A Review of Deep Transfer Learning and Recent Advancements

Mohammadreza Iman
Department of Computer Science, Franklin College of Arts and Sciences
University of Georgia
Athens, GA, USA
m.iman@uga.edu

Khaled Rasheed
Institute for Artificial Intelligence, Franklin College of Arts and Sciences
University of Georgia
Athens, GA, USA
khaled@uga.edu

Hamid R. Arabnia
Department of Computer Science, Franklin College of Arts and Sciences
University of Georgia
Athens, GA, USA
hra@uga.edu

Abstract

A successful deep learning model is dependent on extensive training data and processing power and time (known as training costs). There exist many tasks without enough number of labeled data to train a deep learning model. Further, the demand is rising for running deep learning models on edge devices with limited processing capacity and training time. Deep transfer learning (DTL) methods are the answer to tackle such limitations, e.g., fine-tuning a pre-trained model on a massive semi-related dataset proved to be a simple and effective method for many problems. DTLs handle limited target data concerns as well as drastically reduce the training costs. In this paper, the definition and taxonomy of deep transfer learning is reviewed. Then we focus on the sub-category of network-based DTLs since it is the most common types of DTLs that have been applied to various applications in the last decade.

Introduction

Deep learning turned out to be the answer to many machine learning problems in the last two decades by addressing the nonlinearity of many datasets. Recent advancements in deep learning methods deliver various usages and applications in extremely different areas such as image processing, natural language processing (NLP), numerical data analysis and predictions, and voice recognition. However, deep learning comes with restrictions, such as expensive training processes (time and processing) and the requirement of extensive training data (labeled data) [1].

Since the start of the Machine Learning (ML) era, transfer learning has been a neat exploration for scientists. Before the rise of deep learning models, transfer learning was known as domain adaptation and focused on homogenous data sets and how to relate such sets to each other because of the nature of ML algorithms [2][3]. Traditional ML models have less dependency on dataset size, and usually, their training is less costly than deep learning models since they have been mostly designed for linear problems. Therefore, the motivation for using transfer learning in deep learning is higher than ever in the AI (Artificial Intelligence) and ML fields since it can address the two restraints of extensive training data and training costs.

Recent transfer learning methods on deep learning aim to reduce training process time and cost, and the necessity of extensive training datasets which can be hard to harvest in some areas such as medical images. Moreover, a pre-trained model for a specific job can be run on a simple edge device like a cellphone with limited processing capacity and limited training time [4]. Also, developments in DTL are opening the door to more intuitive and sophisticated AI systems since it considers learning a continuous task. A great example of this idea is Google's deep mind project and advancements such as progressive learning [5].

In this paper, first the definition of DTL is reviewed followed by the taxonomy of DTL. Then, selected practical recent studies of DTL are listed, categorized, and summarized. Moreover, two experimental evaluations of DTL and their conclusions are reviewed.

Deep Learning

Deep learning (DL) or deep neural network (DNN) is a machine learning subcategory, which can deal with nonlinear datasets. DNNs consist of layers of stacked nodes, with activation function and associated weights, (fully/partially) connected and usually trained (weight adjustments) by back-propagation and optimization algorithms. During the past two decades, DNNs were developed rapidly and are used in many aspects of our daily lives today. For instance, Convolutional Neural Network (CNN) layers have improved deep learning models for visual-related tasks since 2011, and as of today, most DLs use CNN layers [1]. For more details about machine learning and deep learning, please refer to [1] since this paper is focused on deep transfer learning, and we assume that the reader should have a thorough understanding of machine learning and deep learning.

Deep Transfer Learning (DTL)

Deep transfer learning is about using the obtained knowledge from another task and dataset (even one not strongly related to the source task or dataset) to reduce learning costs. In many ML problems arranging a large amount of labeled data is impossible, which is mandatory for most DL models. For instance, at the beginning of the Covid-19 pandemic or even a year into it, providing enough chest X-Ray labeled data for training a deep learning model was still challenging, while using deep transfer learning, the AI achieved detecting the disease with very high accuracy with a limited training set [6] [7]. Another application is applying machine learning on edge devices such as phones for variant tasks by taking advantage of deep transfer learning to reduce the need for processing power.

An untrained DL uses a random initializing weight for nodes, and during the expensive training process, those weights adjust to the most optimized values by applying an optimization algorithm for a specific task (dataset). Remarkably, [8] proved that initializing those weights based on a trained network with even a very distant dataset improves training performance compared to the random initialization.

