A Survey of Few-Shot Learning and Related Paradigms

www.surveyx.cn

Abstract

Few-shot learning (FSL) has become a crucial paradigm in machine learning, enabling models to generalize from minimal labeled data, which is vital in domains with data scarcity like rare disease diagnosis. This survey explores FSL's role in mitigating catastrophic forgetting and enhancing meta-learning models, emphasizing its transformative potential in developing adaptable AI systems. Challenges such as overfitting, data scarcity, and catastrophic forgetting are addressed by related paradigms like continuous learning, meta-learning, transfer learning, and incremental learning. Continuous learning focuses on maintaining knowledge while adapting to new data, while meta-learning and transfer learning enhance adaptability and knowledge transfer across tasks. Incremental learning emphasizes updating models with new data without retraining from scratch. The survey categorizes FSL methods into metric-based, model-based, and optimization-based approaches, highlighting innovations such as unsupervised meta-learning and Bayesian frameworks. The integration of these paradigms offers comprehensive solutions to enhance model performance in data-scarce environments. Future research should focus on expanding these frameworks to new settings, optimizing adversarial training, and leveraging unlabeled data to further advance FSL methods. This comprehensive overview underscores the synergistic potential of integrating FSL with related paradigms to address the challenges of learning from limited data, enhancing model robustness and adaptability across diverse domains.

1 Introduction

1.1 Significance of Few-Shot Learning

Few-shot learning (FSL) has emerged as a crucial paradigm within machine learning, addressing the limitations of traditional methods that require extensive labeled datasets [1]. This approach is particularly vital in contexts where obtaining large volumes of labeled data is impractical or costly, such as rare disease diagnosis and personalized medicine [2]. FSL enables models to generalize from a minimal number of examples, which is essential for tackling classification problems with limited labeled training data [3].

FSL's significance is further underscored by its ability to mitigate challenges related to catastrophic forgetting and enhance robustness in meta-learning models, facilitating improved performance on novel tasks with restricted data availability [3]. This adaptability is transformative, allowing models to efficiently adjust to new and unforeseen tasks. In practical applications, particularly in the vision domain, FSL proves invaluable by enabling classifiers to learn from a limited set of examples [2].

Moreover, FSL reduces the dependency on labeled base class sets, a requirement often deemed costly and impractical in various machine learning applications [1]. By harnessing few-shot learning, models can significantly enhance performance in data-scarce environments, thereby broadening the applicability of machine learning across diverse domains and advancing the capabilities of artificial intelligence systems.

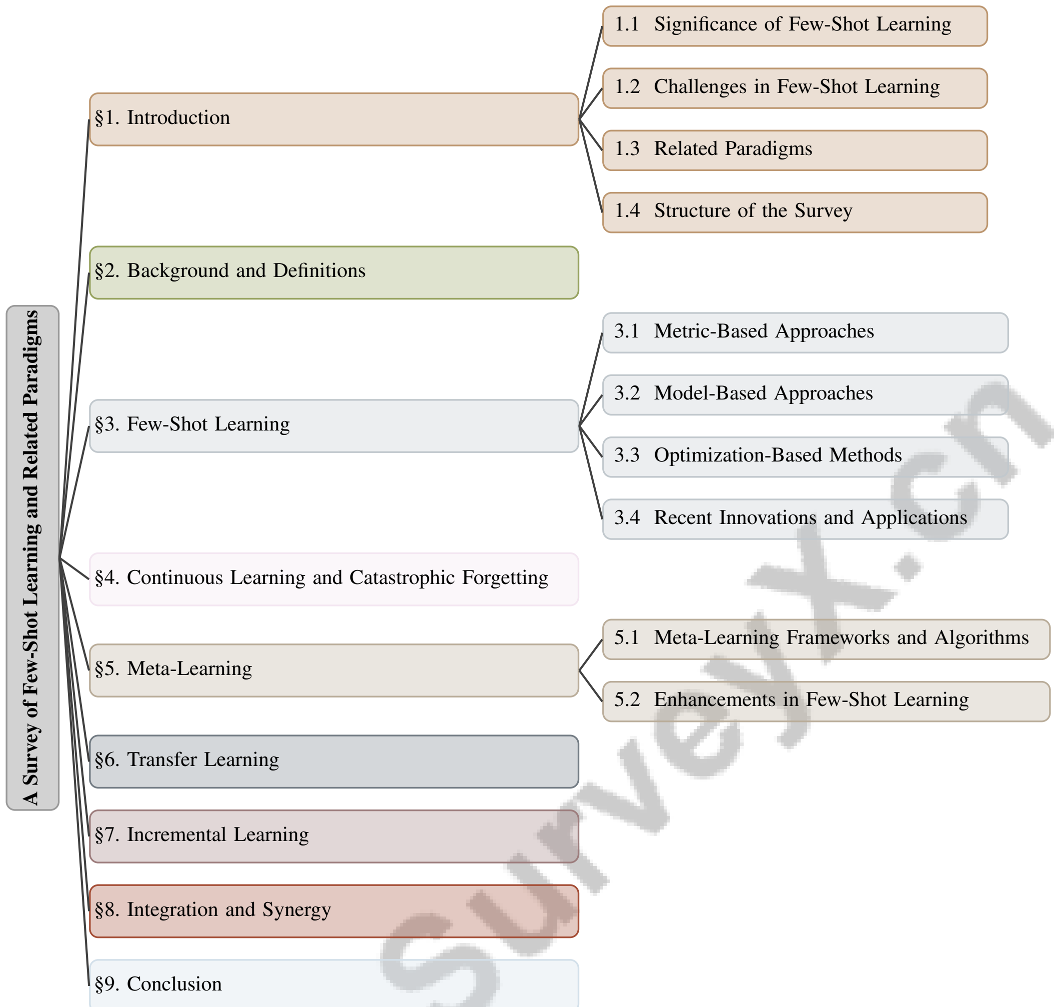


Figure 1: chapter structure

1.2 Challenges in Few-Shot Learning

Few-shot learning (FSL) faces numerous challenges primarily due to the necessity for models to learn effectively from a limited number of labeled examples. A significant issue is the unreliable empirical risk minimizer, which leads to overfitting and poor generalization to new tasks due to small training sample sizes [4]. This challenge is exacerbated by data scarcity, particularly in specialized fields such as rare skin diseases, where long-tailed distribution of datasets further complicates matters [5].

The lack of labeled data for training classifiers poses a critical obstacle, often resulting in overfitting and inadequate generalization in few-shot learning tasks [6]. Additionally, the presence of inductive bias can lead models to confuse closely related categories, compounded by catastrophic forgetting, where previously learned information is lost while acquiring new knowledge.

Traditional neural networks frequently overfit to the limited training data available in FSL contexts, yielding suboptimal performance on unseen classes [7]. Furthermore, existing methods often struggle to function effectively without labeled base classes, limiting their practical applicability [1]. The persistent issue of overfitting highlights the urgent need for innovative approaches to bolster model robustness and adaptability [2].

These challenges underscore the necessity for ongoing research to devise innovative strategies that enhance the generalization capabilities of few-shot learning models. By integrating techniques from continual learning, meta-learning, and the effective utilization of auxiliary datasets, researchers can

AI-generated, for reference only.

develop models that adapt to new tasks with limited examples while maintaining robust performance across various domains and conditions [8, 9, 10, 11].

1.3 Related Paradigms

Few-shot learning (FSL) is intricately linked with several complementary paradigms that address distinct aspects of learning from limited data. Continuous learning, or lifelong learning, emphasizes a model's ability to learn continuously from a data stream while retaining previously acquired knowledge. A significant challenge in this paradigm is catastrophic forgetting, where new learning can disrupt existing knowledge, necessitating novel strategies to ensure effective generalization to new tasks [12].

Meta-learning, or "learning to learn," is pivotal in FSL, enabling models to quickly adapt to new tasks with minimal data. Techniques such as model-agnostic meta-learning (MAML) optimize model initialization for rapid adaptation. The integration of PAC-Bayesian theories provides a theoretical framework for understanding the generalization capabilities of meta-learners, enhancing adaptability in various scenarios. Additionally, meta-representation learning has bridged the gap between pre-training and meta-learning, particularly in contexts lacking global labels [13].

Transfer learning is another paradigm closely associated with FSL, leveraging knowledge from a source domain to improve learning in a target domain. This approach is beneficial when the target domain suffers from data scarcity, allowing models to utilize pre-trained representations from large datasets. However, significant challenges arise when there are substantial differences between source and target domains, necessitating innovative strategies to enhance model generalization. For instance, in document classification, where the shift from natural images to semi-structured documents is considerable, existing techniques may struggle to maintain performance. Recent advancements such as Task Augmented Meta-Learning (TAML) and style transfer-based task augmentation aim to address these challenges by diversifying training styles and improving domain generalization capabilities, thus facilitating effective knowledge transfer across disparate domains [14, 9, 15].

