Lifelong and continual learning - A survey

Naren Sivakumar
UMBC
1000 Hilltop Cir, Baltimore
narens1@umbc.edu

Abstract

While traditional machine learning techniques involve one single iteration of the collation, preparation and training of a model with data, this handicaps models by being a "no updates once deployed" approach. Using this approach not only severely limits the model's performance with respect to real-world data that might change continually, but also makes several assumptions about the widespread availability of data, that might not always be the case. Data also has the tendency to constantly evolve, with several slight changes over time being adding up to make a large dent in the performance of static models. With these limitations in mind, it is in our best interests to create "lifelong learners" - models that are both able to learn new tasks, as well as be good at the old ones. This survey paper aims to understand the basics behind the concept of lifelong and continual learning, as well as learn more about the leading ideas in the same field. Additionally, this paper also aims to understand the core concepts and ideology behind successful lifelong learners, and factors that must be taken into account when building one.

1. Introduction

Continual or lifelong learning refers to the ability of a model to learn new tasks that arise due to complex scenarios in (often real world) data, while not forgetting old tasks that it was initially trained to perform. While traditional models are trained on a single dataset, deployed, and are not updated unless they are retrained to adapt to a different task. While this has seen great success in machine-learning tasks, such as in [6, 9, 10, 12, 16, 39], it is necessary to take into account the rise of the requirement of models that are capable of learning on the job - namely, lifelong or continual learners. Continual learners must be able to effectively negate the requirement of retraining models to adapt to data that changes and evolves over time. They must also be able to learn from new data, while not forgetting old tasks. To formally enumerate the challenges that continual learners must address;

- **Adapt to new data or new tasks:** Continual learners must be able to adapt to new data that is obtained in the real world, which might differ from the style of data that the model has been trained on initially. The idea is to avoid retraining models from scratch, thereby wasting precious resources.
- **Avoid catastrophic forgetting:** Traditional models often "forget" the knowledge obtained from previous tasks, a key challenge that has to be taken into consideration while training continual learners. These models must be able to effectively learn new tasks, while not "forgetting" old knowledge, and constantly add to its knowledge bank.
- **Handle diverse environments:** The model must be robust and flexible when deployed in real-world environments, and must not need to be retrained or manually fixed each time the data varies slightly.

In this survey paper, we will be taking a deep dive into the definition of continual learning, the different types of continual learning and how they differ from each other, how to implement continual learning in real-world scenarios, and take in the research trends in this field over the last few years.

2. Continual Learning

Continual learning, an autological term, is a field that is continually evolving. For example, S Thrun in their work on Explanation-Based Neural Network Learning [37] explores the idea of teaching a model to learn the baseline theory of a domain, that includes certain invariant properties of a domain that will be true regardless of the sub-field of learning in the domain. They further go on to elaborate that this concept integrates two fundamental theories that have been studied extensively in the field of machine learning, namely analytical and inductive learning. Analytical learning, as described by Tom Mitchell et al [24], is the process of generalizing a concept, or domain given very few examples, often in the single digits. The paper further proposes an Explanation Based Generalization concept, where the key concept is that the model has some explanatory capabilities that allow it to generalize a single example into facts about a domain that can be later used to evaluate other examples that are from the same domain. This concept is then empirically proved by Steven Minton's work in the paper Quantitative results concerning the utility of explanation-based learning [22].

Inductive learning is the other cornerstone of such models, that allow it to generate learning patterns without any prior domain context, as shown by J. R. Quinlan in his work Induction of decision trees [27]. This paper breaks down the ID3 algorithm in detail, and shows that it is possible for systems to make inferences from data points while having no previous knowledge about the data. Put together, analytical and inductive learning make a formidable pairing that assist in the creation of continual learning systems, even when there are small incremental amounts of data that are irregularly provided to the system.