Deep transfer learning differs from semi-supervised learning since, in DTL, the source and target datasets can have a different distribution and just be related to each other, while in semi-supervised learning, the source and target data are from the same dataset, only the target set does not have the labels [2]. DTL is also not the same as Multiview learning since Multiview learning uses two or more distinct datasets to improve the quality of one task, e.g., video datasets can be separated into image and audio datasets [2]. Last but not least, DTL differs from Multitask learning despite many shared similarities. The most fundamental difference is that in Multitask learning, the tasks use

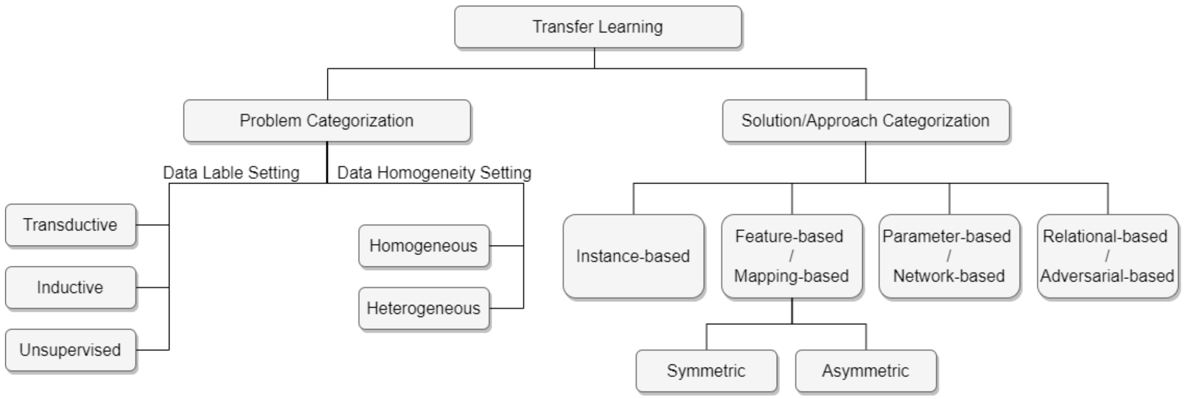

Figure 1. Taxonomy of Transfer Learning which is extendable to Deep Transfer Learning as well.

interconnections to boost each other, and knowledge transfer happens concurrently between related tasks. In contrast in DTL, the target domain is the focus, and the knowledge has already been obtained for target data from source data, and they do not need to be related or function simultaneously [2].

From Transfer Learning to Deep Transfer Learning, Taxonomy

It is possible to categorize Deep Transfer Learnings (DTLs) in different ways by various criteria, similar to Transfer Learnings. DTLs can be divided into two categories of homogeneous and heterogenous based on the homogeneity of source and target data [2]. However, this categorization can be done differently because it is subjective and relative. For example, a dataset of X-Ray photos can be considered heterogeneous to a dataset of tree species photos when the comparison domain is limited to only image data. In contrast, it can be considered homogeneous to the same tree species photo dataset when the domain consists of audio and text datasets.

Also, DTLs can be categorized into three groups based on label-setting aspects: (i) transductive, (ii) inductive, and (iii) unsupervised [2]. Briefly, transductive is when only the source data is labeled; if both source and target data are labeled it is inductive; if none of the data are labeled it is unsupervised deep transfer learning [2].

[2] and [9] mention and define another categorization of DTLs through the aspect of applied approaches. They similarly categorized DTLs into four groups of: (i) instance-based, (ii) feature-based / mapping-based, (iii) parameter-based / network-based, and (iv) relational-based / adversarial-based approaches. Instance-based transfer learning approaches are based on using selected parts of instances (or all) in source data and applying different weighting strategies to be used with target data. Feature-based approaches map instances (or some features) from both source and target data into more homogenous data. Further, the [2] survey divides the feature-based category into asymmetric and symmetric feature-based transfer learning subcategories. "Asymmetric approaches transform the source features to match the target ones. In contrast, symmetric approaches attempt to find a common latent feature space and then transform both the source and the target features into a new feature representation." [2] The network-based

(parameter-based) methods are about using the obtained knowledge in the model (network) with different combinations of pre-trained layers: freezing some and/or fine-tuning some and/or adding some fresh layers. Relational/adversarial-based approaches focus on extracting transferable features from both source and target data either using the logical relationship or rules learned in the source domain or by applying methods inspired by generative adversarial networks (GAN) [2][9]. Figure 1 shows the taxonomy of the above-mentioned categories [2].

Other than the network-based and adversarial-based approaches, all other categories have been explored deeply during the last couple of decades for different ML techniques known as domain adaptation or transfer learning [2][3]. However, most of those techniques are still applicable to deep transfer learning (DTL) as well. Network-based (parameter-based) approaches are the most applied techniques in DTL since they can tackle the domain adaptation between source and target data by adjusting the network (model). In other words, deep transfer learning is mainly focused on network-based approaches. Remarkably, network-based approaches in deep learning models can even tackle the adaptation of a very distant source and target data [2][9].