Incremental learning focuses on updating models with new data without retraining from scratch, striving to balance new and old knowledge. This paradigm is crucial for developing systems capable of evolving over time, assimilating new information while preserving existing capabilities [16]. The Interventional Few-Shot Learning (IFSL) framework exemplifies an innovative approach within this context, utilizing causal interventions to address issues arising from pre-trained knowledge [17].

These paradigms collectively advance few-shot learning, each addressing specific challenges and providing complementary strategies to enhance the efficacy and adaptability of machine learning systems in data-scarce environments. The survey categorizes FSL methods into three main perspectives—data, model, and algorithm—focusing on how prior knowledge can improve learning outcomes [4]. Additionally, discussions on meta-learning, transfer learning, and hybrid approaches to FSL offer a comprehensive understanding of the field [18]. Insights into the relationship between scaling training data and few-shot performance further elucidate how scaling impacts model performance [19].

1.4 Structure of the Survey

This survey is systematically organized to comprehensively explore few-shot learning and its related paradigms. The paper begins with an **Introduction** that underscores the significance of few-shot learning (FSL) in machine learning, emphasizing its capacity to address challenges such as data scarcity. It also introduces related paradigms, including continuous learning, meta-learning, transfer learning, and incremental learning [3].

Following the introduction, the **Background and Definitions** section provides detailed explanations of key concepts, establishing a foundational understanding necessary for subsequent sections. This includes definitions and discussions on the interrelations and distinctions between few-shot learning and paradigms like continuous learning and catastrophic forgetting [12].

The survey then progresses to the **Few-Shot Learning** section, subdivided into discussions on metric-based, model-based, and optimization-based approaches, alongside recent innovations and applications across various domains [19].

In the **Continuous Learning and Catastrophic Forgetting** section, the focus shifts to the challenges of continuous learning and strategies to mitigate catastrophic forgetting, emphasizing the importance of maintaining performance on old tasks while learning new ones [16].

The **Meta-Learning** section explores the concept of 'learning to learn,' discussing frameworks like MAML and their impact on enhancing few-shot learning capabilities.

Subsequently, the **Transfer Learning** section reviews transfer learning principles, highlighting its relevance to few-shot learning and successful applications.

The **Incremental Learning** section defines the paradigm and discusses techniques for updating models with new data while balancing old and new knowledge [17].

Finally, the **Integration and Synergy** section examines how these paradigms can be integrated to enhance learning systems, discussing potential synergies and future research directions [4]. The survey concludes by summarizing key insights and reflecting on potential future research directions in these paradigms.The following sections are organized as shown in Figure 1.

2 Background and Definitions

2.1 Conceptual Overview

Few-shot learning (FSL) empowers models to recognize novel categories from minimal labeled examples, typically one to five, crucial in data-scarce contexts and akin to human learning [20, 21]. The asymptotic equipartition property (AEP) underpins FSL, outlining conditions for small datasets to effectively represent input-output distributions [22, 23]. Meta-learning, or "learning to learn," is integral to FSL, enhancing adaptation to new tasks through diverse frameworks like black-box, metric-based, and Bayesian approaches [24]. Strategies such as Fixed Meta-Learning (FIX-ML) emphasize task diversity, improving learning outcomes with diverse support sets [25]. Frameworks like E3BM optimize architectures for few-shot classification, addressing challenges of limited data [26].

Benchmarks in image classification, particularly for novel classes, are vital for evaluating FSL methodologies [19]. Applications in natural language processing, where models must continuously learn from sequential tasks and generalize with few examples, underscore FSL's importance in dynamic, data-scarce environments [11]. FSL research offers strategies to address data scarcity and rapid adaptation by leveraging meta-learning techniques and optimizing model architectures for enhanced generalization across various tasks [4].

2.2 Interrelations and Distinctions

Understanding the interrelations and distinctions among paradigms like few-shot learning (FSL), meta-learning, transfer learning, and multi-task learning is crucial for grasping their unique contributions and synergies. FSL enhances learning efficiency by utilizing structural conditions, moving beyond traditional i.i.d. assumptions to pool data from source tasks [27]. This reliance on structural data properties sets it apart from other approaches.

Meta-learning focuses on optimizing learning processes across tasks via meta-representations, objectives, and optimizers, contrasting with transfer learning, which emphasizes knowledge transfer from a source to a target domain [24]. Unified benchmarking in transfer and meta-learning is critical, given varied evaluation protocols that often obscure cross-paradigm insights [28].

FSL research is categorized into meta-learning-based, non-meta-learning-based, and hybrid approaches, reflecting diverse strategies to address limited data challenges. The interplay between agent-environment interactions and model adaptation strategies highlights the necessity for efficient knowledge transfer and model generalization [24]. Pre-training and meta-learning interconnect, particularly regarding dataset diversity's influence on performance, emphasizing diverse pre-training datasets for enhanced adaptability [21]. The assumption that a small, well-chosen set of examples can achieve generalization comparable to larger datasets is pivotal in FSL, underscoring the importance of example selection [22].

In graph-based few-shot learning, challenges of label and structure scarcity highlight data limitations [28]. Meta-learning's computational demands, particularly numerous gradient steps for convergence, present challenges influencing meta-update efficiency [4]. These interrelations and distinctions elucidate the nuanced roles each paradigm plays in advancing machine learning, offering complementary strategies to address limited data challenges. The categorization into Euclidean and non-Euclidean frameworks further illustrates methodological diversity within FSL and related paradigms [28].

3 Few-Shot Learning

Category	Feature	Method
Metric-Based Approaches	Prototype and Distance-Based	RN[29], PN[30]
	Causal and Bayesian Integration	CFSL[31]
	Self-Supervised Enhancements	PDANet[6]
Model-Based Approaches	Gradient Optimization Techniques	iMAML[32], MAML++[33]
	Data Consistency Strategies	FIX-ML[25]
	Probabilistic and Uncertainty Modeling	PLATIPUS[34]
	Training Dynamics and Monitoring	ABE[35]
Optimization-Based Methods	Latent Space Techniques	DCN[36], MD[37]
	Model Transfer Strategies	TSMD[38]
	Task Enhancement Approaches	MAE[39], TAM[40]
Recent Innovations and Applications	Data Utilization Techniques	MDR[41], UML[1]
	Classification Strategies	MahiNet[42]
	Scalability and Adaptability	MLTP[43]
	Probabilistic Approaches	MQDA[2]

Table 1: This table provides a comprehensive overview of various Few-Shot Learning (FSL) methods categorized into metric-based, model-based, optimization-based approaches, and recent innovations. Each category is further detailed with specific features and methods, showcasing the diverse strategies employed to enhance learning from limited data. The table also references key studies that contribute to the development and application of these FSL techniques.

Few-shot learning (FSL) addresses the challenge of training models with limited data by categorizing strategies into metric-based, model-based, and optimization-based methods, each contributing uniquely to the field. Table 1 presents a detailed classification of Few-Shot Learning methods, illustrating the diverse approaches and innovations that address the challenges of learning from limited data. Table 3 provides a comparative analysis of different metric-based approaches in Few-Shot Learning, detailing their core mechanisms, adaptation strategies, and key innovations. **??** illustrates the hierarchical structure of these FSL methodologies, highlighting the distinctions between the three primary categories: Metric-Based, Model-Based, and Optimization-Based Approaches. Each category is further divided into specific methods and innovations, reflecting the diverse strategies employed to enhance learning from limited data. Metric-based approaches leverage distance metrics in embedding spaces to enhance classification in data-scarce scenarios. Collectively, these approaches address challenges such as data scarcity, model adaptation, and efficient learning, showcasing the evolution and application of FSL techniques in machine learning.

3.1 Metric-Based Approaches

Metric-based approaches are crucial in FSL, employing distance metrics within embedding spaces to classify data effectively with limited samples. Prototypical Networks, a foundational method, compute class prototypes by averaging embedded support points, facilitating classification through distance measures [30]. Relation Networks extend this by using deep distance metrics to compare query images with labeled examples, enhancing classification without additional updates [29].

As illustrated in Figure 2, the categorization of metric-based approaches in Few-Shot Learning (FSL) highlights foundational methods, adaptive strategies, and innovative techniques. Foundational methods include Prototypical Networks and Relation Networks, which set the groundwork for FSL. Adaptive methods such as Deep Subspace Networks (DSN) and MF-Adapter refine these strategies by incorporating dynamic and causal elements, respectively [44, 45]. Innovative strategies like PDANet and MetaQDA introduce novel concepts such as self-supervised learning and Bayesian inference, enhancing the robustness and efficiency of FSL [6, 2].

Causal Few-Shot Learning Methods (CFSL) incorporate causal mechanisms to enhance interpretability and metric diversity [31]. Self-supervised learning integration, as seen in PDANet, selects discriminative parts from unlabeled images to augment support sets, improving robustness [6]. LSTM-based methods dynamically adjust learning rates, optimizing the training of learner classifiers from few examples [46].

Bayesian perspectives, exemplified by MetaQDA, utilize Bayesian inference on classifier parameters with conjugate priors for efficient learning [2]. Spectral-based regularization aligned with multi-task representation assumptions further enhances metric-based methods [47].

