Volume 11, Issue 3, March – 2026 International Journal of Innovative Science and Research Technology
ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26mar843

A Survey on Advancements in Transfer and Continual Learning: Insights for Modern Computer Vision

Hemanth Sai Kosari[1]; Deeksha Akkati[2]

[1,2]Department of Computer Engineering, San Jose State University San Jose, CA, USA

Publication Date: 2026/03/18

Abstract: Transfer learning and continual learning are pivotal methodologies in current artificial intelligence, offering solutions to enhance computer vision systems. Transfer learning utilizes pretrained models to perform specific tasks efficiently with limited data, while continual learning allows systems to learn new tasks incrementally without forgetting prior knowledge. Vision Transformers (ViTs), leveraging attention mechanisms, have significantly advanced feature representation and task performance in image classification and object detection, outperforming traditional convolutional networks. Despite these advancements, challenges like domain adaptation and catastrophic forgetting remain critical to solve. This paper reviews techniques including fine-tuning, Elastic Weight Consolidation (EWC), and selfsupervised learning, highlighting their applications in fields such as autonomous driving and medical imaging that are closely related to computer vision. It identifies research gaps and provides insights into creating scalable and robust computer vision solutions.

How to Cite: Hemanth Sai Kosari; Deeksha Akkati (2026) A Survey on Advancements in Transfer and Continual Learning: Insights for Modern Computer Vision. *International Journal of Innovative Science and Research Technology*, 11(3), 1286-1291. https://doi.org/10.38124/ijisrt/26mar843

I. INTRODUCTION

Over the last decade, significant developments have been made in the fields of transfer learning and continual learning, as both of them are essential for enhancing computer vision applications. Transfer learning enables models to utilize knowledge acquired from large-scale pretrained networks and datasets to solve specific tasks even with limited datasets, while continual learning equips models to acquire knowledge on new tasks progressively without compromising their existing knowledge. These methodologies are relevant for addressing the complexities of dynamic real-world applications such as autonomous driving, medical imaging, and industrial inspection. This survey aims to sum up the latest advancements in both transfer and continual learning, their integration, and their transformative impact on the territory of computer vision.

One of the most pressing challenges is domain adaptation, where models must generalize effectively from a source domain to a target domain with different characteristics. Although transfer learning has improved this issue through strategies such as fine-tuning and feature reuse, domain shifts continue to pose significant challenges, particularly in specialized areas like medical imaging or remote sensing [1]. Continual learning on the contrary, faces the problem of catastrophic forgetting, where models may lose the knowledge to perform previously learned tasks when learning new knowledge. This becomes critical in rapidly evolving environments, like robotics, where systems must continuously adapt to new tasks while remembering their previous learning [2].

The researchers have introduced several innovative solutions to address these challenges. In the space of transfer learning, techniques such as domain adversarial neural networks (DANNs) and self-supervised learning frameworks, including SimCLR and MoCo, have shown great potential in enhancing domain generalization [3]. In continual learning, methods like Elastic Weight Consolidation (EWC) and experience replay work to reduce forgetting by safeguarding the critical parameters and revisiting prior tasks during training [4]. Notably, the hybridization of these two approaches is gaining traction, as it joins the strengths of both transfer and continual learning to develop more robust and adaptive computer vision models.

This survey will deliver a thorough review of recent advancements, emphasizing the application of these methodologies across various domains, such as autonomous driving, medical image analysis, remote sensing, and surveillance systems. By analyzing the latest research from the last decade, this survey aims to find existing gaps and emerging trends in the integration of transfer and continual learning. It will also address challenges like computational efficiency and data scarcity, which must be cleared to enhance the applicability of these techniques in real-world scenarios [5].

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26mar843

In the research phase, a broad literature review will be undertaken, focusing on publications from top conferences and journals over the past ten years. Key frameworks and algorithms will be categorized, and compared regarding their performance across diverse tasks and domains. The ultimate goal is to offer insights into the future of transfer and continual learning, spotlighting potential breakthroughs that could facilitate the development of more efficient computer vision systems.