In deep transfer learning (DTL), different techniques are applied for network-based approaches, although generally, they are combinations of pre-training, freezing, fine-tuning, and/or adding a fresh layer(s). A deep learning network (DL model) trained on source data is called a pre-trained model consisting of pre-trained layers. Freezing and fine-tuning are techniques using some or all layers of pre-trained models to train the model on target data. Freezing some layers means the parameters/weights will not change and are constant values for frozen layers from a pre-trained model. Fine-tune means the parameters/weights are initialized with the pre-trained values instead of random initialization for the whole network or some selected layers. Another recent DTL technique is based on freezing a pre-trained model and adding new layers to that model for training on target data; Google's deep mind project introduces this technique in 2016 as Progressive Learning / progressive neural networks (PNNs) [5][10].

The concept of progressive learning mimics human skill learning, which is adding a new skill on top of previously learned skills as a foundation to learn a new one. E.g., a child learns how to run after learning to crawl and walk and using all the skills obtained in the process. Similarly, PNNs prevent catastrophic forgetting in DTL versus fine-tuning techniques by freezing the whole pre-trained model and learning (adjusting to) the new task by training the newly added layers on top of previously trained layers [5][10].

In deep learning models, usually, the earlier layers do the feature extraction at a high level of detail, further layers towards the end extract the information and conceptualize the given data, and lateral layers do the classifications or predictions. For instance, in the image-related model, the earlier layers of CNN extract the edges, corners, and tiny patches of a given image. Further layers put those details together to detect objects or faces, and the lateral layers, usually fully connected layers, do the classification [11]. Given this process, the most effective and efficient approach for DTL, to our knowledge, is to freeze the earlier and middle layers from a related pre-trained model and fine-tune the lateral layers for the new task/dataset [12]. Similarly, the new layers are added to the last part of a pre-trained model in progressive learning.

Nonetheless, some other research in this area use combinational and sophisticated methods to tackle transfer learning in deep learning like ensembled networks, weighting strategies, etc. [2]. However, to our knowledge, the search for recent advancements in DTL for practical tasks ends up with methods based on mostly the network-based and limited number of adversarial-based approaches.

Review of Recent Advancement in DTL

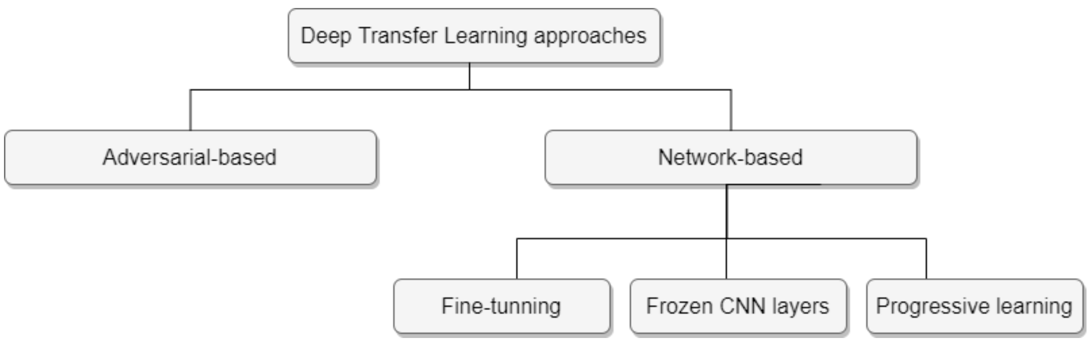

Figure 2. Most common Deep Transfer Learning approaches.

We limited our selection to the last 4-5 years of published studies on deep transfer learning for various tasks and data types. Table 1 shows the list of selected works from more than a hundred reviewed literature sorted by their DTL approaches. We used systematic literature review (SLR) technique [13] for the process of finding and selecting these thirty publications. We found that all reviewed studies mostly fall into three categories of network-based approaches and some into the adversarial-based approach, which are explained in the previous section. We name these approaches as (i) Fine-tuning: fine-tuning a pre-trained model on target data; (ii) Freezing CNN layers: the earlier CNN layers are frozen and only the lateral fully connected layers are fine-tuned; (iii) Progressive learning: some or all layers of a pre-trained model are selected and used frozen, and some fresh layers will be add to the model to be trained on target data; and (iv) Adversarial-based: extracting transferable features from both source and target data using adversarial or relational methods, Figure 2.