Collectively, these methodologies illustrate the evolution of metric-based approaches in FSL, incorporating innovative strategies to address limited data challenges. Techniques from transfer learning, meta-learning, and Bayesian methods enhance generalization to novel classes with minimal samples, especially in complex domains like document classification. Advancements such as the Category Traversal Module improve feature relevance identification, boosting classification performance. FSL offers a promising low-cost solution for machine learning applications, addressing data collection limitations and computational demands of traditional deep learning methods [48, 9, 10, 49]. The refinement of distance metrics and integration of additional learning paradigms remain crucial for adaptability and effectiveness across learning scenarios.

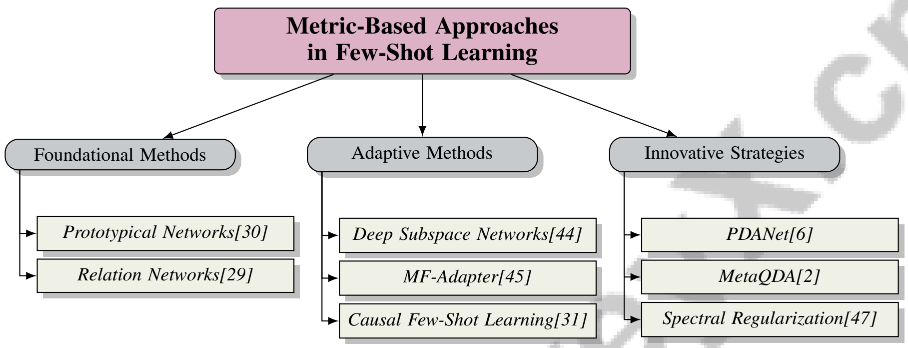


Figure 2: This figure illustrates the categorization of metric-based approaches in Few-Shot Learning (FSL), highlighting foundational methods, adaptive strategies, and innovative techniques. Foundational methods include Prototypical Networks and Relation Networks, which set the groundwork for FSL. Adaptive methods such as Deep Subspace Networks and MF-Adapter refine these strategies by incorporating dynamic and causal elements. Innovative strategies like PDANet and MetaQDA introduce novel concepts like self-supervised learning and Bayesian inference, enhancing the robustness and efficiency of FSL.

3.2 Model-Based Approaches

Method Name	Architectural Innovations	Optimization Techniques	Task Adaptability
iMAML[32]	Meta-learning Methods	Implicit Differentiation	Minimal Data
MAML++[33]	Multi-step Loss	Derivative-order Annealing	Dynamic Learning Rates
PLATIPUS[34]	Probabilistic Framework	Variational Inference Techniques	Sampling Multiple Models
ABE[35]	Activation Dynamics	Activation Statistics	Distributional Shifts
FIX-ML[25]	Fixed Support Pools	Fixed Support Pool	Fixed Support Pool

Table 2: Overview of model-based approaches in few-shot learning, detailing architectural innovations, optimization techniques, and task adaptability. The table compares methods such as iMAML, MAML++, PLATIPUS, ABE, and FIX-ML, highlighting their unique contributions to addressing data scarcity and enhancing model performance.

Model-based approaches in FSL focus on optimizing model architectures for learning from limited data, addressing data scarcity and computational demands. These approaches often incorporate meta-learning and transfer learning techniques to adapt models to new tasks with minimal data. Task-specific meta distillation, for example, involves training a large teacher model alongside a compact student model, enabling knowledge transfer for improved few-shot performance [38, 10].

Implicit Model-Agnostic Meta-Learning (iMAML) uses implicit differentiation for efficient meta-gradient computation, optimizing meta-learning processes [32]. MAML++ enhances the MAML framework by improving stability and efficiency, facilitating better task adaptation [33].

PLATIPUS employs a probabilistic framework, using variational inference to model uncertainty and adapt to new tasks robustly [34]. Activation-Based Early-stopping (ABE) uses neural activation dynamics to guide early-stopping, preventing overfitting and enhancing generalization [35].

The FIX-ML method optimizes learning algorithms using a fixed support pool across tasks, emphasizing consistent support sets for improved generalization [25].

Table 2 provides a comprehensive comparison of various model-based approaches in few-shot learning, illustrating the architectural innovations, optimization techniques, and task adaptability that these methods employ to enhance learning from limited data. These model-based approaches highlight the contributions of architectural innovations and optimization techniques for efficient learning from limited data, enhancing performance across diverse tasks. They address data scarcity and computational demands while identifying task-relevant features, ultimately improving FSL system effectiveness in real-world applications [49, 48, 10]. Efficient model design and dynamic learning strategies enhance adaptability and performance in data-scarce environments.

3.3 Optimization-Based Methods

Optimization-based methods in FSL refine learning processes to enable quick adaptation to new tasks with limited data. These methods leverage meta-learning to determine optimal parameter updates and improve convergence, utilizing auxiliary datasets and frameworks to achieve performance akin to data-rich deep learning approaches [9, 46, 8, 10].

Meta-Dropout introduces a regularization mechanism by perturbing latent features, dynamically adjusting dropout rates to enhance generalization and robustness against overfitting [37]. The Decoder Choice Network (DCN) restricts model parameters to a low-dimensional latent space, facilitating efficient learning through essential parameter updates [36].

Task-Specific Meta Distillation (TSMD) trains a teacher and student model, enabling effective task adaptation by distilling knowledge from the teacher [38]. Meta-Autoencoder (MAE) generates auxiliary tasks, enhancing training diversity and generalization [39]. The Transformer Adaptation Method (TAM) optimizes task-specific parameters in transformer models, leveraging their flexibility for efficient learning [40].

These optimization-based methods illustrate advanced strategies, including meta-learning, transfer learning, and hybrid approaches, enhancing learning processes in few-shot tasks. They address challenges like limited samples, domain shifts, and efficient feature extraction, improving generalization and performance in machine learning applications [10, 9, 46, 49, 47]. By focusing on regularization, dimensionality reduction, knowledge distillation, task augmentation, and model adaptation, these approaches enhance FSL models' efficiency and effectiveness in data-scarce environments.

3.4 Recent Innovations and Applications

Recent advancements in FSL have expanded its applicability across domains, driven by innovative methodologies enhancing performance and generalization. Unsupervised meta-learning (UML) methods outperform existing methods on few-shot classification benchmarks, leveraging unlabeled data effectively [1].

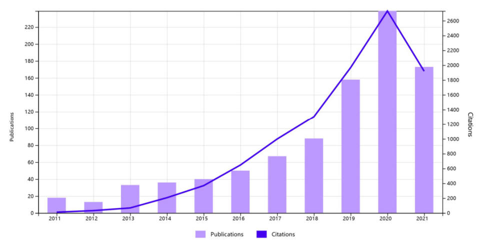
MetaQDA integrates Bayesian principles, demonstrating superior performance in cross-domain and class-incremental scenarios with robust uncertainty modeling [2]. Regularization through theoretical assumptions improves performance and learning rates, emphasizing theoretical frameworks' role in efficient learning strategies [47].

Frameworks like MahiNet address the many-class few-shot problem by utilizing class hierarchy information, outperforming existing methods in supervised and meta-learning scenarios [42]. LSTM-based meta-learners enhance performance in few-shot tasks, optimizing neural network training with limited data [46].

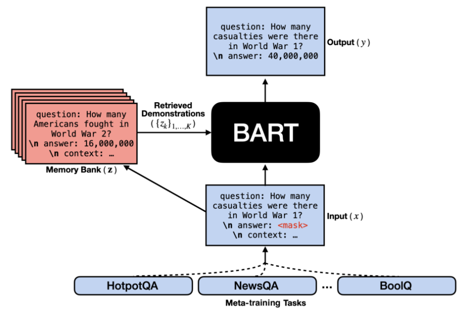
The MLTP framework improves generalization across network sizes on benchmark datasets, highlighting scalable learning strategies' importance [43].

Recent FSL advancements reflect rapid evolution, effectively tackling data scarcity challenges and enhancing model generalization across applications like document classification and image recognition. These innovations draw from transfer learning, meta-learning, and hybrid approaches, enabling models to learn from limited labeled data and leverage auxiliary datasets. While traditional methods struggle with novel domain generalization, recent research addresses these limitations, particularly in semi-structured documents, indicating a promising trajectory for FSL developments

[48, 9, 8, 10]. As novel techniques and evaluation frameworks develop, FSL applications are anticipated to expand, driving advancements in challenging contexts.



