



Universiteit
Leiden
The Netherlands

Exploring deep learning for data-efficient and generalizable visual recognition

He, K.

Citation

He, K. (2026, September 22). *Exploring deep learning for data-efficient and generalizable visual recognition*. Retrieved from <https://hdl.handle.net/1887/4310276>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/4310276>

Note: To cite this publication please use the final published version (if applicable).

Chapter 1

Introduction

1.1 Motivation

With increasing attention toward social safety and security, comprehensive visual surveillance infrastructures are deployed in public areas, e.g., transportation hubs, healthcare facilities and retail centers. On the one hand, manually processing extensive visual data generated by these systems is time consuming and will bring severe economic burdens. On the other hand, in some high-stakes areas, such as autonomous driving and rare disease diagnosis, the availability of labeled training samples is often limited, making conventional supervised learning approaches impractical. Hence, there is an urgent need for few-shot learning (FSL), which aims to generalize and adapt models to novel categories based on a handful of annotated instances. As illustrated in Figure 1.1, a standard FSL pipeline typically includes: 1) Collect and preprocess a labeled dataset of base classes; 2) Train a feature extractor on abundant base-class samples; 3) Construct episodes that mimic few-shot scenarios and train the meta-learner or metric head under few-shot constraints; 4) Given novel classes with only a few labeled support samples, adapt the model using metric-based matching or optimization-based tuning; 5) Evaluate query images from unseen classes to measure model adaptability and robustness. This thesis places particular emphasis on Components 2, 3 and 4.

Thanks to the rapid development of deep learning, FSL has achieved remarkable progress in recent years. Current FSL approaches have demonstrated impressive performance on standard benchmarks, where models are evaluated on novel classes given only a handful of labeled examples [1, 2]. In some constrained cases, the performance of these FSL methods even surpass human-level recognition accuracy [3, 4]. However, in real-world scenarios, data distributions are diverse, which is a big challenge for models' generalization ability and robustness. In this thesis, we aim to bridge the gap between standard benchmarks and real-world scenarios, and improve FSL models' performance by exploring three core research themes: metric learning, representation enhancement and knowledge transfer.

In Figure 1.2 the three themes are conceptually depicted.

- **Metric Learning** plays an essential role in machine learning systems, which aims to learn a highly discriminative embedding space so that the distances between similar samples are as close as possible and the distances between different categories are as large as possible. In general, metric learning significantly improves the performance in classification, clustering, and retrieval tasks. Inspired by this, we explore the similarities between class prototypes for FSL.
- **Feature Enrichment** focuses on refining learned features, boosting models' generalization capabilities and robustness. Through techniques such as attention mechanisms or multi-scale fusion, the enhanced representations help the