Table 1. List of selected recent DTL publications

Ref.	Year	Title	Data Type	Time Series	Approach	CNN Layers	Known Models Used	Dataset Field
[6]	2020	Automated Deep Transfer Learning-Based Approach for Detection of COVID-19 Infection in Chest X-rays	Image	no	Fine-tuning	yes	Inception-Xception	Medical image
[7]	2020	Classification of the COVID-19 infected patients using DenseNet201 based deep transfer learning	Image	no	Fine-tuning	yes	ImageNet, Dense-Net	Medical image
[14]	2019	Enhancing materials property prediction by leveraging computational and experimental data using deep transfer learning	Tabular/ bigdata	yes	Fine-tuning	no	NA	Quantum mechanics
[15]	2019	Application of deep transfer learning for automated brain abnormality classification using MR images	Image	no	Fine-tuning	yes	Res-Net	Medical image

[16]	2019	An adaptive deep transfer learning method for bearing fault diagnosis	Tabular/ bigdata	yes	Fine-tuning	no	LSTM RNN	Mechanic
[17]	2019	Online detection for bearing incipient fault based on deep transfer learning	Image	yes	Fine-tuning	yes	VGG-16	Mechanic
[18]	2019	Towards More Accurate Automatic Sleep Staging via Deep Transfer Learning	Tabular/ bigdata	yes	Fine-tuning	yes	NA	Medical data
[19]	2019	Deep Transfer Learning for Multiple Class Novelty Detection	Image	no	Fine-tuning	yes	Alex-Net, VGG-Net	Vision
[20]	2019	A Digital-Twin-Assisted Fault Diagnosis Using Deep Transfer Learning	Tabular/ bigdata	no	Fine-tuning	no	NA	Mechanic
[21]	2019	Learning to Discover Novel Visual Categories via Deep Transfer Clustering	Image	no	Fine-tuning	yes	NA	Vision
[22]	2018	Deep Transfer Learning for Person Re-identification	Image	no	Fine-tuning	yes	NA	Identification / security
[23]	2018	Deep Transfer Learning for Art Classification Problems	Image	no	Fine-tuning	yes	NA	Art
[24]	2018	Classification and unsupervised clustering of LIGO data with Deep Transfer Learning	Image	no	Fine-tuning	yes	NA	Physics / Astrophysics
[25]	2018	Empirical Study and Improvement on Deep Transfer Learning for Human Activity Recognition	Tabular/ bigdata	yes	Fine-tuning	yes	NA	Human Activity Recognition
[26]	2018	Automatic ICD-9 coding via deep transfer learning	Tabular/ bigdata	no	Fine-tuning	yes	NA	Medical
[27]	2017	Video-based emotion recognition in the wild using deep transfer learning and score fusion	video (audio & visual)	yes	Fine-tuning	yes	VGG-Face	Human science / psychology
[28]	2020	Automated invasive ductal carcinoma detection based using deep transfer learning with whole-slide images	Image	no	Freezing CNN layers	yes	Res-Net, Dense-Net	Medical image
[29]	2019	Deep Transfer Learning for Signal Detection in Ambient Backscatter Communications	Tabular/ bigdata	no	Freezing CNN layers	yes	NA	Tele-communication
[30]	2019	Brain tumor classification using deep CNN features via transfer learning	Image	no	Freezing CNN layers	yes	Google-Net	Medical image
[31]	2018	Comparison of Deep Transfer Learning Strategies for Digital Pathology	Image	no	Freezing CNN layers	yes	NA	Medical image
[32]	2018	Deep transfer learning for military object recognition under small training set condition	Image	no	Freezing CNN layers	yes	NA	Military
[33]	2018	Deep Transfer Learning for Image-Based Structural Damage Recognition	Image	no	Freezing CNN layers	yes	VGG-Net	Civil engineering
[34]	2017	Deep Transfer Learning for Modality Classification of Medical Images	Image	no	Freezing CNN layers	yes	VGG-Net, Res-Net	Medical image
[35]	2017	Folding Membrane Proteins by Deep Transfer Learning	Tabular/ bigdata	no	Freezing CNN layers	yes	Res-Net	Chemistry
[36]	2020	An Evaluation of Progressive Neural Networks for Transfer Learning in Natural Language Processing	NLP / text	no	Progressive learning	no	NA	NLP
[37]	2020	Progressive Transfer Learning and Adversarial Domain Adaptation for Cross-Domain Skin Disease Classification	Image	no	Progressive learning	yes	NA	Medical image

[38]	2017	Progressive Neural Networks for Transfer Learning in Emotion Recognition	Image & audio	yes	Progressive learning	no	NA	Paralinguistic
[39]	2020	A deep transfer learning model with classical data augmentation and CGAN to detect COVID-19 from chest CT radiography digital images	Image	no	Adversarial-based	yes	Alex-Net, VGG-Net16, VGG-Net19, Google-Net, Res-Net50	Medical image
[40]	2019	Diagnosing Rotating Machines with Weakly Supervised Data Using Deep Transfer Learning	Tabular/ bigdata	yes	Adversarial-based	yes	NA	Mechanic
[41]	2017	A New Deep Transfer Learning Based on Sparse Auto-Encoder for Fault Diagnosis	Tabular/ bigdata	yes	Sparse Auto-Encoder	no	NA	Mechanic