II. LITERATURE REVIEW

Transfer learning has significantly increased the adaptability and efficiency of machine learning and deep learning models by leveraging knowledge from pre-trained models. Zhuang et al. [1] highlight the core methodologies of transfer learning, including fine-tuning and feature extraction, which have enabled models to perform well on target tasks with limited data. These methods are particularly impactful in fields like medical imaging, where labelled data is scarce, and models must generalize across diverse datasets. Despite its advantages, challenges such as domain shifts and overfitting remain. Ganin et al. [5] propose Domain-Adversarial Neural Networks (DANNs) to address domain shifts by aligning feature distributions across source and target domains, thereby enhancing generalization. This method has shown significant success in applications such as autonomous driving and industrial inspection, where domain adaptation is critical.

Continual learning aims to address the limitations of traditional static training approaches by enabling models to learn sequential tasks without forgetting prior knowledge. Wang et al. [2] explore various methodologies to mitigate catastrophic forgetting, a key challenge in this paradigm. Elastic Weight Consolidation (EWC), introduced by Kirkpatrick et al. [4], penalizes changes to important parameters identified during earlier tasks, preserving critical information. Fayek et al. [7] extend this approach through progressive learning, which dynamically adjusts the model's capacity to accommodate new tasks while retaining past knowledge. These methods often involve trade-offs between computational scalability and efficiency, particularly in highscale applications like robotics, embedded, and autonomous systems.

Recent advancements in Vision Transformers (ViTs) have restructured computer vision techniques entirely by utilizing attention mechanisms to model global dependencies in image data. Unlike traditional convolutional neural networks (CNNs), which rely on localized filters, ViTs process images as sequences of patches, capturing contextual relationships more effectively. Parvaiz et al. [8] demonstrate that ViTs outperform CNNs in tasks like image classification and object detection, particularly in domains requiring high interpretability, such as medical imaging. Furthermore, Zhou et al. [10] investigate adaptive ViT architectures that optimize parameter utilization for specific tasks, enhancing performance without significant computational overhead. While ViTs have established a new benchmark in computer vision, their reliance on large datasets for pretraining remains a notable limitation, prompting research into more efficient training techniques.

➢ *Overview*

- *Importance of Transfer and Continual Learning*

Transfer Learning and Continual Learning have become crucial paradigms in AI and ML due to their ability to solve problems with limited training data and resources. Transfer Learning enables a model trained on one task to be used again for another related task, saving computational resources and most importantly time. For instance, models trained on large datasets like ImageNet can be fine-tuned for specific tasks like medical image classification, helping healthcare practitioners to use high-quality models despite having fewer labelled images for their domain. Continual Learning, on the other hand, allows models to increase their knowledge over time without forgetting previously learned tasks. This is particularly useful in robotics, where an autonomous robot can learn new skills or adapt to new environments without forgetting prior tasks, such as navigation or object recognition. These paradigms are critical in fields like autonomous vehicles, where continuous learning of new road conditions, weather, and driving scenarios is necessary for safe operations. Both methods push AI toward more flexible, adaptable systems capable of lifelong learning, essential for creating intelligent systems that can evolve with changing environments.