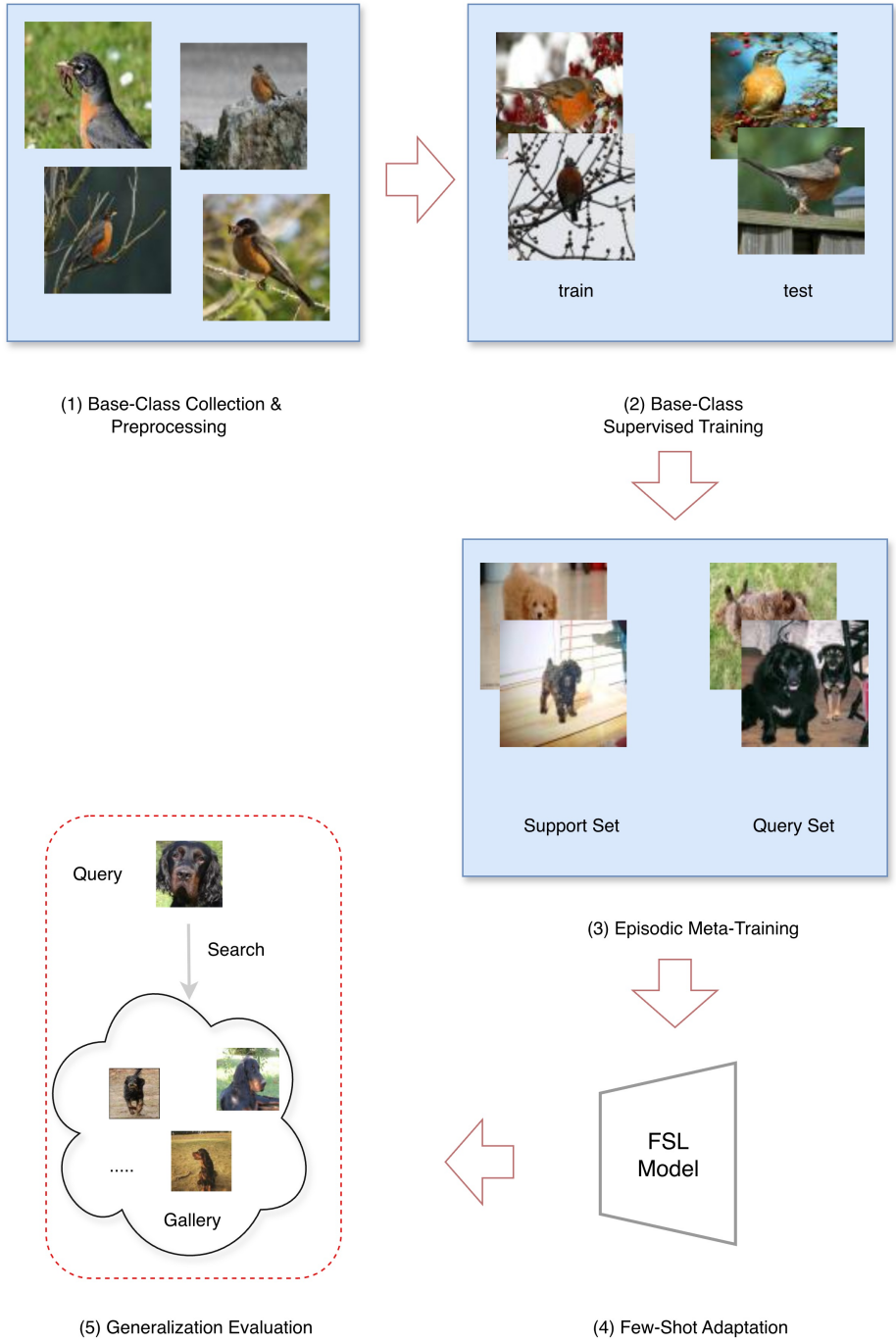


Figure 1.1: Five main stages in the pipeline of FSL.