The definition of lifelong learning has varied over the years, as more and more people have dug deeper into the field and identified new problems, constraints or solutions. The earliest attempts at defining a continual learning model were made in 1994, by MB Ring in his thesis [29] that talked about continual learning in a reinforcement environment. Since then, many stalwarts of the field such as S Thrun and T Mitchell have refined the definition in their paper Lifelong Robot Learning [36], albeit in the field of robotics. They also talk about the advantages of continual learning, and how transfer learning makes it easier or robots learning similar tasks. While their experiments was with a small robot in tasks like unknown room navigation and seek and fetch, they proved the larger point that transfer learning and the usage of an existing knowledge base plays a vital role in improving performance. They also correctly argue that robots, like humans, may encounter several tasks in the same workspace, such as delivering mail, shipping things and so on in a warehouse.

Thrun further defines lifelong continual learning in his paper "Is learning the n-th thing any easier than learning the first?"[35]. If the model has been trained on a sequence of tasks from 1 to N, then it should use relevant information from its knowledge base, ie the information it has accrued from training on N tasks to help it better learn the N+1th task. Once it has achieved a passable score on the N+1th task, the knowledge gained from training on this task is then added to its knowledge base.

This overarching problem formulation can then be broken down into further detail, by classifying it into different, smaller problem formulations, which is done in the next section.

This concept, however, was transferred to machine learning models by Zhiyuan Chen and Bing Liu in their paper Lifelong Machine Learning[5]. In this paper, Chen and Liu discuss the implications of using lifelong learning in many regions of machine learning, such as lifelong unsupervised learning, shown in topic modeling by Pankaj Gupta et al in their paper Neural Topic Modelling with Continual Lifelong Learning [11]. Similarly, this concept is also used with semi-supervised learning in NELL [2]. New definitions have since then been formulated, such as the one crafted by Chen and Liu in their papers Lifelong Machine Learning and Lifelong Machine Learning for Topic Modeling and Beyond [4]. This newer definition is more in line with the tasks these lifelong learners encounter in the real world in the modern era. The definition assumes two things:

- **Once a task is learned, the data used to learn that task is no longer accessible;**
- **and that the new task, N+1, and its data are provided by the user.**

With these assumptions in mind, the goal of lifelong learners is to learn the new task, N+1, incrementally:

- **without catastrophically forgetting the old tasks, or resulting in degradation of the old information learned by the same model;**
- **and with knowledge transfer, or the use of existing knowledge gathered from previous tasks to learn the new task better.**

In this discussion, the two main factors that affect the performance of lifelong learners are catastrophic forgetting and knowledge transfer, and warrant a dive into existing literature to understand them a little better.

2.1. Catastrophic Forgetting

Catastrophic forgetting is a major adverse factor in the performance of lifelong learners. It is formally defined as the loss of performance in older tasks due to the loss or overwriting of previously learned information. In line with the important requirement of ensuring good performance on older tasks while effectively learning new tasks, several papers like Kirkpatrick's seminal work "Overcoming catastrophic forgetting in neural networks" [15], where he introduces the Elastic Weight Consolidation (EWC), a metric designed to selectively slow down learning on important weights while learning new tasks, and AA Rusu's paper "Progressive Neural Networks" Rusu et al. [30] that introduces progressive neural networks, a new style of neural network that draws from a pool of information that it has created, then adds to the pool once it finishes training. The paper claims that this style of network has immunity towards forgetting and further proves this point by testing it on playing several games.

2.2. Knowledge Transfer

On the other side, knowledge transfer is an invaluable tool that helps in making training these models easier than training them from scratch. Papers stressing on using transfer learning techniques in continual learning, like ELLA [31], a paper by P Ruvolo that makes use of a "basis" to train on new tasks, and continually refines this basis as it learns to do more and more. As the field of knowledge transfer gained steam, several in-depth survey papers, like Zhuang et al's "A comprehensive survey on transfer learning" [45] have been written, covering two types of knowledge transfer, broadly classified into problem categorization and solution categorization. Additionally, it can also be segregated by data interpretation, by doing so broadly into strategy and objective classes. There are further subdivisions in both classifications, but are out of the scope of this survey paper.