(a) The graph depicts the growth in publications and citations over a decade.[48]



(b) Meta-training for Question Answering with BART[41]

Figure 3: Examples of Recent Innovations and Applications

As shown in Figure 3, FSL has emerged as a pivotal innovation in machine learning, characterized by its ability to learn from limited examples. This approach has gained significant traction over the past decade, as illustrated by the growth in publications and citations depicted in the accompanying graph. The graph highlights the increasing academic interest and research output in this area, with a notable rise in both publications and citations from 2011 to 2021. Concurrently, recent advancements in FSL have led to innovative applications, such as the meta-training process for question answering using the BART (Bilingual Attention Transformer) model. The flowchart of this process illustrates how BART utilizes a memory bank of demonstrations to effectively answer questions by retrieving relevant demonstrations based on the input question and context. These examples underscore the rapid evolution and broadening application of FSL techniques, which continue to push the boundaries of what is possible with limited data in machine learning [48, 41].

Feature	Prototypical Networks	Relation Networks	Deep Subspace Networks (DSN)
Core Mechanism	Class Prototypes	Deep Metrics	Dynamic Subspaces
Adaptation Strategy	Distance Measures	Query Comparison	Causal Elements
Key Innovation	Embedding Averaging	NO Updates Needed	Adaptive Strategies

Table 3: This table provides a comparative analysis of three prominent metric-based approaches in Few-Shot Learning: Prototypical Networks, Relation Networks, and Deep Subspace Networks (DSN). It highlights the core mechanisms, adaptation strategies, and key innovations of each method, showcasing their distinct contributions to addressing the challenges of learning from limited data.

4 Continuous Learning and Catastrophic Forgetting

4.1 Challenges in Continuous Learning

Continuous learning in few-shot contexts is challenged by data scarcity and dynamic learning environments. Overfitting is prevalent when models trained on small datasets fail to generalize to new tasks, particularly when instance relationships are unclear, leading to inadequate graph constructions [6]. Task-overfitting, which results in models becoming overly specialized, further complicates learning. Mitigating this requires integrating prior knowledge and attention-based meta-learning methods to enhance robustness across scenarios [50]. Misalignment between meta-learning objectives and whole-classification tasks can hinder generalization to novel classes, necessitating aligned learning objectives [51].

Catastrophic forgetting, where models lose previously acquired knowledge when learning new information, is exacerbated by environments with numerous classes and few examples per class, making consistent performance across varying class distributions difficult [46]. Addressing these issues requires techniques that refine knowledge representation and mitigate bias from data scarcity [52].

AI-generated, for reference only.

Model interpretability is another concern, as black-box models lack transparency. Enhancing transparency through interpretable concept-based prototypical networks and leveraging class hierarchy, as seen in MahiNet, can improve prediction accuracy and address challenges associated with large class diversity [53, 42].

Balancing few-shot performance with computational efficiency is crucial. The Transformer Adaptation Method (TAM) exemplifies strategies that achieve a favorable trade-off between performance and efficiency, enabling adaptation without extensive architectural changes [40]. Approaches that reduce the impact of confounding variables and enhance robustness through diverse representation are vital for improving generalization in continuous learning scenarios [31, 7].

The challenges in few-shot learning and continuous adaptation highlight the need for innovative methodologies and robust evaluation frameworks. These are essential for effective learning across diverse tasks, especially in scenarios involving domain shifts like semi-structured document classification and dynamic knowledge accumulation from diverse NLP tasks. Existing models often struggle with generalization due to fixed dataset limitations or catastrophic forgetting, necessitating a unified approach integrating continual learning with few-shot capabilities. Addressing these complexities is crucial for enhancing performance and resilience in varied and unpredictable environments [9, 54, 11].

4.2 Strategies to Mitigate Catastrophic Forgetting

Mitigating catastrophic forgetting in continuous learning, particularly in few-shot contexts, is crucial for models to retain historical knowledge while adapting to new tasks. Specialized training approaches tailored to manage long-tailed distributions are effective in reducing the risk of forgetting previously learned information [5]. The MLCC method exemplifies a strategy that realigns semantic distributions to maintain consistency across tasks, allowing models to adapt without compromising prior knowledge [55].

Incorporating meta-learning frameworks provides another avenue for addressing catastrophic forgetting by retaining historical knowledge while dynamically adjusting to new tasks [3]. MetaQDA emphasizes efficient classifier layer learning without fine-tuning features, crucial for preventing forgetting in few-shot scenarios [2]. Unsupervised meta-learning techniques reduce labeling costs and improve generalization by leveraging unlabeled data, addressing data scarcity and forgetting challenges [1].

Challenges with meta-learning algorithms' reliance on probabilistic learning bounds highlight the need for strategies that effectively utilize all available data [47]. By overcoming these limitations, meta-learning frameworks can better support continuous learning environments.

Dynamic adaptation, semantic alignment, and efficient data utilization are crucial in addressing catastrophic forgetting, as shown by continual learning frameworks like CLIF, which enable knowledge accumulation from diverse tasks while maintaining performance on learned tasks. Techniques such as task-and-layer-wise attenuation in meta-learning help mitigate conflicts during adaptation, enhancing generalization from limited examples across domains. This multifaceted approach underscores the significance of these strategies in improving learning efficiency and performance in few-shot scenarios [56, 9, 54, 57, 11]. Integrating these techniques allows continuous learning systems to achieve sustained performance across tasks, ensuring robust knowledge retention and adaptability.

5 Meta-Learning

Meta-learning has become a cornerstone in artificial intelligence, particularly for addressing the challenges of learning from limited data. This section explores frameworks and algorithms that form the foundation of meta-learning, enhancing adaptability and efficiency in machine learning models. By examining these methodologies, we gain insights into mechanisms that facilitate effective learning in few-shot scenarios, setting the stage for discussing frameworks and algorithms that have significantly advanced this domain.

5.1 Meta-Learning Frameworks and Algorithms

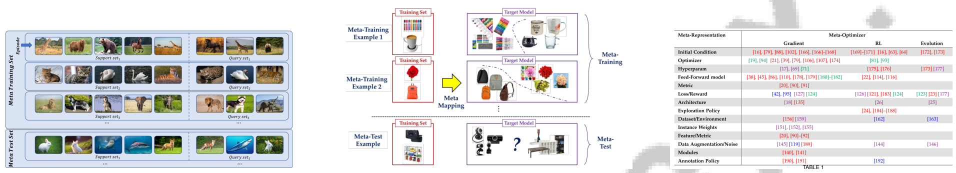
Meta-learning, or "learning to learn," encompasses frameworks and algorithms designed to boost model adaptability and efficiency, especially in few-shot contexts. A pivotal approach is the Model-

AI-generated, for reference only.

Agnostic Meta-Learning (MAML), which optimizes initial parameters for rapid task adaptation with minimal updates. MAML++ further refines this by enhancing stability, generalization, and reducing computational costs [33]. The Task Attended Meta-Learning (TAML) framework uses attention mechanisms to prioritize tasks, improving the meta-learning process [58]. Hierarchical meta-learning (HML) introduces structures for efficient task adaptation across varying task structures [24].

Probabilistic frameworks, such as those employing hierarchical Bayesian models, enhance adaptability by tailoring hyperparameters to task-specific characteristics, improving knowledge transfer [47]. The Meta-Learning Based Training Procedure (MLTP) treats training batches as tasks, optimizing performance on both current and new tasks [43]. The FIX-ML framework exemplifies innovation by optimizing learning algorithms with a fixed support pool, enhancing model generalization [25].

The BOUNCE GRAD algorithm uses modular approaches, combining neural network modules for new tasks via simulated annealing and gradient descent [59]. Shin et al.'s method allows frequent meta-updates by shifting learning trajectories based on parameter changes, boosting adaptability [60]. These diverse methodologies advance meta-learning, offering innovative solutions for swift adaptation and effective learning from minimal datasets, applicable in few-shot learning, natural language processing, and robotics [25, 24, 61, 62, 63].



(a) Meta-Training and Meta-Testing for Animal Image Classification[64]

(b) Meta-Training and Meta-Testing for Fine-Grained Image Classification[65]

(c) Meta-Representation and Meta-Optimizer in Machine Learning[66]

Figure 4: Examples of Meta-Learning Frameworks and Algorithms

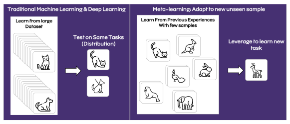
Figure 4 illustrates the versatility of meta-learning across contexts, showcasing frameworks like "Meta-Training and Meta-Testing for Animal Image Classification," which enhances classification through dual-phase structure. Similarly, "Meta-Training and Meta-Testing for Fine-Grained Image Classification" employs distinct training sets to refine classification. "Meta-Representation and Meta-Optimizer in Machine Learning" highlights the role of meta-representations and optimizers in refining processes through gradient, reinforcement learning, and evolutionary strategies [64, 65, 66].

5.2 Enhancements in Few-Shot Learning

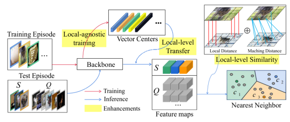
Meta-learning has propelled few-shot learning by introducing methodologies that enhance adaptability, data efficiency, and robustness. The L2F framework, with its selective forgetting mechanism, improves generalization and optimization landscapes [56]. Modular approaches, like Attentive Independent Mechanisms (AIM), use mixtures of experts to enhance learning efficiency and mitigate forgetting [67]. Task similarity is crucial for effective adaptation, emphasizing strategies that leverage this for optimal outcomes [24].