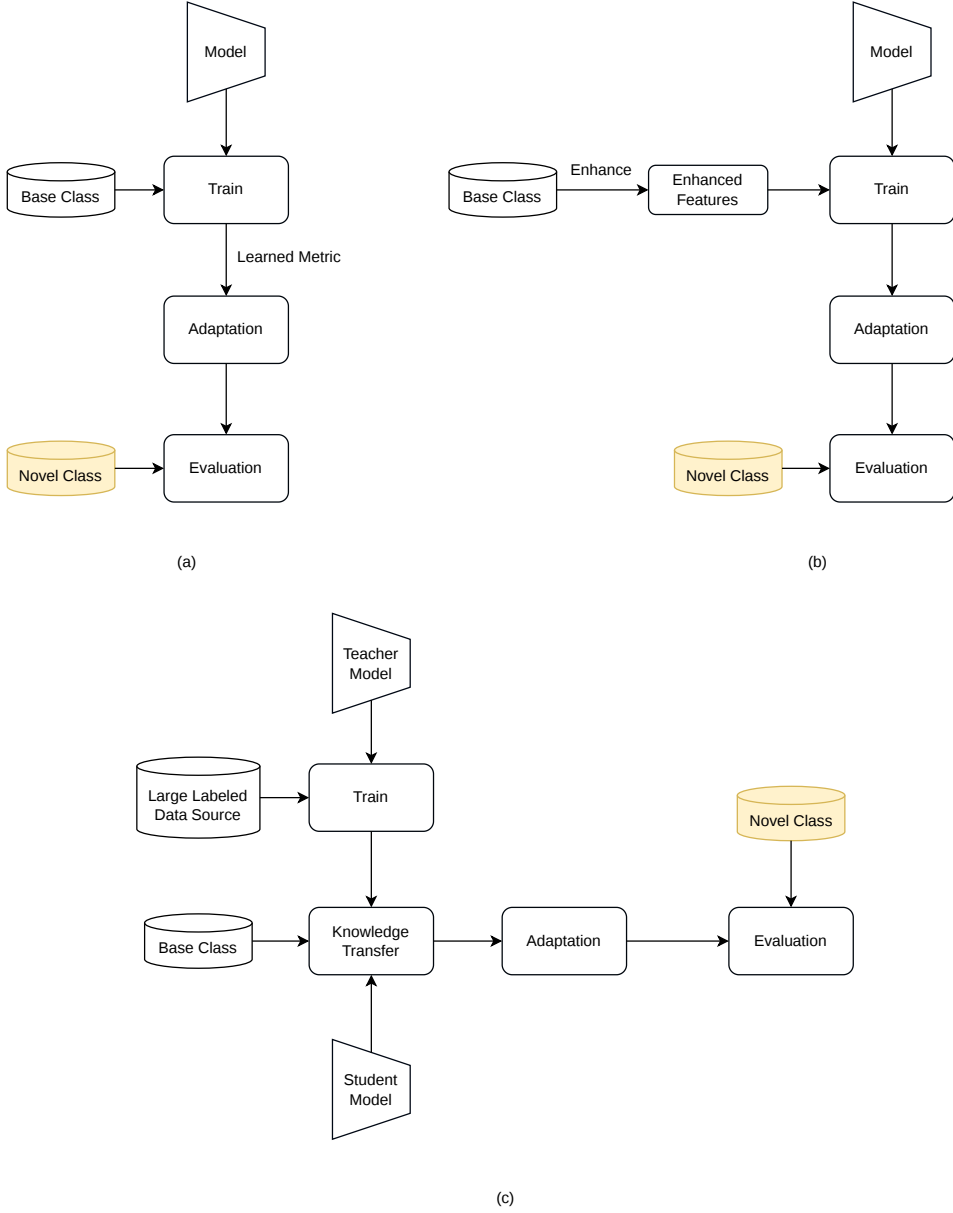


Figure 1.2: Conceptual illustration of the three research themes in this thesis, including (a) metric learning, (b) feature enrichment and (c) knowledge transfer.

model better capture subtle but discriminative characteristics, strengthening the model’s performance under challenging real-world conditions.

- **Knowledge Transfer** addresses the challenge of leveraging previously acquired knowledge to improve the model’s performances in new tasks. It seeks

to bridge the gap between domains by facilitating knowledge sharing, which is particularly effective when labeled data are scarce or expensive to collect. To the best of our knowledge, our work is the first study to explore the dual selective knowledge transfer framework for FSL, which aims to learn effective knowledge from both the teacher model and current samples

1.2 Background and Related Work

As one of the important machine learning applications, few-shot learning (FSL) has attracted extensive attention as a pivotal research direction of learning new concepts from very limited examples. Given a handful of labeled samples—the support set—for each new class, the goal of FSL is to build a model with good robustness and generalization ability by leveraging prior knowledge from related tasks.

In order to comprehensively explore FSL capabilities, over the past decades, a wide variety of benchmark datasets specific for FSL tasks have been proposed. These datasets cover different domains (e.g., natural images, medical scans, buildings) and vary in scale, complexity, and data distribution. The diversity among tasks and domains leads to enormous generalization challenges, often referred to as distribution or domain gaps. A detailed overview of these FSL benchmarks and their characteristics has been presented in the recent literature [5].

As discussed in the survey proposed by Song *et al.* [6], recent FSL researches have rapidly evolved beyond conventional meta-learning setups and shifted toward more realistic and open-world scenarios. This transition emphasizes the need for adaptable and scalable solutions, for example, 1) enhancing discriminability in low-data scenarios by using metric learning to optimize feature embeddings; 2) the use of representation enhancement to improve feature robustness and to mitigate overfitting; 3) the use of knowledge transfer to boost the model’s performance on new tasks by leveraging prior learned knowledge.