Continual learning can be further classified into different smaller problem formulations to enable better understanding and problem-solving. While there are several, like Task-Free Continual Learning, Class Incremental Learning, Blurred Boundary Continual Learning, and so on, we shall focus on:

- Class Incremental Learning
- Task Incremental Learning
- Domain Incremental Learning

The reason for choosing these three sub-fields is quite simply the fact that there is a lot more peer-reviewed research in these sub-domains than there is for other subdomains. This might be due to the fact that it is easier to source data to train on for CIL by simply splitting existing datasets like CIFAR-100 and MNIST into classes, and training models based on these [38]. While papers like Min-Yeong Park's "Versatile Incremental Learning: Towards Class and Domain-Agnostic Incremental Learning" [25] are trying to develop class and domain agnostic techniques in incremental learning by introducing intra and inter-class confusion, the predominant focus right now is on class incremental learning, where the new data is assumed to be obtained in distinctive classes.

3. Different Types of Incremental Learning

As discussed before, the three types of incremental learning we will focus on are class, task and domain incremental learning. These classes can be explained as varying the point of focus in between performing the same classification between different classes, completely different tasks, and tasks that may be similar in different domains.

3.1. Class Incremental Learning

In the real world, introducing novel classes with little to no data (as is the case when new plants are discovered, for example) often pose a large problem to traditional learners. In Class Incremental Learning (referred to as CIL for ease of use from here onwards), the introduction of new classes, in theory, does not disrupt anything – in fact, the new class and features are learned more easily than before due to the existing knowledge bank that the model already possesses from learning older classes. Zhou et al define CIL in their paper "Class-Incremental Learning: A Survey" [43] as an "aim to learn from an evolutive stream of information, without overlapping classes. The goal is to preserve knowledge and learn the new task simultaneously. Once the new task has been learned, the model is tested on all seen classes." They also go on to say that class overlap is a common occurrence, and this version of CIL is known as blurry class incremental learning [1].

Another important development in the CIL field was the use of exemplar sets during continual training. Exemplar sets are smaller samples of the data sets that the model has seen during previous training tasks, that are then used to "few-shot" the older instances of data while training the model on the new class, thereby making sure that the model doesn't forget what it has learned before. Exemplar sets have been a staple of creating good CIL models since their introduction in the iCaRL paper, written by S A Rebuffi et al [28]. The exemplar sets are managed by the model, with some samples of the new data getting added to the exemplar set once training is complete. These data samples are usually randomly sampled from the larger data set.

However, exemplar sets raise many ethical concerns, with some governmental legislation preventing the long-term storage of data in central locations, as stated in Marc Masana's survey on Class Incremental Learning [21]. The usage of exemplar sets also raises several memory concerns, since storing exemplar sets consumes a lot of memory, as shown in Zhou's paper [43]. With this in mind, there have been several successful approaches to solving class incremental learning without using exemplar sets, with some examples of approaches being splitting the learner into a feature extractor and a classifier, and training each separately on alignment loss and stored "prototypes", or the feature means of the old classes respectively, as done in the PASS paper written by Zhu et al [44]. Expanding on the prototype concept, G Petit et al [26] propose a method where they simply replace the old prototype with the nearest feature center that includes the new classes.

However, Libo Huang et al in their paper "Exemplar-Free Class Incremental Learning via Incremental Representation" [13] argue that these methods do not always lead to good results in practice, and instead propose a simple incremental representation, where a good initial feature representation coupled with the building and maintaining of near-perfect feature spaces is more than enough to alleviate forgetting in the learner while making the overall understanding of learning more comprehensible since pseudo features are not required. The feature space is constructed through dataset augmentation strategies and is maintained by a simple L2 loss classifier.

To sum up, class incremental learning can be achieved either by using exemplar sets, which are dominant in the field, but are more computationally expensive, or by using exemplar-free class incremental learning methods, which is a nascent field that is quickly gaining popularity because of its ease of understanding and reduced computational complexities.

