregard an outsized tiny labeled set. A more comprehensive unlabeled set results in a portion of the labels. With a given SSL methodology, a deep learning model is learned and, then, the results are reported over separate and ordered parts of names examples on the first test range. For this to be specified, this protocol for real-world environments, Oliver et al.[19] proposed the following methods for developing this experimental technique: a collective implementation.

When conducting an assessment, as both sets come from an identical dataset, The potential mismatch in delivery between labeled and unlabeled examples is often ignored. In real-world implementations, such a discrepancy is widespread; compared to categorize data, unlabeled data seem to have specific classification results. For improved SSL acceptance in the real world, we discuss the modifications labeled and unlabeled data quantity. The number of labeled instances differs from a standard SSL process. Still, it may also provide additional insights into SSL approaches efficacy by simulating practical situations, such as practicing on a comparatively limited unlabeled range, to vary the dimensions during a systematic process. Small Validation Sets Realistically. In such cases, If a thoroughly annotated dataset ends with robust hyperparameter tuning can result in overfitting, with a test data set that is considerably more significant than the labeled set used for testing or assessment.

7. Self- Supervision For Semi Supervised Learning

A kind of unsupervised learning is self-supervised learning[48], where a normal supervised error is used to train the model, but on a pretext assignment for which supervision comes again from knowledge itself. In this case, the goal is not to optimize final output mostly on pretext task, but then to acquire rich and transferable characteristics for extracted features. A number of pretext tasks are suggested, where the model is first trained using unlabeled examples with one or maybe more tasks, the final output is either used to produce raw data representations, which are used to train a shallow classifier on Dl Or explicitly quite well with named images for a specific task. Instances of such data acquisition pretext tasks are:

Specimen-CNN[49]. A set of N patches are formed with random mutations for an input dataset, most of such patches are then treated as a different class, and the classifier is intended to predict an input data patch., the right class.

The [50] rotation. A provided rotation is performed on the input object out of following four iterations of multiples with 90 and the model is equipped to predict the right rotation that has been applied.

The [51]patches. A first pattern is isolated from the reference map uniformly, such patch is referred to with minor volatility at eight neighboring sites, as its base, and eight distinct neighboring and semi trends are separated, so the model becomes taught to approximate the presence of many of the neighboring patches first. Additional variants of this context challenge have been proposed, including the jigsaw puzzle[52] in which the patches being randomly permutated into some other system and the aim is to anticipate the appropriate permutation added to achieve the proper permutation. that has been applied to get the patches ordered correctly.

The [53] colorization. First, the input image is modified from RGB to Grayscale, the input is fed further into prototype just with the gray - level information found inside the L component or even the pixel values with components, and the aim is to estimate the remainder including its information. By utilizing the test color space, the challenge will either be viewed as a correlation issue or a classification problem.

Predictive Coding Contrastive[54]. Using a compare contributes to failure of Noise Contrastive Assessment[60] as well as its current derivatives along with Momentum Contrast[55] and SimCLR[56], the positive could be an input data picture and its modified iterations or a defined patch and also its neighboring patches, while the neutral is randomly chosen images. The prototype is taught to discriminate between nice and negative samples.

Such contextual practices can be effectively used for SSL, in which the network is always trained on the context aspect by self-supervision and self-supervision on the entire dataset as well, but instead tuned to just the labelled collection using usual cross-entropy loss, both continuously as seen by[57] with rotation as that of the pretext task, or iteratively, first before building the classifier with self-supervision and then precise refining it on.

8. Conclusion

Just a few of the various semi-supervised learning methods were explored during this survey. The strengths and limitations of existing methods[4] must be understood to allow full use of the capacity of unlabeled data. Data labeled earlier, as mentioned, is costly and challenging to get. Unlabeled data, on the other side, are comparatively easy to obtain. Semi-supervised learning often does not identify unlabeled data, and better classifiers are also often not created. Semi-supervised learning requires less human labor and delivers a much better result than its unsupervised and supervised counterparts. Because of this benefit, both in theory and in practice, semi-supervised learning is of great importance.

Acknowledgment

We would wish to thank Dr. Rahamatkar, Director, Amity School of Engineering and Technology, who motivated and provided the research facilities. We are very grateful to the anonymous referees for his or her valuable suggestions to enhance this paper. We are also thankful to the Amity University Chhattisgarh, Raipur, India, to promote research activities.