1.2.1 Metric Learning

Human beings inherently distinguish similarities and differences among observed instances, and effectively categorizing these objects based on these measurements. This innate ability significantly contributes to our ability for rapid learning and adaptation to new tasks. Analogously, machine learning algorithms also explore systematic strategies to evaluate similarities among samples, which is crucial to classification and retrieval tasks. Unlike traditional supervised learning methods that are trained on extensive datasets, metric learning designs specialized distance functions to classify samples effectively, which is especially valuable in scenarios where only limited labeled samples are available, such as FSL.

Metric learning has also been named similarity learning [7] and deep metric learning [8] in previous works. Existing methods focus on computer vision tasks like image classification [9], object detection [10] and person re-identification [11]. Metric learning requires the model to find an embedding space where similar instances are clustered together. The key challenge in metric Learning, especially within FSL scenarios, is accurately defining and learning meaningful measures from limited samples to ensure good generalization to novel classes. However, conventional metric learning methods are not suitable for FSL for various reasons:

- The number of training samples per class in FSL is much lower than that in conventional metric learning tasks, e.g., popular benchmarks such as CIFAR-100 [12] and ImageNet [13], which provides sufficient data for the model’s training. In contrast, benchmark for FSL scenarios like miniImageNet [14] or Omniglot [3] often provides as few as one to five examples per class, making it challenging for classification tasks on these new classes, as illustrated in Figure 1.3.
- FSL datasets are more imbalanced [15, 16, 17], that means there exist significant intra-class differences and subtle inter-class differences, which makes it difficult to learn reliable metrics that can effectively distinguish categories.
- Moreover, in real-world FSL scenarios, the training (base) classes and the testing (novel) classes follow different data distributions. This increases the difficulty of learning a transferable metric.

To address the above-mentioned difficult yet practical challenges, we propose a novel prototype similarity framework. In contrast to prior works that treat feature extraction, metric refinement, and regularization as independent components, our approach integrates a cosine sharpening transformation with elastic regularization into a unified representation learning framework. This joint design encourages the model to simultaneously refine decision margins, preserve the geometric consistency of the embedding manifold, and enhance robustness to subtle inter-class similarities. Extensive experiments on FSL benchmarks demonstrate that our approach effectively improves generalization to unseen categories and achieves state-of-the-art performance (Chapter 3).

1.2.2 Feature Enrichment

In order to adapt the model to novel classes, extracting discriminative features from the limited number of labeled samples plays an important role in FSL scenarios. To resolve the limitation of insufficient representation caused by scarce training samples, feature enrichment has been exploited to address these practical challenges. In recent FSL works, representation enhancement has been shown effective to extracting discriminative features from limited data, leading to better FSL performance