3.2. Task Incremental Learning

Task incremental learning the process of learning new tasks, while ensuring the older tasks are not forgotten. The definition of task incremental learning is further elaborated in the paper "Three types of incremental learning", written by Gido M et al [38]. In this definition, van de Ven goes on to say that each learner must learn distinguishable tasks, and it is always clear to the algorithm which task has to be executed at what time. He further writes that since there are different methods of training for each task, varying from using a different output layer to using a different neural network, the main objective is not only to prevent catastrophic forgetting but also to promote the sharing of learned information between tasks to optimize the trade-off between performance and computational complexity. This could include positive forward or even backward information transfer between tasks.

For example, one paper that focuses on *minimizing* negative back transfer is the paper by David Lopez-Paz, "Gradient Episodic Memory for Continual Learning" [20]. This model favors an episodic memory, which both minimizes forgetting and promotes positive back transfer, ie using the new task to improve performance on older tasks. They also stress the usage of rich task descriptors, which can be used to create zero-shot learners.

The paper "Multi-Task Incremental Learning for Object Detection" by Xialei Liu et al [18] takes task incremental learning a step forward into the realm of object detection. The authors correctly identify that most object detectors are static, and those that aren't are incrementally taught on a single domain, as shown in [19, 34]. They show that continually learning object detection can be broken down into four categories, three of which are non-trivial, spanning all possible combinations of domains and classes being changed.

According to Van de Ven and S Farquhar[7, 38], however, the community should attempt to move away from the notion that context identities will always be provided at test time, which sets the bar for task incremental learning too low. Indeed, in the experimental results of their paper, they have proven that while all the methods (CIL, TIL and DIL) perform better than single shared neural networks that are sequentially trained on different contexts, they are outperformed by single separate neural network being trained for each context.

3.3. Domain Incremental Learning

According to Haizhou Shi [33], Domain incremental learning aims to adapt to a sequence of domains with access to only a small subset of data from previous domains. While previous work in the domain incremental learning field like [8, 14, 23, 40] uses different methods suited to their particular tasks, this is a domain that does not have one standard way of approaching the task [33]. While this is good to inspire creativity, it often leads to confusing approaches. This is why Shi proposes a unified DIL technique, one which uses three different bounds to create a single adaptive error bound, that performs equivalent to the state of the art models in the same space, such as [3, 17, 32, 42]. This goes to show that while publications in the field thus far adopt their own methods, a standard method can be introduced and used as a benchmark standard, or even to further the domain while improving upon the adaptive error bound technique.

On the other hand, there are published works trying to extend incremental learning to multiple domains, such as X Wei et al [41] who uses a multi-level feature alignment to ensure domain alignment can be reached at various levels, and a progressive training strategy is adapted to gradually introduce the model to multiple domains, rather than all at once.

4. Conclusion

This short review, in conclusion, says that continual or incremental learning is a field of research that has an ever-growing impact as our world leans more and more towards a one-size-fits-all model that learns continuously with a large data stream to support it. It also results in benefits when applied correctly to models that are implemented in the real world in terms of understanding and computational complexity. This field has been theorized since 1994 [29] and worked on with several different approaches coming to the forefront. However, The field is unfortunately still heavily reliant on existing datasets to train continual learners, which leads to an unhealthy bias toward class incremental learning. Additionally, while there are methods being developed in Domain Incremental Learning, they are quite disjoint, and do not build up on prior work. There have been groundbreaking inventions in the field, such as with iCaRL [28], Elastic Weighted Consideration [15] and GEM[20] and related works. Since it is such a nascent field, there is definitely more to come, with newer entrants focusing towards classless and domainless incremental learning [25], and exemplar free class incremental learning [13, 26].