Benchmarks, such as those by [68], provide insights into few-shot learning's practical applications. LGM-Net exemplifies advancements by achieving competitive performance without extensive parameter tuning [20]. Bi-level Hypernetworks for Adapters with Regularization mitigate catastrophic forgetting and enhance capabilities [11]. Integrating few-shot learning with transfer learning enhances performance in specialized domains, like rare skin disease classification [5]. The MLTP consistently improves generalization, highlighting meta-learning's role in optimizing strategies [43]. MetaQDA's fixed feature extraction framework underscores the importance of stable representations for rapid adaptation [2].

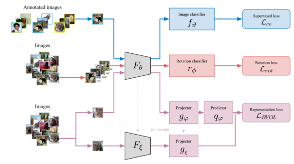
Advancements in meta-learning offer diverse strategies to tackle challenges of learning from limited data, improving model adaptability and performance. These strategies enable rapid adaptation to unseen tasks, addressing traditional deep learning's limitations in generalization and extensive data requirements. Recent research categorizes meta-learning approaches into metric-based, memory-based, and learning-based methods, facilitating models that perform well with minimal samples across applications, including natural language processing and robotics [69, 10, 24, 64, 51].



(a) Meta-learning: Adapt to new unseen sample[64]



(b) Local-level Transfer for Local-agnostic Image Retrieval[70]



(c) Deep Learning Framework for Image Rotation Classification[71]

Figure 5: Examples of Enhancements in Few-Shot Learning

As depicted in Figure 5, few-shot learning addresses training models with minimal data. Meta-learning enhances few-shot capabilities, adapting to unseen samples by leveraging prior experiences. Innovations in local-level transfer methods for image retrieval allow effective retrieval in a local-agnostic manner, while deep learning frameworks for image rotation classification enhance model understanding and prediction of orientations. These examples highlight the strides in few-shot learning, driven by meta-learning and novel methodologies, creating robust models capable of thriving with limited data [64, 70, 71].

6 Transfer Learning

6.1 Principles of Transfer Learning

Transfer learning is a critical machine learning paradigm that enhances model performance by utilizing knowledge from one domain to improve learning in another related domain. This is particularly beneficial in few-shot learning, where models generalize from specific classes to unseen ones. Meta-transfer learning exemplifies this by adapting neural networks to new tasks through learned weight transfer strategies. Style transfer-based task augmentation further bolsters model generalization by diversifying training styles, effectively addressing domain shifts [15, 72, 73, 74]. These strategies are invaluable when labeled data is scarce, enabling models to leverage pre-trained knowledge for improved performance on new tasks.

A fundamental principle of transfer learning is ensuring class separability and model expressiveness, providing a theoretical basis for how models trained on one set of classes can perform effectively on another [74]. This principle emphasizes the need for learned representations to capture target domain nuances while effectively separating distinct classes.

Transfer learning typically involves two stages: pre-training and fine-tuning. In the pre-training stage, a model learns general features from a large source domain dataset, which are then fine-tuned on a smaller target domain dataset. This addresses domain shifts and improves performance without extensive labeled data [75, 9, 76, 77].

Addressing domain shift discrepancies is crucial in transfer learning, as effective approaches must mitigate these to ensure applicable transferred knowledge. Techniques like domain adaptation and generalization maintain model performance across diverse domains, especially in few-shot learning and document classification. Recent advancements integrate transfer learning with meta-learning algorithms, enhancing adaptability and accuracy. Fine-tuning on specialized datasets improves few-shot learning performance, suggesting that these techniques' effectiveness may stem from broader adaptations to few-shot scenarios rather than solely domain adaptation [76, 9, 54].

The modularity and flexibility of deep learning architectures facilitate seamless adaptation to new tasks and datasets. Reusable neural network modules allow combinatorial generalization, leveraging prior knowledge from related tasks to reduce training data requirements and improve learning efficiency. Advancements in meta-transfer learning, particularly in few-shot learning, illustrate these architectures' potential to generalize across diverse applications, including classification tasks and robotics [59, 72, 73, 74]. This adaptability extends the applicability of pre-trained models across various real-world tasks, enhancing their practical utility.

Transfer learning principles focus on strategic knowledge reuse, domain variation management, and model architecture optimization to improve performance across diverse tasks. This approach not only facilitates model generalization to unseen classes but also integrates meta-learning strategies that adapt neural networks for few-shot learning. By utilizing large-scale pre-trained models and techniques like weight scaling and shifting, transfer learning maximizes efficiency and accuracy in challenging environments [73, 74]. These principles highlight the paradigm's potential to significantly reduce data requirements and computational resources for training effective machine learning models in new domains.

6.2 Applications and Techniques

Transfer learning has achieved significant success across various applications, especially where labeled data is limited. In computer vision, pre-trained models on large datasets like ImageNet are fine-tuned for specific tasks such as object detection and segmentation, yielding impressive results with minimal task-specific data by transferring expressive and separable learned features [74].

In natural language processing (NLP), transfer learning has transformed tasks like sentiment analysis, machine translation, and question answering. Models such as BERT and GPT, pre-trained on extensive datasets, undergo fine-tuning for specific tasks, significantly enhancing accuracy and efficiency. Fine-tuning on targeted datasets, like software documentation, can improve performance by up to 7.5

In healthcare, transfer learning aids in developing diagnostic models, particularly where medical imaging data is scarce, such as with rare diseases. This approach effectively utilizes pre-trained models to enhance feature representation and classification performance, even with limited labeled examples. Combining transfer learning with few-shot learning techniques can significantly improve accuracy, especially in skin disease classification datasets. By leveraging auxiliary information from existing labeled data and employing data augmentation strategies, transfer learning addresses challenges posed by long-tailed data distributions and enhances diagnostic models' robustness [5, 78]. By transferring knowledge from models trained on related datasets, transfer learning aids in creating robust models for disease detection and classification, essential in medical domains where data collection is hindered by privacy concerns and rarity of conditions.

In autonomous driving, transfer learning enables models trained in simulated environments to adapt to real-world scenarios, significantly reducing the need for extensive real-world data collection and expediting autonomous systems' development. Techniques like meta-learning and few-shot learning allow models to generalize from limited data, making the training process more efficient and feasible, particularly in domains where data acquisition is challenging or impractical [79, 80, 9, 48, 81].

Transfer learning techniques often include domain adaptation strategies that fine-tune models to mitigate domain shift effects between source and target domains, crucial for enhancing model generalization across different classes, especially in few-shot learning scenarios where performance with limited labeled samples is essential. Recent advancements include methods leveraging style transfer for task augmentation and meta-learning approaches integrating transfer learning with fine-tuning to improve performance under significant domain shifts, facilitating more robust classification in diverse applications [77, 9, 15, 76, 74]. This is achieved through techniques like adversarial training and domain-invariant feature extraction, enhancing model generalization across various environments.

The modular architecture of deep learning models enables seamless incorporation of transfer learning techniques, facilitating the reuse and adaptation of pre-trained components. This modularity allows for developing reusable neural network modules that can be combined in various configurations to efficiently address new tasks, enhancing generalization capabilities. By leveraging previously learned modules, deep learning systems achieve combinatorial generalization, akin to flexible sentence construction in language, ultimately reducing training data requirements for new tasks and improving performance across related domains, including robotics and few-shot learning scenarios [59, 73]. This modularity is particularly advantageous in multi-task learning scenarios, where shared representations can be fine-tuned for multiple related tasks, optimizing resource utilization and improving overall model performance.

AI-generated, for reference only.

7 Incremental Learning

7.1 Definition and Objectives of Incremental Learning

Incremental learning is a machine learning paradigm focused on the continuous adaptation of models to new data while preserving previously acquired knowledge. This approach is essential in dynamic environments where data is constantly generated, allowing models to evolve without the need for complete retraining. Its primary goal is to integrate new information into the model's knowledge base while mitigating catastrophic forgetting, crucial for applications like few-shot learning, where models must recognize new categories with limited examples while maintaining performance on established tasks. Advanced methodologies such as Continual Learning of Few-Shot Learners (CLIF) and federated few-shot class-incremental learning enhance generalization capabilities and preserve existing knowledge in decentralized or data-scarce settings. Strategies focusing on training within flat minima have shown promise in reducing the impact of catastrophic forgetting, thereby improving learning efficiency and adaptability [82, 59, 83, 11].

A critical aspect of incremental learning is balancing the integration of new knowledge with the retention of existing knowledge to maintain model performance across diverse tasks, especially as new classes are introduced over time. The Knowe model exemplifies an effective strategy for class-incremental learning, normalizing and freezing classifier weights after learning from coarse labels, which allows adaptation to new fine classes without forgetting previously learned information [84].