The most common DTL method is using a trained model on a highly related dataset to target data and fine-tune it on target data (fine-tuning). This method can also be considered a specific weight initialization technique. The simplicity of applying this technique makes it the most popular DTL method; 16 of 30 selected works have used this method. This method can improve training on target data in various ways, such as reducing training costs and tackling the need for an extensive target dataset. However, it is still prone to catastrophic forgetting. Needless to say, it is a very effective DTL method for many tasks and datasets in various fields such as medical, mechanics, art, physics, security, etc. Also, it has been applied for both image datasets and tabular (numerical) datasets as listed in Table 1.

The second popular approach in DTL is freezing CNN layers in a pre-trained model and fine-tune only lateral fully connected layers (Freezing CNN layers). CNN layers extract features from the given dataset, and the fully connected layers are responsible for classification, which in this method will be fine-tuned to the new task for target data. [28...35] are the sample research publications, which have used this method for different data types such as image and tabular data as listed in Table 1. This technique is specific to the models consisting of CNN layers; however, it can be extended to other deep learning models by assuming the earlier and middle layers are acting similar to CNN layers for feature extraction.

Using well-known models such as VGG-Net, Alex-Net, and Res-Net, which has already been trained on ImageNet datasets [42], is a general approach for both of the techniques mentioned above since they are easily accessible, and they are pre-trained to the highest possible accuracy. It is worth mentioning that such training can take days of processing time even with clusters of GPUs/TPUs and the mentioned methods are skipping the pre-training step by simply downloading a publicly available pre-trained model.

[36][37][38] are based on the progressive learning method, also known as progressive neural networks (PNNs), described earlier. [36] evaluates progressive learning effectiveness for common natural language processing (NLP) tasks: sequence labeling and text classification. Through evaluation and comparison of applying PNNs to various models, datasets, and tasks, they show how PNNs improve DL models' accuracy by avoiding catastrophic forgetting in fine-tuning techniques. [37][38] use PNNs for image and audio datasets and similarly finds tangible improvements in comparison to other DTL techniques.

[39] and [40] are examples of adversarial-based approaches that we found in the literature. In [39], they used conditional generative adversarial networks (CGAN) to expand limited target data of chest X-Ray images for detecting Covid-19 DTL model. [40] applies domain adversarial training to obtain the shared features between multiple source datasets.

Moreover, we found some tailored DTL methods for specific tasks and datasets like [41]. The proposed method in [41] as they describe is based on "three-layer sparse auto-encoder to extract the features of raw data, and applies the maximum mean discrepancy term to minimizing the discrepancy penalty between the features from training data and target data." They tailor that method for smart industry fault diagnosis problems and achieve 99.82% accuracy which is better than other approaches like deep belief network, sparse filter, deep learning, and support vector machine. Such tailored DTL approaches are not usually easy to generalize for different tasks or datasets. Nonetheless, they can open the door to interesting and new techniques in deep transfer learning's future.

Experimental Studies in Deep Transfer Learning

In this section we review two remarkable experimental evaluations of DTL techniques. The tests' setup, analysis, and conclusions are noteworthy.

"What is being transferred in transfer learning?" [43] is a recent experimental study which uses a series of tests on visual domain and deep learning models and tries to investigate what makes a successful transfer and which part of the network is responsible for that. To do so, they analyze networks in four different cases: (i) pre-trained network, (ii) random initialized network, (iii) fine-tuned network on target domain after pretraining on source domain, (iv) trained network from random initialization [43]. Moreover, to characterize the role of feature reuse, they use a source (pre-train) domain containing natural images (IMAGENET), and a few target (downstream) domains with decreasing visual similarities from natural images: DOMAINNET real, DOMAINNET clipart, CHEXPERT (medical chest X-Rays) and DOMAINNET quickdraw [43].

The study shows that feature reuse plays a key role in deep transfer learning as a pre-trained model on IMAGENET shows the largest performance improvement on real domain, which shares similar visual features (natural images) with IMAGENET in comparison to randomly initialized models. Also, they run a series of experiments by shuffling the image blocks (different block sizes). These experiments prove that feature reuse plays a very important role in transfer learning, particularly when the target domain shares visual features with the source domain. However, they realize that feature reuse is not the only reason for deep transfer learning success since even for distant targets such as CHEXPERT and quickdraw, they still observe performance boosts from deep transfer learning. Additionally, in all cases pre-trained models converge way faster than random initialized models. [43]