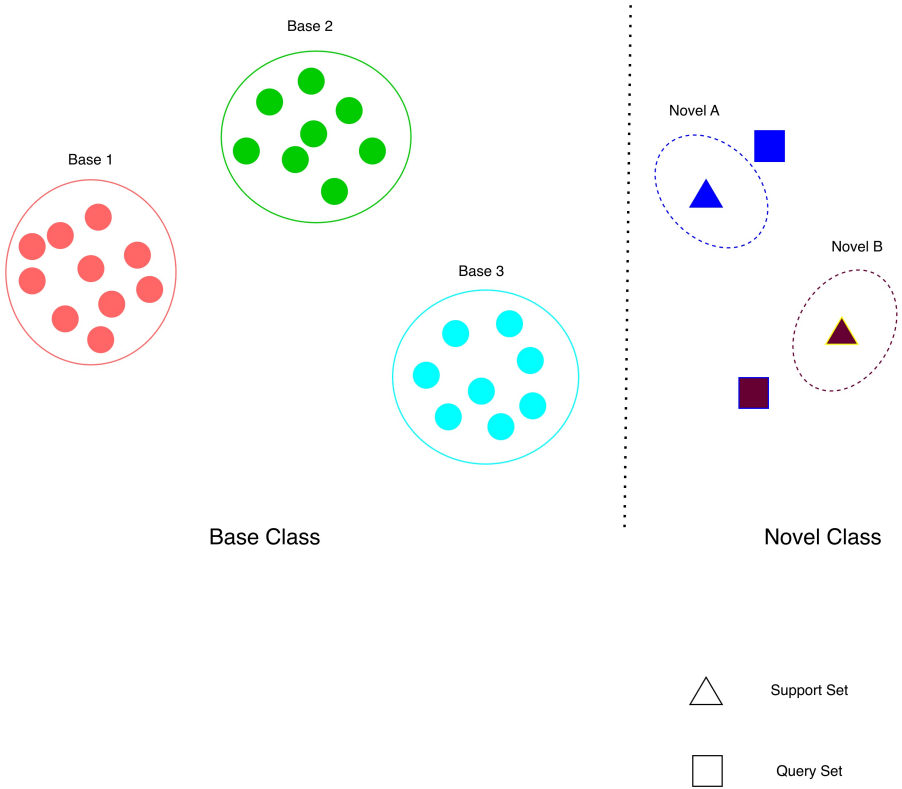


Figure 1.3: Metric learning challenge in FSL: for novel classes, the support set often consists of only 1-5 examples

Driven by the development of deep learning, current representation enhancement methods in FSL can be generally divided into two main branches. The first branch explores self-supervised learning to pretrain models on auxiliary tasks without manual annotations, thereby learning generalized and robust representations. For example, Chen *et al.* [18] proposed a contrastive learning framework that maximizes the agreement between augmented instances of the same image, while Lu *et al.* [19] demonstrated that self-supervised pretraining can outperform supervised pretraining in few-shot scenarios under appropriate conditions, such as deeper backbones, larger image sizes, and stronger data augmentations. Similarly, Gidaris *et al.* [20] introduced a rotation prediction task to learn semantic features that are applicable to FSL. The second branch focuses on feature refinement through attention mechanisms and meta-learning, effectively enhancing the discriminativeness of embeddings for FSL tasks. Tseng *et al.* [21] proposed a feature-wise transformation layer to simulate various feature distributions during meta-training, improving the model’s generalization ability to novel classes. Meanwhile, Hou *et al.* [22] introduced a cross-attention

module to dynamically highlight task-relevant features, reducing the interference by irrelevant background information. Recent work proposed by Oreshkin *et al.* [23] further demonstrated that task-adaptive feature modulation can significantly boost few-shot performance by aligning representations with task-specific requirements. Despite these notable advancements, existing methods still face two main challenges in FSL: 1) Feature overfitting, where enhanced representations generalize poorly to novel classes due to the scarcity of training samples; 2) Domain shift, where representations learned from base classes fail to adapt to novel classes following different data distributions. To mitigate these issues, recent studies have explored hybrid approaches that combine representation enhancement with metric learning [24] or data augmentation [25]. For example, Li *et al.* [26] propose adversarial feature hallucination networks to synthesize diverse training samples, while Wang *et al.* [27] employ information fusion rectification to rectify the distribution of the support set in novel classes.

Different from previous works that mainly focus on external data augmentation or simple feature aggregation strategies, we analyze the limitations that arise when few-shot models are forced to learn discriminative representations from extremely limited data. To address these limitations, we introduce an attention-guided feature pyramid network that explicitly enhances feature richness by integrating complementary information from multiple backbone layers through a learnable top-down fusion mechanism (Chapter 4). This design strengthens the discriminativeness of few-shot embeddings by preserving both local details and global semantic structure. In addition, we develop an adaptive multi-scale feature fusion framework that further enriches representations by expanding multi-branch feature diversity and dynamically re-weighting cross-scale responses according to their task relevance (Chapter 5). Together, these two directions form a comprehensive approach to feature enrichment in FSL, allowing models to form more discriminative, richer, and more transferable embeddings with very limited supervision.

1.2.3 Knowledge Transfer

The rapid development of deep neural networks has obtained advanced performance in many vision tasks. Meanwhile, it is still challenging to train models that can quickly adapt to new tasks with limited labeled examples. Knowledge transfer, aiming to leverage knowledge acquired from previous learned tasks or abundant labeled data to boost performance on new tasks, has emerged as a crucial paradigm in FSL. Given a handful of labeled samples in a target task, knowledge transfer seeks to utilize transferable knowledge from auxiliary domains or tasks to learn robust representations for novel tasks. The key challenge is how to filter useful knowledge, as presented in Figure 1.4.