References

1. Bengio, Yoshua (2009). "Learning Deep Architectures for AI". Foundations and Trends in Machine Learning. (1):1127. CiteSeerX 10.1.1.701.9550. doi:10.1561/2200000006. Archived from the original (PDF) on 2016-03-04. Retrieved 2015-09-03.
2. X. Bi, C. Zhang, X. Zhao, D. Li, Y. Sun and Y. Ma, "CODES: Efficient Incremental Semi-Supervised Classification Over Drifting and Evolving Social Streams," in IEEE Access, vol. 8, pp. 14024-14035, 2020.
3. L. Wang, N. Guan, D. Shi, Z. Fan, and L. Su, "Robust Semi-Supervised Non-Negative Matrix Factorization With Structured Normalization," in IEEE Access, vol. 7, pp. 133996-134013, 2019
4. R. Dupre, J. Fajtl, V. Argyriou, and P. Remagnino, "Improving Dataset Volumes and Model Accuracy With Semi-Supervised Iterative Self-Learning," in IEEE Transactions on Image Processing, vol. 29, pp. 4337-4348, 2020.
5. J. Wan and Y. Wang, "Cost-Sensitive Label Propagation for Semi-Supervised Face Recognition," in IEEE Transactions on Information Forensics and Security, vol. 14, no. 7, pp. 1729-1743, July 2019.
6. R. Fierimonte, S. Scardapane, A. Uncini, and M. Panella, "Fully Decentralized Semi-supervised Learning via Privacy-preserving Matrix Completion," in IEEE Transactions on Neural Networks and Learning Systems, vol. 28, no. 11, pp. 2699-2711, Nov. 2017.
7. J. Zhang and Y. Peng, "SSDH: Semi-Supervised Deep Hashing for Large Scale Image Retrieval," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 29, no. 1, pp. 212-225, Jan. 2019.
8. L. Zhang, D. Zhang, X. Yin, and Y. Liu, "A Novel Semi-Supervised Learning Approach in Artificial Olfaction for E-Nose Application," in IEEE Sensors Journal, vol. 16, no. 12, pp. 4919-4931, June 15, 2016.
    A. H. Akbarnejad and M. S. Baghshah, "An Efficient Semi-Supervised Multi-label Classifier Capable of Handling Missing Labels," IEEE Transactions on Knowledge and Data Engineering, vol. 31, no. 2, pp. 229-242, February 1 2019.
9. Yanbei Chen, Xiatian Zhu, Shaogang Gong, Semi-Supervised Deep Learning with Memory, ECCV'18

10. Christian Biemann. Unsupervised and knowledge-free natural language processing in the structure discovery paradigm. Ph.D. thesis, Leipzig University, Germany, 2007
11. Ellen Riloff. Automatically generating extraction patterns from untagged text. In Proceedings of the national conference on artificial intelligence, pages 1044–1049, 1996
12. Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks. In Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining, pages 855–864, 2016
13. Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. Deepwalk: Online learning of social representations. In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 701–710, 2014.
14. Olivier Chapelle, Bernhard Scholkopf, and Alexander Zien. Semi-supervised learning (chapelle, o. et al., eds.; 2006)[book reviews]. IEEE Transactions on Neural Networks, 20(3):542–542, 2009
15. Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition, pages 248–255. Ieee, 2009
16. Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. 2009.
17. Yuval Netzer, Tao Wang, Adam Coates, Alessandro Bissacco, Bo Wu, and Andrew Y Ng. Reading digits in natural images with unsupervised feature learning. 2011.
18. Avital Oliver, Augustus Odena, Colin A Raffel, Ekin Dogus Cubuk, and Ian Goodfellow. Realistic evaluation of deep semi-supervised learning algorithms. In Advances in Neural Information Processing Systems, pages 3235–3246, 2018.
19. Avrim Blum and Tom Mitchell. Combining labeled and unlabeled data with co-training. In Proceedings of the eleventh annual conference on Computational learning theory, pages 92–100, 1998
20. C. Trabelsi and S. Yahia, "A probabilistic approach for events identificationfrom social media rss feeds," Database Systems for Advanced Applications, vol. 7827, pp. 139–152, 2013.
    A. Zubiaga, D. Spina, R. Martínez, and V. Fresno, "Real-time classification of twitter trends," journal of the Association for Information Science and Technology, vol. 66, no. 3, pp. 462–473, 2015.
    B. Cortes and V. Vapnik, "Support vector networks," Machine Learning,vol. 20, pp. 273–297, 1995.
21. Z. Fu, K. Ren, J. Shu, X. Sun, and F. Huang, ''Enabling personalized search over encrypted outsourced data with efficiency improvement,'' IEEE Trans. Parallel Distrib. Syst., vol. 27, no. 9, pp. 2546–2559, Sep. 2016
22. J. Zhao, X. Shang, and H. Zhang, ''Recovering seabed topography from sonar image with constraint of sounding data,'' J. China Univ. Mining echnol., vol. 46, no. 2, pp. 443–448, 2017
23. J. Weston, F. Ratle, H. Mobahi, and R. Collobert, "Deep Learning viaSemi-Supervised Embedding," in Neural Networks: Tricks of the Trade.Springer, 2009, pp. 639–655.
24. X. Zhu and Z. Ghahramani, "Learning from Labeled and Unlabeled Data with Label Propagation," Carnegie Mellon University Technical Report, no. CMU-CALD-02-107, 2002.
25. M. Belkin, P. Niyogi, and V. Sindhwani, "Manifold regularization: A geometric framework for learning from labeled and unlabeled examples," Journal of machine learning research, vol. 7, no. Nov, p. 2399–2434, 2006.
26. Oscar Beijbom. Domain adaptations for computer vision applications. arXiv preprint arXiv:1211.4860, 2012
27. Martin Cooke, Phil Green, Ljubomir Josifovski, and Ascension Vizinho. Robust automatic speech recognition with missing and unreliable acoustic data. Speech communication, 34(3):267–285, 2001
28. Jifeng Dai, Kaiming He, and Jian Sun. Boxsup: Exploiting bounding boxes to supervise convolutional networks for semantic segmentation. 2015 IEEE International Conference on Computer Vision (ICCV), Dec 2015