Further, they manually analyze common and uncommon mistakes in the training of randomly initialized versus pre-trained models. They observe that data samples marked incorrect in the pre-trained model and correct in the randomly initialized model are mostly ambiguous samples. On the other hand, the majority of the samples that a pre-trained model marked correct and a randomly initialized model marked incorrect are straightforward samples. This means that a pre-trained model has a stronger prior, and it is harder to adapt to the target domain. Moreover, using centered kernel alignment to measure feature similarities, they conclude that the initialization point drastically impacts feature similarity, and two networks with high accuracy can have a different feature space. Also, they discover similar results for distance in parameter space, which two random-initialized models are farther from each other compared to two pre-trained models. [43]

In regard to performance barriers and basins in the loss landscape, they have concluded that the network stays in the same basin of solution when fine-tuning a pre-trained network. They reach to this conclusion by training pre-trained models from two random runs as well as training random initialized models twice and comparing. Even when training a random initialized model two times with the same random values the models end up in different basins. [43]

Module criticality is an interesting analysis of deep learning models. Usually, in a deep CNN model each layer of CNN considers a module, while in some models a component of network can be considered as a module. To measure criticality of a module, it is possible to take a trained model and re-initialize each module at once and compare the amount of model accuracy drop. Adopting this technique, the authors of [4] discovered: (i) fully connected layers (near to model output) become critical for P-T model, and (ii) module criticality increases moving from the input side of model towards output, which is consistent with the concept of earlier layers (near input) extracting more general features while lateral layers have features that are more specialized for the target domain. [43]

[44] is another experimental analysis of transfer learning in visual tasks with the title of "Factors of Influence for Transfer Learning across Diverse Appearance Domains and Task Types". Three factors of influence are investigated in this study: (i) image domain, the difference in image domain between source and target tasks, (ii) task type, the difference in task type, and (iii) dataset size, the size of the source and target training sets. They perform over 1200 transfer learning experiments on 20 datasets spanning seven diverse image domains (consumer, driving, aerial, underwater, indoor, synthetic, closeups) and four task types (semantic segmentation, object detection, depth estimation, keypoint detection). [44]

They use data normalization (e.g., Illumination normalization) and augmentation techniques to improve models' accuracy. They adopt recent high-resolution backbone HRNetV2, which consists of 69M parameters. This backbone is easily adjustable for different datasets by simply replacing the head of the backbone. To make a fair comparison they pre-trained (to be used for transfer learning) their models from scratch and evaluated their performance using top-1 accuracy on the ILSVRC'12 validation set. [44]

The transfer learning experiments are mainly divided into two settings of (i) transfer learning with small target training set and (ii) with the full target set. The evaluation of transfer learning models is based on the gain obtained from fine-tuning from a specific source model compared to fine-tuning from ILSVRC'12 image classification with the main question of "are additional gains possible, by picking a good source?". Furthermore, they added a series of experiments for multi-source training to investigate the impact of using multi-source training for a specific task. [44]

Such an exhaustive experimental analysis resulted in following observations: (i) all experiments proved that transfer learning outperforms training from scratch (random initialization); (ii) for 85% of target tasks there exists a source task which tops ILSVCR'12 pre-training; (iii) the most transfer gain happens when the source and target tasks are in the same image domain (within-domain), which is even more important than source size; (iv) positive transfer gain is possible when the source image domain includes the target domain; (v) although multisource models bring good transfer, they are outperformed by the largest within-domain source; (vi) "for 65% of the targets within the same image domain as the source, cross-task-type transfer results in positive transfer gains"; (vii) as naturally expected, the larger datasets positively transfer towards the smaller datasets; (viii) transfer effects are stronger for a small target training set, which helps the process of choosing the transfer learning model by testing several models with a small section of target data. [44]

Conclusion

In this paper the taxonomy of transfer learning and deep transfer learning (DTL) has been reviewed. The dynamic nature of deep learning networks brought forth the possibility of adjusting pre-trained models for different tasks and datasets in various ways. Hence, the number of successful studies in this area is raising exponentially. Thirty articles out of more than a hundred recent publications in deep transfer learning has been listed and categorized based on their DTL approaches. The three sub-categories of network-based DTL approaches are the most widely used methods in the literature: (i) Fine-tuning, (ii) Freezing CNN Layers, and (iii) Progressive learning. These techniques have proven their ability and effectiveness for various machine learning problems. The simplicity of fine-tuning publicly available pre-trained models on extensive datasets is the reason for it being the most common transfer learning technique. For instance, several DTL techniques have been designed with high accuracy to detect an infected lung with Covid-19 despite being trained on very limited number of available labeled X-Rays since the start of pandemic. Last but not least, two interesting experimental studies and their discoveries on DTL has been reviewed in this paper.

All this is bringing DTL to the forefront of research in artificial intelligence and machine learning. Moreover, progressive learning, which is a very recent method in DTL, is developing rapidly and widely. This method can be considered a solid step towards artificial general intelligence since it is based on continual learning methodology.