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26mar843

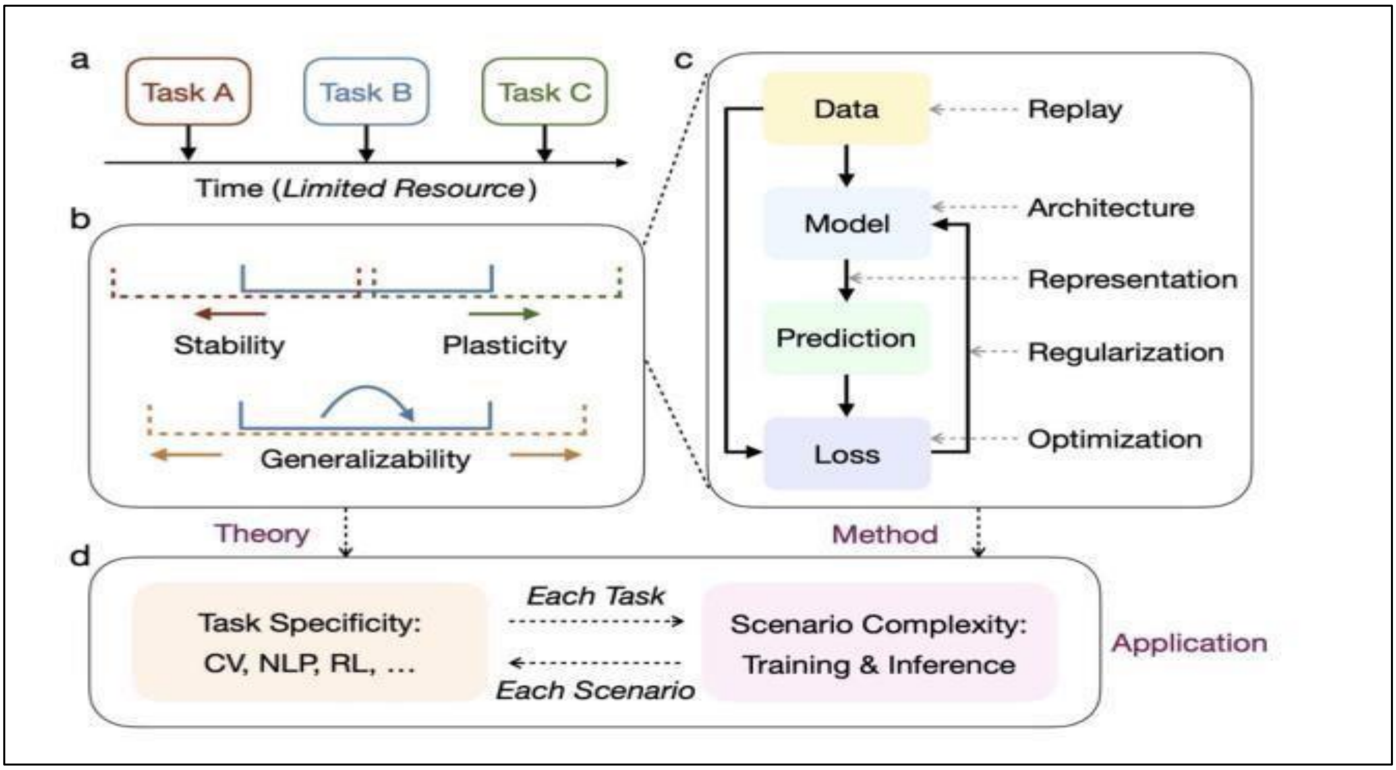

Fig 1 Architecture of Continual Learning

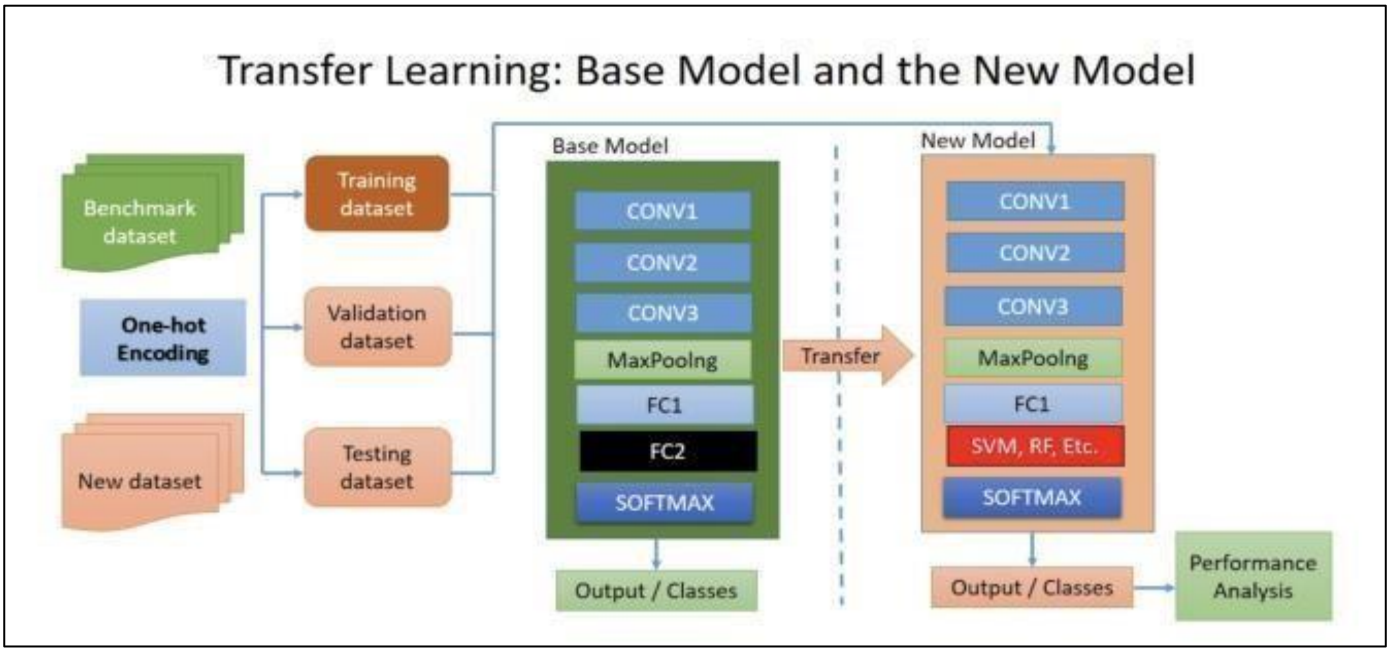

Fig 2 Architecture of Transfer Learning

- *Catastrophic Forgetting*

Catastrophic forgetting refers to the phenomenon where a neural network, forgets the earlier tasks after learning to do a newer task. This issue comes when a model is trained continuously on various datasets or tasks, causing the network to overwrite the knowledge gained from earlier tasks. As a result, the model's performance on previously learned tasks decreases or may fully forget it. This problem especially occurs in continual learning scenarios, where the model must adapt to new information without losing proficiency on prior knowledge, making it a significant challenge in developing lifelong learning systems.

✓ *Elastic Weight Consolidation (EWC)*:

Elastic Weight Consolidation is a popular technique used to overcome catastrophic forgetting phenomenon in continual learning. The main idea behind Elastic Weight Consolidation is to penalize large changes in weights that are essential for previously learned tasks while allowing flexibility for new learning. This is achieved by introducing a regularisation term based on the Fisher information matrix,

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26mar843