29. Steve Hanneke. Theoretical foundations of active learning. Technical report, CARNEGIE-MELLON UNIV PITTSBURGH PA MACHINE LEARNING DEPT, 2009
30. Burr Settles. Active learning literature survey. Technical report, University of Wisconsin-Madison Department of Computer Sciences, 2009.
31. Sanjoy Dasgupta. Analysis of a greedy active learning strategy. In Advances in neural information processing systems, pages 337–344, 2005.
32. Sheng-Jun Huang, Rong Jin, and Zhi-Hua Zhou. Active learning by querying informative and representative examples.In Advances in neural information processing systems, pages 892–900, 2010
33. Thomas Drugman, Janne Pylkkonen, and Reinhard Kneser. Active and semi-supervised learning in asr: Benefits on the acoustic and language models. arXiv preprint arXiv:1903.02852, 2019
34. Phill Kyu Rhee, Enkhbayar Erdenee, Shin Dong Kyun, Minhaz Uddin Ahmed, and Songguo Jin. Active and semisupervised learning for object detection with imperfect data. Cognitive Systems Research, 45:109–123, 2017.
35. Xiaojin Zhu, John Lafferty, and Zoubin Ghahramani. Combining active learning and semi-supervised learning using gaussian fields and harmonic functions. In ICML 2003 workshop on the continuum from labeled to unlabeled data in machine learning and data mining, volume 3, 2003.
36. Mingfei Gao, Zizhao Zhang, Guo Yu, Sercan O Arik, Larry S Davis, and Tomas Pfister. Consistency-based semi-supervisedactive learning: Towards minimizing labeling cost. arXiv preprint arXiv:1910.07153, 2019.
37. Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. IEEE Transactions on knowledge and data engineering,22(10):1345–1359, 2009.
38. Vishal M Patel, Raghuraman Gopalan, Ruonan Li, and Rama Chellappa. Visual domain adaptation: A survey of recent advances. IEEE signal processing magazine, 32(3):53–69, 2015.
39. Christoph Mayer, Matthieu Paul, and Radu Timofte. Adversarial feature distribution alignment for semi-supervised learning. arXiv preprint arXiv:1912.10428, 2019
40. Geoffrey French, Michal Mackiewicz, and Mark Fisher. Self-ensembling for visual domain adaptation. arXiv preprintarXiv:1706.05208, 2017
41. Alex Ratner, P Varma, B Hancock, and Chris Ré. Weak supervision: A new programming paradigm for machine learning (2019).
42. Amy Bearman, Olga Russakovsky, Vittorio Ferrari, and Li Fei-Fei. What's the point: Semantic segmentation with point supervision. In European conference on computer vision, pages 549–565. Springer, 2016.
43. Jifeng Dai, Kaiming He, and Jian Sun. Boxsup: Exploiting bounding boxes to supervise convolutional networks for semantic segmentation. 2015 IEEE International Conference on Computer Vision (ICCV), Dec 2015
44. Salvador García, Julián Luengo, and Francisco Herrera. Data preprocessing in data mining. Springer, 2015
45. Lilian Weng. Self-supervised representation learning. lilianweng.github.io/lil-log, 2019.
46. Alexey Dosovitskiy, Philipp Fischer, Jost Tobias Springenberg, Martin Riedmiller, and Thomas Brox. Discriminative unsupervised feature learning with exemplar convolutional neural networks. IEEE transactions on pattern analysis and machine intelligence, 38(9):1734–1747, 2015.
47. Spyros Gidaris, Praveer Singh, and Nikos Komodakis. Unsupervised representation learning by predicting image rotations. arXiv preprint arXiv:1803.07728, 2018
48. Carl Doersch, Abhinav Gupta, and Alexei A Efros. Unsupervised visual representation learning by context prediction.In Proceedings of the IEEE international conference on computer vision, pages 1422–1430, 2015
49. Mehdi Noroozi and Paolo Favaro. Unsupervised learning of visual representations by solving jigsaw puzzles. In European Conference on Computer Vision, pages 69–84. Springer, 2016

50. Richard Zhang, Phillip Isola, and Alexei A Efros. Colorful image colorization. In European conference on computer vision, pages 649–666. Springer, 2016.
51. Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. arXivpreprint arXiv:1807.03748, 2018
52. Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. arXiv preprint arXiv:1911.05722, 2019.
53. Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. arXiv preprint arXiv:2002.05709, 2020.
54. Xiaohua Zhai, Avital Oliver, Alexander Kolesnikov, and Lucas Beyer. S4l: Self-supervised semi-supervised learning. In Proceedings of the IEEE international conference on computer vision, pages 1476–1485, 2019.