References

[1] M. Iman, H. R. Arabnia, and R. M. Branchinst, "Pathways to Artificial General Intelligence: A Brief Overview of Developments and Ethical Issues via Artificial Intelligence, Machine Learning, Deep Learning, and Data Science," Springer, Cham, 2021, pp. 73–87.

[2] F. Zhuang *et al.*, "A Comprehensive Survey on Transfer Learning," *Proc. IEEE*, 2020.

[3] A. Farahani, S. Voghoei, K. Rasheed, and H. R. Arabnia, "A Brief Review of Domain Adaptation," Oct. 2020.

[4] S. Voghoei, N. Hashemi Tonekaboni, J. G. Wallace, and H. R. Arabnia, "Deep learning at the edge," *Proc. - 2018 Int. Conf. Comput. Sci. Comput. Intell. CSCI 2018*, pp. 895–901, Dec. 2018.

[5] H. S. Chang, M. C. Fu, J. Hu, and S. I. Marcus, "Google Deep Mind's AlphaGo," *OR/MS Today*, vol. 43, no. 5, pp. 24–29, 2016.

[6] N. Narayan Das, N. Kumar, M. Kaur, V. Kumar, and D. Singh, "Automated Deep Transfer Learning-Based Approach for Detection of COVID-19 Infection in Chest X-rays," *Irbm*, vol. 1, pp. 1–6, 2020.

[7] A. Jaiswal, N. Gianchandani, D. Singh, V. Kumar, and M. Kaur, "Classification of the COVID-19 infected patients using DenseNet201 based deep transfer learning," *J. Biomol. Struct. Dyn.*, pp. 1–8, Jul. 2020.

[8] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?," *Adv. Neural Inf. Process. Syst.*, vol. 4, no. January, pp. 3320–3328, Nov. 2014.

[9] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, "A survey on deep transfer learning," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11141 LNCS, pp. 270–279, 2018.

[10] A. A. Rusu *et al.*, "Progressive Neural Networks," 2016.

[11] J. Yosinski, J. Clune, A. Nguyen, T. Fuchs, and H. Lipson, "Understanding neural networks through deep visualization," *arXiv Prepr. arXiv1506.06579*, 2015.

[12] H. Ravishankar *et al.*, "Understanding the mechanisms of deep transfer learning for medical images," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2016, vol. 10008 LNCS, pp. 188–196.

[13] B. Kitchenham, O. Pearlbrereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, "Systematic literature reviews in software engineering – A systematic literature review," *Inf. Softw. Technol.*, vol. 51, no. 1, pp. 7–15, Jan. 2009.

[14] D. Jha *et al.*, "Enhancing materials property prediction by leveraging computational and experimental data using deep transfer learning," *Nat. Commun.*, vol. 10, no. 1, pp. 1–12, Dec. 2019.

[15] M. Talo, U. B. Baloglu, Ö. Yıldırım, and U. Rajendra Acharya, "Application of deep transfer learning for automated brain abnormality classification using MR images," *Cogn. Syst. Res.*, vol. 54, pp. 176–188, May 2019.

[16] Z. Wu, H. Jiang, K. Zhao, and X. Li, "An adaptive deep transfer learning method for bearing fault diagnosis," *Measurement*, vol. 151, p. 107227, Feb. 2020.

[17] W. Mao, L. Ding, S. Tian, and X. Liang, "Online detection for bearing incipient fault based on deep transfer learning," *Meas. J. Int. Meas. Confed.*, vol. 152, p. 107278, Feb. 2020.

[18] H. Phan *et al.*, “Towards more accurate automatic sleep staging via deep transfer learning,” *arXiv*, vol. XX, no. XX, pp. 1–12, 2019.

[19] P. Perera and V. M. Patel, “Deep Transfer Learning for Multiple Class Novelty Detection,” 2019.

[20] Y. Xu, Y. Sun, X. Liu, and Y. Zheng, “A Digital-Twin-Assisted Fault Diagnosis Using Deep Transfer Learning,” *IEEE Access*, vol. 7, pp. 19990–19999, 2019.

[21] K. Han, A. Vedaldi, and A. Zisserman, “Learning to Discover Novel Visual Categories via Deep Transfer Clustering,” 2019.

[22] M. Geng, Y. Wang, T. Xiang, and Y. Tian, “Deep Transfer Learning for Person Re-identification,” *2018 IEEE 4th Int. Conf. Multimed. Big Data, BigMM 2018*, Nov. 2016.

[23] M. Sabatelli, M. Kestemont, W. Daelemans, and P. Geurts, “Deep Transfer Learning for Art Classification Problems,” 2018.

[24] D. George, H. Shen, and E. A. Huerta, “Deep transfer learning: A new deep learning glitch classification method for advanced ligo,” *arXiv*, vol. 97, no. 10. arXiv, p. 101501, 22-Jun-2017.