which estimates the importance of each parameter. When the model learns a new task, EWC reduces the risk of catastrophic forgetting by protecting critical parameters while allowing others to adapt the new task. For example, in reinforcement learning, EWC helps ensure that the agent does not forget previously learned actions while learning new strategies. This method has been successful in various domains, such as robotics and natural language processing, where models often need to continually learn without losing their ability to perform previous tasks. It requires computation of the Fisher matrix, which can be computationally expensive for large networks.

✓ *Quantum Machine Learning*:

Quantum Machine Learning (QML) is an emerging field that may provide solutions to catastrophic forgetting by utilizing quantum computing concepts to enhance adaptation of models. Quantum computers' ability to perform complex computations using qubits allows for more efficient storing and retrieval of knowledge, potentially overcoming the issue of catastrophic forgetting. QML could enable neural networks to retain previous knowledge while incorporating new data by exploiting quantum entanglement, superposition, and interference. Quantum algorithms, such as quantum annealing, may optimize the learning process, helping neural networks adapt to new tasks without forgetting previously learned tasks. While still in the early stages, researchers have proposed that quantum-enhanced algorithms could better preserve the knowledge retained from earlier tasks, especially in dynamic and evolving environments. The use of QML for continual learning presents exciting possibilities, but practical implementations remain limited by current hardware capabilities and the need for further theoretical development.