Incremental learning methodologies utilize a variety of advanced techniques, including rehearsal strategies, regularization methods, and architectural adjustments, to tackle catastrophic forgetting. Recent studies emphasize the importance of these approaches, particularly in incremental few-shot learning settings, where models must adapt to new categories with limited examples while retaining prior knowledge. For instance, searching for flat minima during base training can help maintain model performance across tasks, while novel regularization techniques enhance generalization capabilities. Rehearsal methods reinforce previous knowledge by replaying a subset of past data, while regularization techniques impose constraints on model updates to preserve learned weights. Architectural approaches may involve dynamic network expansion or modular design to accommodate new information without disrupting existing structures [56, 82, 65, 85, 11].

Incremental learning seeks to develop models that are adaptable, scalable, and proficient in maintaining high performance as they encounter new data. This approach enables models to continuously expand their knowledge base and generalize rapidly to new tasks, addressing challenges such as catastrophic forgetting and data scarcity. By employing techniques like searching for flat minima during training, models can effectively learn new categories while preserving previously acquired knowledge, thus enhancing their overall generalization capabilities in dynamic environments [82, 65, 61, 11]. This paradigm is crucial for applications in fields such as robotics, autonomous systems, and personalized recommendation systems, where continuous learning from evolving data streams is essential for operational efficacy.

7.2 Challenges in Balancing New and Old Knowledge

Incremental learning faces unique challenges in maintaining a balance between integrating new information and preserving existing knowledge. A primary difficulty is catastrophic forgetting, where model performance on previously learned tasks degrades as new knowledge is acquired. This issue is exacerbated in scenarios with significant changes in data distribution, necessitating strategies to manage the stability-plasticity dilemma [84].

The stability-plasticity trade-off is fundamental in incremental learning, with stability referring to the retention of past knowledge and plasticity denoting the capacity to incorporate new information. Achieving an optimal balance between representing task-relevant features and adapting to diverse tasks is crucial for maintaining robust performance across few-shot learning scenarios, enabling effective leverage of intra-class similarities and inter-class distinctions while minimizing sample complexity [86, 47, 49]. Techniques such as regularization, which constrains model parameter updates, and rehearsal, which involves replaying past data, are commonly employed to address this trade-off.

A significant challenge in machine learning is the risk of overfitting to new data, particularly when that data is limited or imbalanced. This can heighten the model's sensitivity to randomness and variability in training outcomes, as observed in few-shot learning and meta-learning approaches that aim to generalize from minimal labeled samples [87, 9, 28, 10]. This may lead to a model overly specialized in recent tasks at the expense of generalizing to previously learned tasks. Methods like elastic weight consolidation (EWC) mitigate this risk by selectively updating model parameters based on their importance to past tasks, preserving essential knowledge while allowing for the integration of new information.

Moreover, the dynamic nature of incremental learning environments often requires models to adapt to new classes or tasks without access to the complete dataset. This constraint necessitates efficient memory management strategies to retain the most relevant information for future learning. Exemplar-based methods, which retain representative samples of previous tasks, are valuable in few-shot learning, effectively balancing the integration of new information with the preservation of existing knowledge while minimizing memory usage and computational demands. This approach is particularly advantageous in scenarios with scarce or costly data, enabling efficient learning from limited examples and reducing the time and resources typically required for machine learning applications [18, 10].

The challenges of balancing new and existing knowledge in incremental learning underscore the urgent need for innovative strategies that dynamically adapt to evolving data streams while ensuring robust retention of previously acquired knowledge. Recent advancements, such as the Continual Learning of Few-Shot Learners (CLIF) framework, illustrate the potential to integrate continual learning with rapid generalization across diverse tasks, addressing issues like catastrophic forgetting and data scarcity. Furthermore, approaches like searching for flat local minima during base training can enhance the model's ability to efficiently learn new categories while preserving performance on previously learned tasks. These developments suggest that a unified framework, drawing insights from both meta-learning and traditional supervised learning, could significantly improve generalization and performance in few-shot learning scenarios [82, 65, 61, 11]. Addressing these challenges is critical for the successful deployment of incremental learning systems in real-world applications, where continuous adaptation and high performance are essential.

8 Integration and Synergy

8.1 Enhancing Few-Shot Learning with Meta-Learning and Transfer Learning

Meta-learning and transfer learning integration has become a powerful approach to enhance few-shot learning by improving model adaptability and leveraging pre-trained knowledge. Meta-learning optimizes learning processes across tasks, significantly boosting generalization and adaptation speed in few-shot contexts [46]. The synergy between meta-learning and few-shot learning is exemplified by LGM-Net, which uses prior knowledge to swiftly adapt to new tasks [20]. Transfer learning complements this by enabling models to utilize pre-trained features from related tasks, enhancing generalization without extensive labeled data. Techniques such as scaling and shifting functions for adapting pre-trained networks in meta-transfer learning (MTL) exemplify this approach, effectively transferring knowledge to improve few-shot learning capabilities [5]. Simulated data augmentation further showcases the synergy between simulated data and few-shot learning [88].

Task-adaptive architecture search techniques, like those in MetAdapt, emphasize dynamically tailoring architectures to improve learning efficiency and adaptability [3]. Moreover, integrating FIX-ML with few-shot learning paradigms enhances learning by focusing on fixed support sets, providing a novel perspective on support set diversity [25]. The KCL method can also be integrated with few-shot and zero-shot learning methods, enhancing their capabilities without generative models or auxiliary databases [52]. Collectively, these methodologies underscore the transformative potential of integrating meta-learning and transfer learning in addressing few-shot learning challenges [47]. By leveraging shared representations and optimizing learning processes, these approaches significantly enhance the adaptability and performance of few-shot learning models across various domains.

8.2 Mitigating Catastrophic Forgetting through Continuous and Incremental Learning

Addressing catastrophic forgetting in continuous and incremental learning is crucial for maintaining model performance over time, especially in dynamic data environments. Incremental learning techniques, which preserve previously acquired knowledge while integrating new information, are promising. The Knowe model exemplifies this with a coarse-to-fine strategy that normalizes and freezes classifier weights after learning from coarse labels, enabling adaptation to new fine classes without losing prior knowledge [84]. Continuous learning approaches often employ regularization techniques to constrain model updates, preserving essential knowledge while integrating new tasks. Rehearsal methods, which replay subsets of past data, reinforce previous knowledge, reducing the risk of forgetting [87, 25].

Integrating continuous and incremental learning paradigms establishes a robust framework for mitigating catastrophic forgetting, especially in incremental few-shot learning, where models adapt to new categories with limited data while maintaining performance on previously learned tasks. This approach emphasizes training strategies that prioritize knowledge preservation and incorporates techniques like regularized adapter generation and task-dependent hyperparameter selection to enhance generalization and adaptability across diverse tasks [82, 67, 89, 85, 11]. By combining dynamic adaptation strategies with robust knowledge retention mechanisms, these approaches offer promising solutions for developing models capable of sustaining high performance across various tasks and conditions.

8.3 Synergies between Few-Shot Learning and Multi-Modal Data Integration

Integrating multi-modal data into few-shot learning frameworks presents a promising avenue for enhancing model performance and adaptability. Leveraging multiple modalities—such as visual, auditory, and textual information—enables models to achieve a more comprehensive understanding of complex concepts, improving generalization from limited examples. This multi-modal approach aligns with principles of human-like learning, where diverse sensory inputs facilitate more robust learning processes [21]. By combining diverse modalities, such as images and textual descriptions, models can enhance feature representations, leading to substantial performance improvements [90, 91, 49].

The synergy between few-shot learning and multi-modal integration also reduces reliance on large labeled datasets. By utilizing complementary information, models can achieve high performance with fewer labeled examples, addressing data scarcity challenges in real-world applications. This integration is particularly significant in fields like healthcare and autonomous systems, where diverse data types—images, text, sensor readings—are frequently available. Effectively combining these modalities enhances decision-making processes, leading to improved outcomes; for example, leveraging textual descriptions alongside visual data has shown to increase classification accuracy by approximately 30

The integration of multi-modal data into few-shot learning marks a significant advancement, enhancing performance by leveraging complementary information from different modalities. Research indicates that attention-based fusion techniques can boost classification accuracy by around 30

8.4 Future Directions and Challenges in Paradigm Integration

Integrating few-shot learning with related paradigms such as meta-learning, transfer learning, and continuous learning offers numerous opportunities for advancing machine learning capabilities. Future research should focus on developing hybrid models that effectively combine data augmentation, embedding techniques, and semantic information to enhance learning efficiency and accuracy [4]. Exploring sophisticated knowledge graph integration techniques and emerging trends in artificial intelligence will be essential for enhancing existing methods and addressing challenges related to data scarcity and task diversity.

In unsupervised learning, future research will aim to develop better algorithms and increase the quantity of unlabeled data to enhance unsupervised meta-learning (UML), indicating further research directions in paradigm integration. Additionally, improving part selection mechanisms and exploring various augmentation strategies in few-shot contexts are promising avenues, particularly for complex datasets or tasks [6]. Refining initialization strategies and investigating the underlying causes of

AI-generated, for reference only.

optimization strategies' effectiveness in meta-learning are essential for enhancing the robustness and adaptability of these approaches. Future efforts should also focus on expanding problem setups to include multi-modal learning and improving data augmentation techniques, critical for advancing theoretical foundations and exploring new applications of few-shot learning [4].