[25] R. Ding *et al.*, “Empirical Study and Improvement on Deep Transfer Learning for Human Activity Recognition,” *Sensors*, vol. 19, no. 1, p. 57, Dec. 2018.

[26] M. Zeng, M. Li, Z. Fei, Y. Yu, Y. Pan, and J. Wang, “Automatic ICD-9 coding via deep transfer learning,” *Neurocomputing*, vol. 324, pp. 43–50, 2019.

[27] H. Kaya, F. Gürpınar, and A. A. Salah, “Video-based emotion recognition in the wild using deep transfer learning and score fusion,” *Image Vis. Comput.*, vol. 65, pp. 66–75, Sep. 2017.

[28] Y. Celik, M. Talo, O. Yildirim, M. Karabatak, and U. R. Acharya, “Automated invasive ductal carcinoma detection based using deep transfer learning with whole-slide images,” *Pattern Recognit. Lett.*, vol. 133, pp. 232–239, May 2020.

[29] C. Liu, Z. Wei, D. W. K. Ng, J. Yuan, and Y. C. Liang, “Deep Transfer Learning for Signal Detection in Ambient Backscatter Communications,” *IEEE Trans. Wirel. Commun.*, 2020.

[30] S. Deepak and P. M. Ameer, “Brain tumor classification using deep CNN features via transfer learning,” *Comput. Biol. Med.*, vol. 111, no. June, p. 103345, 2019.

[31] R. Mormont, P. Geurts, and R. Marée, “Comparison of deep transfer learning strategies for digital pathology,” 2018.

[32] Z. Yang *et al.*, “Deep transfer learning for military object recognition under small training set condition,” *Neural Comput. Appl.*, vol. 31, no. 10, pp. 6469–6478, Oct. 2019.

[33] Y. Gao and K. M. Mosalam, “Deep Transfer Learning for Image-Based Structural Damage Recognition,” *Comput. Civ. Infrastruct. Eng.*, vol. 33, no. 9, pp. 748–768, Sep. 2018.

[34] Y. Yu, H. Lin, J. Meng, X. Wei, H. Guo, and Z. Zhao, "Deep Transfer Learning for Modality Classification of Medical Images," *Information*, vol. 8, no. 3, p. 91, Jul. 2017.

[35] S. Wang, Z. Li, Y. Yu, and J. Xu, "Folding Membrane Proteins by Deep Transfer Learning," *Cell Syst.*, vol. 5, no. 3, pp. 202-211.e3, Sep. 2017.

[36] G. Hagerer *et al.*, "An Evaluation of Progressive Neural Networks for Transfer Learning in Natural Language Processing," 2020.

[37] Y. Gu, Z. Ge, C. P. Bonnington, and J. Zhou, "Progressive Transfer Learning and Adversarial Domain Adaptation for Cross-Domain Skin Disease Classification," *IEEE J. Biomed. Heal. Informatics*, vol. 24, no. 5, pp. 1379–1393, May 2020.

[38] J. Gideon, S. Khorram, Z. Aldeneh, D. Dimitriadis, and E. M. Provost, "Progressive Neural Networks for Transfer Learning in Emotion Recognition," *Proc. Annu. Conf. Int. Speech Commun. Assoc. INTERSPEECH*, vol. 2017-August, pp. 1098–1102, Jun. 2017.

[39] M. Loey, G. Manogaran, and N. E. M. Khalifa, "A deep transfer learning model with classical data augmentation and CGAN to detect COVID-19 from chest CT radiography digital images," *Neural Comput. Appl.*, pp. 1–13, Oct. 2020.

[40] X. Li, W. Zhang, Q. Ding, and X. Li, "Diagnosing Rotating Machines with Weakly Supervised Data Using Deep Transfer Learning," *IEEE Trans. Ind. Informatics*, vol. 16, no. 3, pp. 1688–1697, Mar. 2020.

[41] L. Wen, L. Gao, and X. Li, "A new deep transfer learning based on sparse auto-encoder for fault diagnosis," *IEEE Trans. Syst. Man, Cybern. Syst.*, vol. 49, no. 1, pp. 136–144, Jan. 2019.

[42] M. Simon, E. Rodner, and J. Denzler, "ImageNet pre-trained models with batch normalization," Dec. 2016.

[43] B. Neyshabur, H. Sedghi, and C. Zhang, "What is being transferred in transfer learning?," *Adv. Neural Inf. Process. Syst.*, vol. 2020-Decem, no. NeurIPS, 2020.

[44] T. Mensink, J. Uijlings, A. Kuznetsova, M. Gygli, and V. Ferrari, "Factors of Influence for Transfer Learning across Diverse Appearance Domains and Task Types," 2021.