- *Self-Supervised Learning in Transfer Learning*

A strategy that bridges the gaps between supervised and unsupervised learning, especially regarding Transfer Learning, is self-supervised learning. A model learns to produce significant features from unlabelled input in selfsupervised learning. Self-supervised learning allows models to learn from unlabelled datasets and then refine them on smaller datasets when used in Transfer Learning. Better feature representations that translate well to a range of tasks are provided by this method, which increases the adaptability of Transfer Learning. Through self-supervised pretraining on big datasets, computer vision models can identify valuable patterns and structures that can then be optimized for uses, such as autonomous driving or facial recognition. Selfsupervised learning increases Transfer Learning's scalability to domains with less reliance on labelled data.

- *Domain Adversarial training on Neural Networks*

Domain-adversarial training (DAT) is a technique used to improve the generalization ability of neural networks when transferring knowledge across various domains with different data distributions. In many Transfer Learning scenarios, models trained on a single dataset may not perform well when applied to a new, unseen domain due to domain shift — a difference between the source and target domains. Domain adversarial training addresses this by training a neural network to learn domain-invariant features that are useful for both source and target domains. This is done by introducing a domain discriminator that attempts to classify whether a feature comes from the source or the target domain, while the feature extractor is trained to fool this discriminator, forcing the model to learn features that are agnostic to the domain. For example, in computer vision, a model trained on synthetic images can be adapted to work on real-world images, overcoming the challenge of domain shift. DAT has been applied successfully in fields like medical imaging, where models trained on one type of medical data can be transferred to other datasets, such as those from different hospitals or geographical regions, without significant performance loss.