Existing knowledge transfer approaches applied in FSL can be broadly divided into two main groups. The first group focuses on parameter transfer, where knowledge

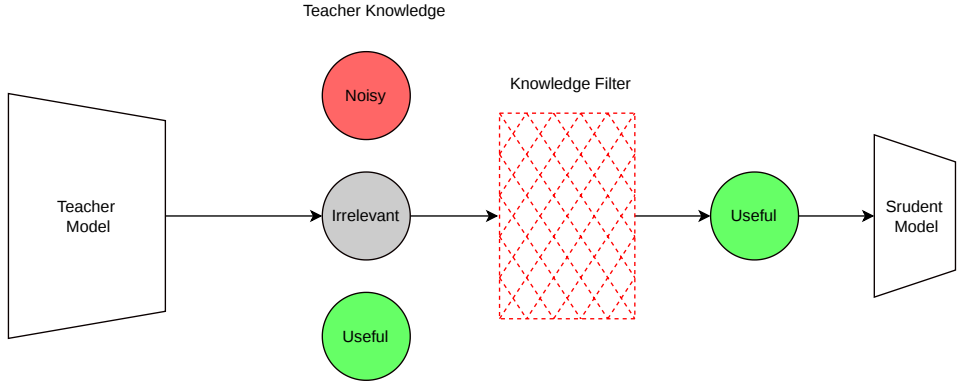


Figure 1.4: Knowledge Transfer challenge.

is transferred through the weights of pre-trained models. These approaches mainly employ meta-learning frameworks to learn good initialization parameters that can quickly adapt to new tasks. For example, Finn *et al.* [2] proposed Model-Agnostic Meta-Learning (MAML), which learns model parameters that can be fine-tuned with a small number of gradient steps on new tasks. Antoniou *et al.* [28] extended this idea by incorporating a meta-learning optimizer to improve the adaptation efficiency. Despite their effectiveness, these methods often require to tune the adaptation process carefully and may suffer from overfitting when there is a large domain gap between base and novel classes. In contrast, the second group [12,20,21,22,23] emphasizes feature-level knowledge transfer, which attempts to learn transferable representations that generalize well to novel classes. Jamal *et al.* [29] proposed a task-agnostic meta-learning approach that explicitly separates knowledge transfer from task-specific adaptation, enabling more effective cross-task generalization. To address severe domain shift in cross-domain few-shot learning, Paeedeh *et al.* [30] introduce ADAPTER, which uses a transformer with bidirectional cross-attention to learn transferable features between base-domain and target-domain representations.

Despite these advances, knowledge transfer in FSL still faces several key challenges: The semantic gap between base and novel classes may lead to negative knowledge transfer when the knowledge is not properly aligned; Limited supervision in novel tasks makes it difficult to preserve the discriminative power of transferred knowledge; The diversity of potential novel tasks requires the transfer mechanism to be flexible and adaptive. To address these challenges, recent works have explored more sophisticated knowledge transfer strategies. Cheraghian *et al.* [31] propose a semantic-aware knowledge distillation framework for few-shot class-incremental learning, which leverages word embeddings and an attention mechanism over multiple visual embeddings to align visual and semantic representations and mitigate

catastrophic forgetting. Chen *et al.* [32] develop a knowledge graph transfer network to explicitly model relationships between base and novel classes.

Different from previous approaches, we introduce a dual selective knowledge transfer (DSKT) framework that explicitly preserves the most transferable components of the teacher’s representation while suppressing task-irrelevant or noisy signals, as detailed in Chapter 6. DSKT performs selective distillation from both the perspective of global semantic knowledge and the task-specific representations extracted from the limited support samples. By selectively integrating the most beneficial components from both sources, DSKT enables the student model to maintain a more stable and discriminative embedding space when learning new categories with minimal annotated data.

1.3 Thesis Outline and Research Questions

In Section 1.2, we have briefly reviewed the recent progress from three research directions: 1) metric learning; 2) feature enrichment and 3) knowledge transfer. Few-shot learning has become an increasingly significant topic in the image understanding area. For a better understanding of FSL, we first present a comprehensive review of existing researches related to FSL. Our review includes the main challenges, benchmarks and strategies specialized for model robustness in FSL tasks. Chapter 2 is based on the published journal paper [33]:

He, K., Pu, N., Lao, M., and Lew, M.S. “Few-shot and meta-learning methods for image understanding: a survey.” *International Journal of Multimedia Information Retrieval*, 2023.