References

[1] Jihwan Bang, Heesu Kim, YoungJoon Yoo, Jung-Woo Ha, and Jonghyun Choi. Rainbow memory: Continual learning with a memory of diverse samples. In *Proceedings of*

the IEEE/CVF conference on computer vision and pattern recognition, pages 8218–8227, 2021. 3

[2] Andrew Carlson, Justin Betteridge, Bryan Kisiel, Burr Settles, Estevam Hruschka, and Tom Mitchell. Toward an architecture for never-ending language learning. In *Proceedings of the AAAI conference on artificial intelligence*, pages 1306–1313, 2010. 2

[3] Arslan Chaudhry, Marc'Aurelio Ranzato, Marcus Rohrbach, and Mohamed Elhoseiny. Efficient lifelong learning with a-gem. *arXiv preprint arXiv:1812.00420*, 2018. 4

[4] Zhiyuan Chen. Lifelong machine learning for topic modeling and beyond. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Student Research Workshop*, pages 133–139, 2015. 2

[5] Zhiyuan Chen and Bing Liu. *Lifelong machine learning*. Springer Nature, 2022. 2

[6] Jacob Devlin. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018. 1

[7] Sebastian Farquhar and Yarin Gal. Towards robust evaluations of continual learning, 2019. 4

[8] Prachi Garg, Rohit Saluja, Vineeth N Balasubramanian, Chetan Arora, Anbumani Subramanian, and CV Jawahar. Multi-domain incremental learning for semantic segmentation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 761–771, 2022. 4

[9] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020. 1

[10] Alex Graves and Alex Graves. Long short-term memory. *Supervised sequence labelling with recurrent neural networks*, pages 37–45, 2012. 1

[11] Pankaj Gupta, Yatin Chaudhary, Thomas Runkler, and Hinrich Schuetze. Neural topic modeling with continual lifelong learning. In *Proceedings of the 37th International Conference on Machine Learning*, pages 3907–3917. PMLR, 2020. 2

[12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015. 1

[13] Libo Huang, Zhulin An, Yan Zeng, Chuanguang Yang, Xinqiang Yu, and Yongjun Xu. Exemplar-free class incremental learning via incremental representation. *arXiv preprint arXiv:2403.16221*, 2024. 3, 4

[14] Tobias Kalb, Masoud Roschani, Miriam Ruf, and Jürgen Beyerer. Continual learning for class-and domain-incremental semantic segmentation. In *2021 IEEE Intelligent Vehicles Symposium (IV)*, pages 1345–1351. IEEE, 2021. 4

[15] James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, et al. Overcoming catastrophic forgetting in neural networks. *Proceedings of the national academy of sciences*, 114(13):3521–3526, 2017. 2, 4

[16] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012. 1

[17] Zhizhong Li and Derek Hoiem. Learning without forgetting. *IEEE transactions on pattern analysis and machine intelligence*, 40(12):2935–2947, 2017. 4

[18] Xialei Liu, Hao Yang, Avinash Ravichandran, Rahul Bhotika, and Stefano Soatto. Continual universal object detection. *CoRR*, abs/2002.05347, 2020. 4

[19] Yaoyao Liu, Bernt Schiele, Andrea Vedaldi, and Christian Rupprecht. Continual detection transformer for incremental object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 23799–23808, 2023. 4

[20] David Lopez-Paz and Marc' Aurelio Ranzato. Gradient episodic memory for continual learning. In *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2017. 4

[21] Marc Masana, Xialei Liu, Bartłomiej Twardowski, Mikel Menta, Andrew D. Bagdanov, and Joost van de Weijer. Class-incremental learning: Survey and performance evaluation on image classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(5):5513–5533, 2023. 3

[22] Steven Minton. Quantitative results concerning the utility of explanation-based learning. *Artificial Intelligence*, 42(2-3): 363–391, 1990. 2

[23] M. Jehanzeb Mirza, Marc Masana, Horst Possegger, and Horst Bischof. An efficient domain-incremental learning approach to drive in all weather conditions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 3001–3011, 2022. 4

[24] Tom M Mitchell, Richard M Keller, and Smadar T Kedar-Cabelli. Explanation-based generalization: A unifying view. *Machine learning*, 1:47–80, 1986. 1

[25] Min-Yeong Park, Jae-Ho Lee, and Gyeong-Moon Park. Versatile incremental learning: Towards class and domain-agnostic incremental learning, 2024. 3, 4

[26] Grégoire Petit, Adrian Popescu, Hugo Schindler, David Picard, and Bertrand Delezoide. Fetril: Feature translation for exemplar-free class-incremental learning. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 3911–3920, 2023. 3, 4