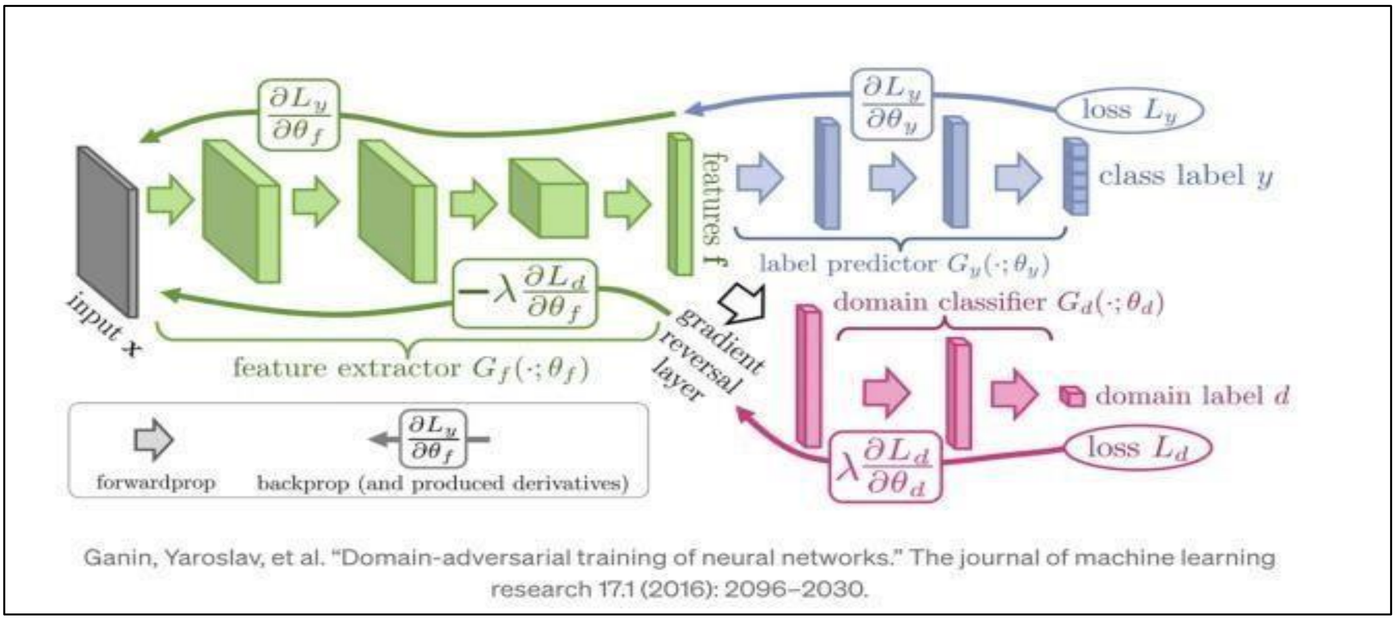

Fig 3 Basic Architecture of DANN

Volume 11, Issue 3, March – 2026 International Journal of Innovative Science and Research Technology
ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26mar843

- *Transformers, ViTs and Attention Mechanism*

By allowing models to concentrate on the most crucial portions of input sequences, transformers have altered deep learning and overcome the drawbacks of previous designs such as CNNs and RNNs. A key part of transformers, the attention mechanism enables models to efficiently process long-range relationships by weighing the significance of various input pieces. Tasks including text generation, machine translation, and natural language processing have greatly improved because of this design. Transformers are very useful in a variety of applications because of their versatility and capacity to handle sequential data without depending on recurring connections.

By segmenting images into patches and interpreting these patches as sequences, Vision Transformers (ViTs) apply the transformer architecture to computer vision. ViTs outperform conventional convolutional neural networks (CNNs) in a variety of tasks, including object detection and picture classification, by utilizing the attention mechanism to capture spatial linkages and global dependencies. ViTs are particularly useful for applications requiring a comprehensive knowledge of the input since they examine the entire visual context, unlike CNNs, which depend on localized filters. Because of this feature, ViTs are a well-liked option for computer vision applications such as picture labeling and segmentation.

Both Transformers and Vision Transformers rely on an attention system that enables models to constantly focus on the most crucial portions of the input while ignoring less significant portions. The model's capacity to learn long-range dependencies and linkages is improved by this dynamic weighting of input data, which is essential for tasks like multimodal learning and visual question answering (VQA). In applications that demand context-aware processing, attention methods have proven very helpful, allowing models to perform exceptionally well in tasks like picture production and video comprehension.