Although deep learning has driven state-of-the-art results in many FSL benchmarks, it still faces many limitations and challenges, such as sensitivity to label scarcity, reliance on large-scale pre-training, limited interpretability of learned metrics, and difficulties in generalizing to highly diverse or shifting task distributions. There remains considerable space to further improve the flexibility, robustness, and theoretical understanding of FSL models. In the following research chapters (Chapters 3–6), we present novel approaches to address the research questions (RQs) and challenges in terms of three research themes, including more discriminative metric formulations, enriched feature representations under data scarcity, and more effective knowledge transfer strategies. In Chapter 7, we summarize our main findings, limitations, potential solutions, and promising directions for future research.

- In Chapter 3, we conduct an in-depth investigation into the fundamental limitations of current metric learning frameworks in FSL scenarios. Through extensive empirical analysis, we find two primary reasons: (1) Most existing

methods employ static metric spaces that fail to dynamically accentuate discriminative features when presented with novel categories. (2) The lack of training samples make the model prone to overfitting and unstable calibration across episodes. These findings lead to our first research question **RQ 1: How to design a unified framework that simultaneously enhances metric discriminability and ensures training stability?** To address this challenge, we introduce a novel approach that strategically combines cosine sharpening transformation with elastic regularization mechanism. The cosine sharpening transformation operates as an adaptive feature recalibration module that dynamically sharpens angular decision boundaries between similar categories, effectively amplifying subtle discriminative signals that are crucial for few-shot identification. Concurrently, our elastic regularization framework provides multi-scale constraint adaptation, allowing the model to maintain representation stability across episodes while preserving the flexibility necessary for rapid task-specific adaptation. This dual mechanism enables our framework to achieve superior performance on both initial adaptation and cross-episode generalization, effectively bridging the gap between metric sensitivity and training stability. This work is based on the accepted paper:

He, K., Pu, N., Bakker, E. M., and Lew, M. S. “ProtoSimNet: Exploiting Prototype Similarity for Few-shot Learning.” accepted by the Neurocomputing journal, 2026.

- In order to enhance the model’s capability for rapid adaptation under limited supervision, our investigations in Chapter 2 and 3 primarily focus on improving feature robustness through metric-based and meta-learning paradigms. In Chapter 4, however, as we further explored the limitations of existing FSL techniques, we observed that many conventional strategies—especially those relying heavily on single-scale representations—tend to impose overly rigid constraints on feature adaptation. Such constraints restrict the model’s adaptation to the diverse scale distributions across different episodes. This observation leads us to the second research question **RQ 2: How to develop a flexible multi-scale representation framework that can dynamically reconcile feature hierarchies while preserving discriminative capability under limited supervision?** Given the inherent multi-scale nature of visual recognition tasks, we propose to exploit an attention-guided feature pyramid construction as the fundamental architecture. However, traditional feature pyramid networks typically rely on fixed fusion strategies and extensive parameterization, which are particularly prone to overfitting in few-shot tasks. To address these challenges, we propose a novel attention-guided feature pyramid network (AFPN) that integrates two innovative components: a simple,

parameter-free attention module (SimAM) for efficient channel-level feature refinement, and a cross-layer spatial attention mechanism (CSAM) for adaptive multi-scale feature fusion. This integrated approach enables our framework to maintain architectural simplicity while achieving remarkable adaptability across varying scale distributions and domain characteristics. This work is based on the published conference paper [34]:

He, K., Bakker, E. M., and Lew, M. S. “Attention-guided Feature Pyramid Network for few-shot learning.” Benelux Conference on Artificial Intelligence and Machine Learning, 2024.