Advanced knowledge distillation methods and optimizations are necessary to enhance the efficiency of task-specific meta-distillation approaches, particularly in few-shot learning contexts. Balancing online performance with the challenge of catastrophic forgetting, alongside investigating diverse environments and tasks, is crucial for validating benchmarks and improving practical applicability. This includes addressing task-level evaluation for reliable performance estimations and refining meta-learning frameworks to enhance knowledge precision and generalization capabilities [92, 61, 77, 49, 85].

9 Conclusion

Few-shot learning, alongside paradigms such as meta-learning, transfer learning, and incremental learning, holds significant promise in overcoming challenges of data scarcity and enhancing model versatility across diverse applications. These methodologies collectively improve model adaptability, enabling superior performance across various tasks. In particular, meta-learning frameworks, like MLCC, demonstrate efficacy in addressing issues such as inductive bias and catastrophic forgetting, thereby ensuring robust performance across a wide range of datasets. The application of w-dropout further strengthens the generalization abilities of few-shot learning models, suggesting that integrating regularization strategies could be highly beneficial.

The integration of attention mechanisms and the strategic use of prior knowledge are pivotal in advancing meta-learning techniques, achieving leading results in few-shot learning evaluations. Moreover, unsupervised approaches such as TrainPro-FSL reveal the potential to outperform traditional supervised methods, especially as evidenced by notable advancements on datasets like miniImageNet.

The strategic use of synthetic data emerges as a critical factor in reducing error rates for rare classes, emphasizing the importance of managing data quantity and variability to optimize model performance. This underscores the potential of synthetic data to enhance model robustness in few-shot learning contexts.

References

[1] Han-Jia Ye, Lu Han, and De-Chuan Zhan. Revisiting unsupervised meta-learning via the characteristics of few-shot tasks, 2022.

[2] Xueting Zhang, Debin Meng, Henry Gouk, and Timothy Hospedales. Shallow bayesian meta learning for real-world few-shot recognition, 2021.

[3] Yadan Luo, Zi Huang, Zheng Zhang, Ziwei Wang, Mahsa Baktashmotlagh, and Yang Yang. Learning from the past: Continual meta-learning via bayesian graph modeling, 2019.

[4] Yaqing Wang, Quanming Yao, James T Kwok, and Lionel M Ni. Generalizing from a few examples: A survey on few-shot learning. *ACM computing surveys (csur)*, 53(3):1–34, 2020.

[5] Zeynep Özdemir, Hacer Yalim Keles, and Ömer Özgür Tanrıöver. Meta-transfer derm-diagnosis: Exploring few-shot learning and transfer learning for skin disease classification in long-tail distribution, 2024.

[6] Wentao Chen, Chenyang Si, Wei Wang, Liang Wang, Zilei Wang, and Tieniu Tan. Few-shot learning with part discovery and augmentation from unlabeled images, 2021.

[7] Shaobo Lin, Xingyu Zeng, and Rui Zhao. Explore the power of dropout on few-shot learning, 2023.

[8] Reza Esfandiarpoor, Amy Pu, Mohsen Hajabdollahi, and Stephen H. Bach. Extended few-shot learning: Exploiting existing resources for novel tasks, 2021.

[9] Jaya Krishna Mandivarapu, Eric bunch, and Glenn fung. Domain agnostic few-shot learning for document intelligence, 2021.

[10] Archit Parnami and Minwoo Lee. Learning from few examples: A summary of approaches to few-shot learning, 2022.

[11] Xisen Jin, Bill Yuchen Lin, Mohammad Rostami, and Xiang Ren. Learn continually, generalize rapidly: Lifelong knowledge accumulation for few-shot learning, 2022.

[12] Alessia Bertugli, Stefano Vincenzi, Simone Calderara, and Andrea Passerini. Few-shot unsupervised continual learning through meta-examples, 2020.

[13] Ruohan Wang, Isak Falk, Massimiliano Pontil, and Carlo Ciliberto. Robust meta-representation learning via global label inference and classification, 2023.

[14] Brando Miranda, Patrick Yu, Yu-Xiong Wang, and Sanmi Koyejo. The curse of low task diversity: On the failure of transfer learning to outperform maml and their empirical equivalence, 2022.

[15] Shuzhen Rao, Jun Huang, and Zengming Tang. Exploiting style transfer-based task augmentation for cross-domain few-shot learning, 2023.

[16] Nihar Bendre, Hugo Terashima Marín, and Peyman Najafirad. Learning from few samples: A survey, 2020.

[17] Zhongqi Yue, Hanwang Zhang, Qianru Sun, and Xian-Sheng Hua. Interventional few-shot learning. *Advances in neural information processing systems*, 33:2734–2746, 2020.

[18] L earning from f ew e xamples: A.

[19] Gabriele Prato, Simon Guiroy, Ethan Caballero, Irina Rish, and Sarath Chandar. Scaling laws for the few-shot adaptation of pre-trained image classifiers, 2021.

[20] Huaiyu Li, Weiming Dong, Xing Mei, Chongyang Ma, Feiyue Huang, and Bao-Gang Hu. Lgm-net: Learning to generate matching networks for few-shot learning, 2019.

[21] Kevin Ellis. Human-like few-shot learning via bayesian reasoning over natural language, 2023.

[22] Deborah Pereg, Martin Villiger, Brett Bouma, and Polina Golland. Less is more: Rethinking few-shot learning and recurrent neural nets, 2023.

[23] Etienne Bennequin. Meta-learning algorithms for few-shot computer vision, 2019.

[24] Huimin Peng. A comprehensive overview and survey of recent advances in meta-learning, 2020.

[25] Amrith Setlur, Oscar Li, and Virginia Smith. Is support set diversity necessary for meta-learning?, 2021.

[26] Yaoyao Liu, Bernt Schiele, and Qianru Sun. An ensemble of epoch-wise empirical bayes for few-shot learning, 2020.

[27] Simon S. Du, Wei Hu, Sham M. Kakade, Jason D. Lee, and Qi Lei. Few-shot learning via learning the representation, provably, 2021.

[28] Xiaofeng Cao, Weixin Bu, Shengjun Huang, Minling Zhang, Ivor W. Tsang, Yew Soon Ong, and James T. Kwok. A survey of learning on small data: Generalization, optimization, and challenge, 2023.

[29] Flood Sung, Yongxin Yang, Li Zhang, Tao Xiang, Philip HS Torr, and Timothy M Hospedales. Learning to compare: Relation network for few-shot learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1199–1208, 2018.

[30] Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. *Advances in neural information processing systems*, 30, 2017.

[31] Guoliang Lin, Yongheng Xu, Hanjiang Lai, and Jian Yin. Revisiting few-shot learning from a causal perspective, 2024.

[32] Aravind Rajeswaran, Chelsea Finn, Sham Kakade, and Sergey Levine. Meta-learning with implicit gradients, 2019.

[33] Antreas Antoniou, Harrison Edwards, and Amos Storkey. How to train your maml, 2019.

[34] Chelsea Finn, Kelvin Xu, and Sergey Levine. Probabilistic model-agnostic meta-learning, 2019.

[35] Simon Guiroy, Christopher Pal, Gonçalo Mordido, and Sarath Chandar. Improving meta-learning generalization with activation-based early-stopping, 2022.

[36] Jialin Liu, Fei Chao, Longzhi Yang, Chih-Min Lin, and Qiang Shen. Decoder choice network for meta-learning, 2019.

[37] Hae Beom Lee, Taewook Nam, Eunho Yang, and Sung Ju Hwang. Meta dropout: Learning to perturb latent features for generalization. In *International Conference on Learning Representations*, 2020.

[38] Yong Wu, Shekhor Chanda, Mehrdad Hosseinzadeh, Zhi Liu, and Yang Wang. Few-shot learning of compact models via task-specific meta distillation, 2022.

[39] Mohammad Rostami, Atik Faysal, Huaxia Wang, and Avimanyu Sahoo. Meta-task: A method-agnostic framework for learning to regularize in few-shot learning, 2025.

[40] Lajanugen Logeswaran, Ann Lee, Myle Ott, Honglak Lee, Marc'Aurelio Ranzato, and Arthur Szlam. Few-shot sequence learning with transformers, 2020.

[41] Aaron Mueller, Kanika Narang, Lambert Mathias, Qifan Wang, and Hamed Firooz. Meta-training with demonstration retrieval for efficient few-shot learning, 2023.

[42] Lu Liu, Tianyi Zhou, Guodong Long, Jing Jiang, and Chengqi Zhang. Many-class few-shot learning on multi-granularity class hierarchy, 2020.

[43] Xiang Deng and Zhongfei Zhang. Is the meta-learning idea able to improve the generalization of deep neural networks on the standard supervised learning?, 2020.