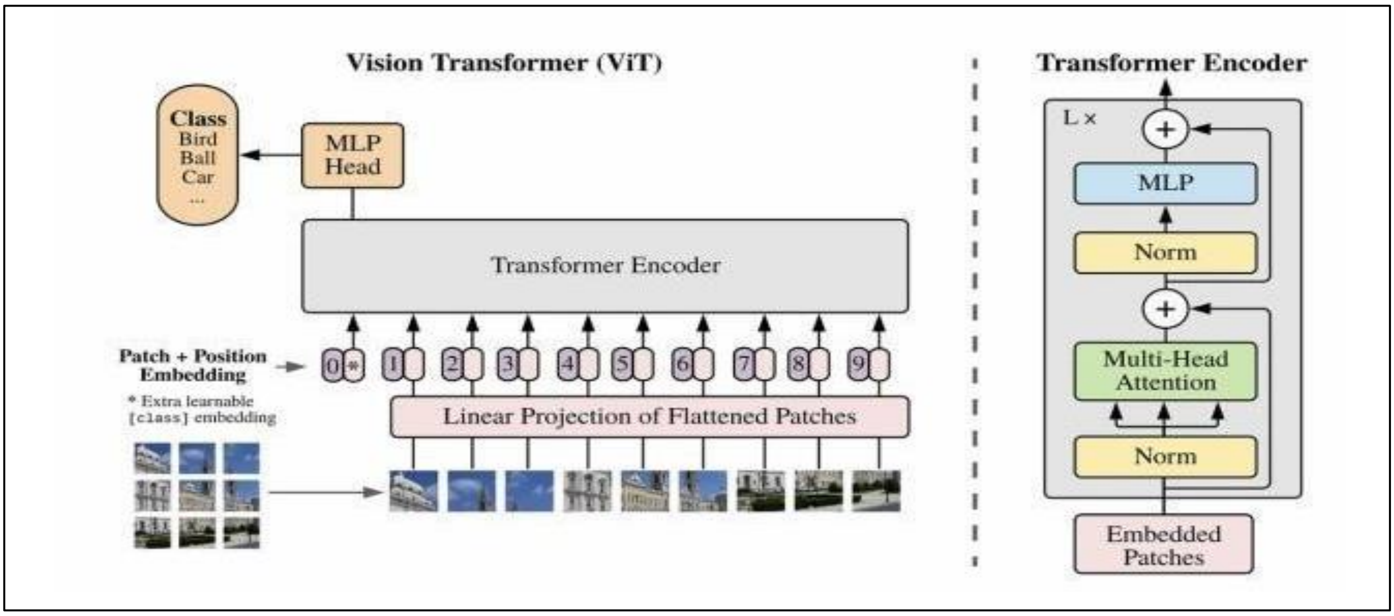

Fig 4 Architecture of a Transformer
Source: https://medium.com/@jduodu5/attention-mechanism-and-vision-transformer-8955f93c53d2

III. CONCLUSION

In order to overcome obstacles like sparse data and changing task settings, transfer learning and continuous learning have become essential approaches in the development of computer vision systems. The important contributions of cutting-edge methods such as Vision Transformers (ViTs), domain-adversarial training, selfsupervised learning, and Elastic Weight Consolidation (EWC) have been examined in this study. In a variety of applications, such as autonomous driving, medical imaging, and robotics, these approaches have shown impressive advancements in addressing issues like domain adaptation, catastrophic forgetting, and feature representation.

However, challenges persist, particularly in achieving computational efficiency, addressing domain shifts, and balancing scalability with performance. The integration of hybrid approaches, such as combining transfer and continual learning paradigms, offers a promising direction for developing robust, adaptive systems capable of lifelong learning. Future research should focus on optimizing resource utilization, exploring quantum machine learning for continual learning, and improving the interpretability and fairness of AI systems.

This survey highlights the revolutionary potential of transfer and continuous learning in influencing the direction of computer vision by combining current developments and pointing out gaps. In addition to improving the methodologies' applicability, closing these gaps will open the door for the development of intelligent systems that can handle.

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26mar843

REFERENCES

[1]. F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, "A Comprehensive Survey on Transfer Learning," arXiv preprint arXiv:1911.02685, 2019. doi: 10.48550/arXiv.1911.02685.

[2]. L. Wang, X. Zhang, H. Su, and J. Zhu, "A Comprehensive Survey of Continual Learning: Theory, Method and Application," arXiv preprint arXiv:2302.00487, 2023. doi: 10.48550/arXiv.2302.00487.

[3]. J. Gui, T. Chen, J. Zhang, Q. Cao, Z. Sun, H. Luo, and D. Tao, "A Survey on Self-supervised Learning: Algorithms, Applications, and Future Trends," arXiv preprint arXiv:2301.05712, 2023. doi:10.48550/arXiv.2301.05712.

[4]. J. Kirkpatrick, R. Pascanu, N. Rabinowitz, J. Veness, G. Desjardins, A. A. Rusu, K. Milan, J. Quan, T. Ramalho, A. Grabska-Barwinska, D. Hassabis, C. Clopath, D. Kumaran, and R. Hadsell, "Overcoming catastrophic forgetting in neural networks," arXiv preprint arXiv:1612.00796, 2016. doi:10.48550/arXiv.1612.00796.

[5]. Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, "Domain-Adversarial Training of Neural Networks," Journal of Machine Learning Research, vol. 17, pp. 135, 2016. doi: 10.48550/arXiv.1505.07818.

[6]. A. Hosna, E. Merry, and J. Gyalmo, "Transfer learning: a friendly introduction," Journal of Big Data, vol. 9, no. 102, 2022. doi: 10.1186/s40537022-00652-w.