[27] J. Ross Quinlan. Induction of decision trees. *Machine learning*, 1:81–106, 1986. 2

[28] Sylvestre-Alvise Rebuffi, Alexander Kolesnikov, Georg Sperl, and Christoph H Lampert. icarl: Incremental classifier and representation learning. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 2001–2010, 2017. 3, 4

[29] Mark Bishop Ring. *Continual learning in reinforcement environments*. The University of Texas at Austin, 1994. 2, 4

[30] Andrei A Rusu, Neil C Rabinowitz, Guillaume Desjardins, Hubert Soyer, James Kirkpatrick, Koray Kavukcuoglu, Razvan Pascanu, and Raia Hadsell. Progressive neural networks. *arXiv preprint arXiv:1606.04671*, 2016. 2

[31] Paul Ruvolo and Eric Eaton. Ella: An efficient lifelong learning algorithm. In *International conference on machine learning*, pages 507–515. PMLR, 2013. 2

[32] Jonathan Schwarz, Wojciech Czarnecki, Jelena Luketina, Agnieszka Grabska-Barwinska, Yee Whye Teh, Razvan Pascanu, and Raia Hadsell. Progress & compress: A scalable framework for continual learning. In *International conference on machine learning*, pages 4528–4537. PMLR, 2018. 4

[33] Haizhou Shi and Hao Wang. A unified approach to domain incremental learning with memory: Theory and algorithm, 2023. 4

[34] Jeng-Lun Shieh, Qazi Mazhar ul Haq, Muhamad Amirul Haq, Said Karam, Peter Chondro, De-Qin Gao, and Shanq-Jang Ruan. Continual learning strategy in one-stage object detection framework based on experience replay for autonomous driving vehicle. *Sensors*, 20(23), 2020. 4

[35] Sebastian Thrun. Is learning the n-th thing any easier than learning the first? *Advances in neural information processing systems*, 8, 1995. 2

[36] Sebastian Thrun and Tom M Mitchell. Lifelong robot learning. *Robotics and autonomous systems*, 15(1-2):25–46, 1995. 2

[37] Sebastian Thrun and Sebastian Thrun. *Explanation-based neural network learning*. Springer, 1996. 1

[38] Gido M Van de Ven, Tinne Tuytelaars, and Andreas S Tolias. Three types of incremental learning. *Nature Machine Intelligence*, 4(12):1185–1197, 2022. 3, 4

[39] A Vaswani. Attention is all you need. *Advances in Neural Information Processing Systems*, 2017. 1

[40] Yabin Wang, Zhiwu Huang, and Xiaopeng Hong. S-prompts learning with pre-trained transformers: An occam's razor for domain incremental learning. *Advances in Neural Information Processing Systems*, 35:5682–5695, 2022. 4

[41] Xing Wei, Shaofan Liu, Yaoci Xiang, Zhangling Duan, Chong Zhao, and Yang Lu. Incremental learning based multi-domain adaptation for object detection. *Knowledge-Based Systems*, 210:106420, 2020. 4

[42] Friedemann Zenke, Ben Poole, and Surya Ganguli. Continual learning through synaptic intelligence. In *International conference on machine learning*, pages 3987–3995. PMLR, 2017. 4

[43] Da-Wei Zhou, Qi-Wei Wang, Zhi-Hong Qi, Han-Jia Ye, De-Chuan Zhan, and Ziwei Liu. Class-incremental learning: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, page 1–20, 2024. 3

[44] Fei Zhu, Xu-Yao Zhang, Chuang Wang, Fei Yin, and Cheng-Lin Liu. Prototype augmentation and self-supervision for incremental learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5871–5880, 2021. 3

[45] Fuzhen Zhuang, Zhiyuan Qi, Keyu Duan, Dongbo Xi, Yongchun Zhu, Hengshu Zhu, Hui Xiong, and Qing He. A comprehensive survey on transfer learning. *Proceedings of the IEEE*, 109(1):43–76, 2021. 3