[44] Christian Simon, Piotr Koniusz, Richard Nock, and Mehrtash Harandi. Adaptive subspaces for few-shot learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4136–4145, 2020.

[45] Jiaying Shi, Xuetong Xue, and Shenghui Xu. Learning to adapt category consistent meta-feature of clip for few-shot classification, 2024.

[46] Sachin Ravi and Hugo Larochelle. Optimization as a model for few-shot learning. In *International conference on learning representations*, 2017.

[47] Quentin Bouniot, Ievgen Redko, Romaric Audigier, Angélique Loesch, and Amaury Habrard. Improving few-shot learning through multi-task representation learning theory, 2022.

[48] Yisheng Song, Ting Wang, Puyu Cai, Subrota K Mondal, and Jyoti Prakash Sahoo. A comprehensive survey of few-shot learning: Evolution, applications, challenges, and opportunities. *ACM Computing Surveys*, 55(13s):1–40, 2023.

[49] Hongyang Li, David Eigen, Samuel Dodge, Matthew Zeiler, and Xiaogang Wang. Finding task-relevant features for few-shot learning by category traversal, 2019.

[50] Yunxiao Qin, Weiguo Zhang, Chenxu Zhao, Zezheng Wang, Xiangyu Zhu, Guojun Qi, Jingping Shi, and Zhen Lei. Prior-knowledge and attention-based meta-learning for few-shot learning, 2020.

[51] Yinbo Chen, Zhuang Liu, Huijuan Xu, Trevor Darrell, and Xiaolong Wang. Meta-baseline: Exploring simple meta-learning for few-shot learning. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 9062–9071, 2021.

[52] Yaohui Li, Qifeng Zhou, Haoxing Chen, Jianbing Zhang, Xinyu Dai, and Hao Zhou. The devil is in the few shots: Iterative visual knowledge completion for few-shot learning, 2024.

[53] Mohammad Reza Zarei and Majid Komeili. Interpretable concept-based prototypical networks for few-shot learning, 2022.

[54] Jun Shern Chan, Michael Pieler, Jonathan Jao, Jérémy Scheurer, and Ethan Perez. Few-shot adaptation works with unpredictable data, 2022.

[55] Bingzhi Chen, Haoming Zhou, Yishu Liu, Biqing Zeng, Jiahui Pan, and Guangming Lu. Enhancing few-shot classification without forgetting through multi-level contrastive constraints, 2024.

[56] Sungyong Baik, Seokil Hong, and Kyoung Mu Lee. Learning to forget for meta-learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2379–2387, 2020.

[57] Kaidi Cao, Maria Brbic, and Jure Leskovec. Concept learners for few-shot learning, 2021.

[58] Aroof Aimen, Sahil Sidheekh, Vineet Madan, and Narayanan C. Krishnan. Stress testing of meta-learning approaches for few-shot learning, 2021.

[59] Modular meta-learning.

[60] Jaewoong Shin, Hae Beom Lee, Boqing Gong, and Sung Ju Hwang. Large-scale meta-learning with continual trajectory shifting, 2022.

[61] Runxi Cheng, Yongxian Wei, Xianglong He, Wanyun Zhu, Songsong Huang, Fei Richard Yu, Fei Ma, and Chun Yuan. Learn to learn more precisely, 2024.

[62] Joaquin Vanschoren. Meta-learning: A survey. *arXiv preprint arXiv:1810.03548*, 2018.

[63] Chi Zhang, Henghui Ding, Guosheng Lin, Ruibo Li, Changhu Wang, and Chunhua Shen. Meta navigator: Search for a good adaptation policy for few-shot learning, 2021.

[64] Hassan Gharoun, Fereshteh Momenifar, Fang Chen, and Amir H. Gandomi. Meta-learning approaches for few-shot learning: A survey of recent advances, 2023.

[65] Wei-Lun Chao, Han-Jia Ye, De-Chuan Zhan, Mark Campbell, and Kilian Q Weinberger. Revisiting meta-learning as supervised learning. *arXiv preprint arXiv:2002.00573*, 2020.

[66] Timothy Hospedales, Antreas Antoniou, Paul Micaelli, and Amos Storkey. Meta-learning in neural networks: A survey, 2020.

[67] Eugene Lee, Cheng-Han Huang, and Chen-Yi Lee. Few-shot and continual learning with attentive independent mechanisms, 2021.

[68] Joshua Ball. Few-shot learning for image classification of common flora, 2021.

[69] Wenpeng Yin. Meta-learning for few-shot natural language processing: A survey. *arXiv preprint arXiv:2007.09604*, 2020.

[70] Junying Huang, Fan Chen, Keze Wang, Liang Lin, and Dongyu Zhang. Enhancing prototypical few-shot learning by leveraging the local-level strategy, 2021.

[71] Nathaniel Simard and Guillaume Lagrange. Improving few-shot learning with auxiliary self-supervised pretext tasks, 2021.

[72] Vincent Dumoulin, Neil Houlsby, Utku Evci, Xiaohua Zhai, Ross Goroshin, Sylvain Gelly, and Hugo Larochelle. Comparing transfer and meta learning approaches on a unified few-shot classification benchmark, 2021.

[73] Qianru Sun, Yaoyao Liu, Zhaozheng Chen, Tat-Seng Chua, and Bernt Schiele. Meta-transfer learning through hard tasks, 2019.

[74] Raphael Baena, Lucas Drumetz, and Vincent Gripon. On transfer in classification: How well do subsets of classes generalize?, 2024.

[75] Zhiqiang Shen, Zechun Liu, Jie Qin, Marios Savvides, and Kwang-Ting Cheng. Partial is better than all: Revisiting fine-tuning strategy for few-shot learning, 2021.

[76] John Cai and Sheng Mei Shen. Cross-domain few-shot learning with meta fine-tuning, 2020.

[77] Yujin Kim, Jaehoon Oh, Sungnyun Kim, and Se-Young Yun. How to fine-tune models with few samples: Update, data augmentation, and test-time augmentation, 2022.

[78] Zhongjie Yu, Lin Chen, Zhongwei Cheng, and Jiebo Luo. Transmatch: A transfer-learning scheme for semi-supervised few-shot learning, 2020.

[79] Shengyu Feng and Hanghang Tong. Concept discovery for fast adapatation, 2023.

[80] Tasmia Shahriar, Kelly Ramos, and Noboru Matsuda. Assertion enhanced few-shot learning: Instructive technique for large language models to generate educational explanations, 2024.

[81] Yulin Zhou, Yiren Zhao, Ilia Shumailov, Robert Mullins, and Yarin Gal. Revisiting automated prompting: Are we actually doing better?, 2023.

[82] Guangyuan Shi, Jiaxin Chen, Wenlong Zhang, Li-Ming Zhan, and Xiao-Ming Wu. Overcoming catastrophic forgetting in incremental few-shot learning by finding flat minima, 2021.

[83] Cuiwei Liu, Siang Xu, Huaijun Qiu, Jing Zhang, Zhi Liu, and Liang Zhao. Few-shot class-incremental learning with non-iid decentralized data, 2024.

[84] Xiang Xiang, Yuwen Tan, Qian Wan, and Jing Ma. Coarse-to-fine incremental few-shot learning, 2021.

[85] Mayank Lunayach, James Smith, and Zsolt Kira. Lifelong wandering: A realistic few-shot online continual learning setting, 2022.

[86] Andrei Boiarov, Oleg Granichin, and Olga Granichina. Simultaneous perturbation stochastic approximation for few-shot learning, 2020.

[87] Branislav Pecher, Ivan Srba, and Maria Bielikova. A survey on stability of learning with limited labelled data and its sensitivity to the effects of randomness, 2024.

[88] Sara Beery, Yang Liu, Dan Morris, Jim Piavis, Ashish Kapoor, Markus Meister, Neel Joshi, and Pietro Perona. Synthetic examples improve generalization for rare classes, 2019.

[89] Ghassen Jerfel, Erin Grant, Thomas L. Griffiths, and Katherine Heller. Reconciling meta-learning and continual learning with online mixtures of tasks, 2019.

[90] Zilun Zhang, Shihao Ma, and Yichun Zhang. Will multi-modal data improves few-shot learning?, 2021.

[91] Simon S Du, Wei Hu, Sham M Kakade, Jason D Lee, and Qi Lei. Few-shot learning via learning the representation, provably. *arXiv preprint arXiv:2002.09434*, 2020.

[92] Luísa Shimabucoro, Timothy Hospedales, and Henry Gouk. Evaluating the evaluators: Are current few-shot learning benchmarks fit for purpose?, 2023.

Disclaimer:

SurveyX is an AI-powered system designed to automate the generation of surveys. While it aims to produce high-quality, coherent, and comprehensive surveys with accurate citations, the final output is derived from the AI's synthesis of pre-processed materials, which may contain limitations or inaccuracies. As such, the generated content should not be used for academic publication or formal submissions and must be independently reviewed and verified. The developers of SurveyX do not assume responsibility for any errors or consequences arising from the use of the generated surveys.