[7]. H. M. Fayek, L. Cavedon, and H. R. Wu, "Progressive learning: A deep learning framework for continual learning," Neural Networks, vol. 128, pp. 345-357, 2020. doi: 10.1016/j.neunet.2020.05.011.

[8]. A. Parvaiz, M. A. Khalid, R. Zafar, H. Ameer, M. Ali, and M. M. Fraz, "Vision Transformers in medical computer vision—A contemplative retrospection," Engineering Applications of Artificial Intelligence, vol. 120, 2023. doi: 10.1016/j.engappai.2023.106126.

[9]. J. Kang, R. Fernandez-Beltran, P. Duan, S. Liu, and A. J. Plaza, "Deep Unsupervised Embedding for Remotely Sensed Images Based on Spatially Augmented Momentum Contrast," IEEE Transactions on Geoscience and Remote Sensing, vol. 59, no. 3, pp. 2598-2610, Mar. 2021. doi: 10.1109/TGRS.2020.3007029.

[10]. Y. Zhou, X. Tian, C. Zhang, Y. Zhao, and T. Li, "Elastic weight consolidation-based adaptive neural networks for dynamic building energy load prediction modeling," Energy and Buildings, vol. 265, p. 112098, 2022. doi: 10.1016/j.enbuild.2022.112098.

[11]. D. Kudithipudi et al., "Biological underpinnings for lifelong learning machines," Nat. Mach. Intell., vol. 4, pp. 196–210, 2022.

[12]. G. I. Parisi et al., "Continual lifelong learning with neural networks: A review," Neural Netw., vol. 113, pp. 54–71, 2019.

[13]. R. Hadsel et al., "Embracing change: Continual learning in deep neural networks," Trends Cogn. Sci., vol. 24, pp. 1028–1040, 2020.

[14]. G. M. Van de Ven and A. S. Tolias, "Three scenarios for continual learning," 2019, arXiv:1904.07734.

[15]. P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation," IEEE Trans. Pattern Anal. Mach. Intell., vol. 33, pp. 898–916, 2011.

[16]. A. Babenko, A. Slesarev, A. Chigorin, and V. S. Lempitsky, "Neural codes for image retrieval," in Proc. Eur. Conf. Comput. Vis. (ECCV), 2014, pp. 584–599.

[17]. M. Baktashmotlagh, M. T. Harandi, B. C. Lovell, and M. Salzmann, "Unsupervised domain adaptation by domain invariant projection," in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), 2013, pp. 769–776.

[18]. S. Ben-David, J. Blitzer, K. Crammer, and F. Pereira, "Analysis of representations for domain adaptation," in Proc. Neural Inf. Process. Syst. (NIPS), 2006, pp. 137–144.

[19]. S. Kundu et al., "Quantum computation: From Church-Turing thesis to Qubits," 2016 IEEE 7th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), New York, NY, 2016, pp.15. doi: 10.1109/UEMCON.2016.7777805.

[20]. E. Alpaydin, Introduction to machine learning (The MIT Press, Cambridge, UK, 2014), Ed. 3, Chap. 1 and 11.

[21]. S. Shalev-Shwartz and S. Ben-David, Understanding Machine Learning from Theory to Algorithms (Cambridge University Press, Cambridge, UK, 2014), Chap.1.

[22]. M. A. Nielsen and I. L. Chuang, Quantum Computation and Quantum Information (Cambridge University Press, Cambridge, UK, 2000), Chap. 1 and 6.

[23]. R. P. Feynman, Simulating Physics with Computers, Int. J. Theor. Phys. 21, 467 (1982).

[24]. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," Commun. ACM, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.

[25]. K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA: IEEE, Jun. 2016, pp. 770– 778. doi: 10.1109/CVPR.2016.90.

[26]. R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," in 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA: IEEE, Jun.2014, pp.580–587. doi: 10.1109/CVPR.2014.81.

[27]. S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards RealTime Object Detection with Region Proposal Networks," IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 6, pp. 1137–1149, Jun. 2017, doi: 10.1109/TPAMI.2016.2577031.