- In Chapter 5, we aim to investigate the role of multi-scale feature representation learning in addressing the limitations of few-shot classification by answering the third research question **RQ 3: How can we effectively enhance feature richness and discriminativeness under extremely data scarcity?** As discussed in Section 1.2.1, most contemporary FSL approaches mainly focus on either metric optimization or meta-learning strategies. While these methods have demonstrated considerable successes in standard benchmark scenarios, they often overlook the fundamental challenge of insufficient feature diversity caused by limited training samples, particularly when dealing with fine-grained categories or significant intra-class variations. To address this gap, we introduce a novel Adaptive Multi-scale Feature Fusion (AMFF) framework specifically designed for FSL. AMFF integrates two key components: a Feature Enrichment Module (FEM) that expands feature diversity via multi-branch convolutional extraction, and a Feature Fusion Module (FFM) that adaptively reweights and aggregates multi-scale representations to enhance discriminative capability. These mechanisms allow the model to learn more diverse visual representations and stronger category-discriminative information from only a few labeled samples. This chapter is based on the published conference paper [35]:

He, K., Bakker, E. M., and Lew, M. S. “Adaptive Multi-scale Feature Fusion Network for Few-Shot Learning.” International Conference on Intelligent Computing, 2025.

- Although enhancing representation learning can improve a model’s ability to recognize new categories from limited data, conventional feature extractors often fail to obtain good generalizable feature representations for FSL as the representations learned from limited samples may not adequately capture the diversity that required to distinguish novel classes. In Chapter 6, we therefore introduce our fourth research question **RQ4: How can transferable**

knowledge be selectively leveraged to build more generalizable feature representations for few-shot tasks? Motivated by the effectiveness of teacher-student learning paradigms in transferring learned knowledge, we believe that a well-designed selective transfer mechanism can enable models to learn more robust and discriminative features. To this end, we propose a Dual Selective Knowledge Transfer (DSKT) framework that consists of two stages. First, we employ a multi-task pre-training methodology to develop a teacher model with robust, transferable feature representations; Second, we implement a dual selective knowledge distillation mechanism that enables the student model to intelligently integrate knowledge from both the teacher model and current task-specific samples. This selective knowledge transfer effectively strengthens the student model’s capability to generalize to unseen categories. This chapter is based on the published journal paper [36]:

He, K., Pu, N., Lao, M., Bakker, E. M., and Lew, M.S. “Dual selective knowledge transfer for few-shot classification.” *Applied Intelligence*, 2023.

- Finally, Chapter 7 summarizes the main findings from the research of this thesis. Also, we discuss the limitations and potential solutions, and point out directions for future research.

Additionally, this thesis draws on insights and experiences from the related work in other publications during my PhD studies:

Lao, M., Pu, N., Liu, Y., **He K.**, Bakker, E. M., and Lew, M. S. “COCA: COllaborative CAusal Regularization for Audio-Visual Question Answering.” *AAAI Conference on Artificial Intelligence*, 2023.

1.4 Main Contributions

The contributions of this thesis can be summarized and characterized as: 1) Task Contributions; and 2) Technical Contributions.

1.4.1 Task Contributions

A new and challenging FSL task. We introduce a new and challenging FSL task that specially emphasizes recognition under multi-scale and structural variations, which is often overlooked in existing FSL benchmarks. Unlike common FSL settings which assume relatively stable object scales, we require models to generalize to novel categories whose visual characteristics may shift dramatically across spatial resolutions. This setting better reflects real-world scenarios, where class instances

frequently face scale fluctuations, local distortions, and fine-grained discriminative differences, yet only a few labeled examples are available. Different from conventional FSL tasks that are conducted on fixed-resolution features, our setup gives rise to several additional challenges:

1. Large intra-class scale variation disrupts the stability of feature representations learned from scarce samples.
2. Subtle differences between classes require models to extract fine-grained local features, where conventional single-scale baselines are unable to capture effective representations..
3. Imbalanced scale distributions across categories exacerbate data scarcity in FSL.

To achieve a comprehensive analysis of these challenges, we develop a task formulation designed to assess FSL models under substantial multi-scale variations—a factor that is typically underrepresented in existing benchmarks. Conventional few-shot settings tend to normalize object scales, either implicitly through dataset bias or explicitly through preprocessing, which fails to address the fundamental difficulties introduced by scale inconsistencies. In contrast, our formulation treats multi-scale variability as a core aspect of the problem and examines how models deal with diverse feature resolutions when supervision is extremely limited. This setting allows us to further explore several questions: how a model can combine information from different hierarchical feature levels, to what extent scale shifts degrade generalization to novel classes. Overall, this task aims to encourage the development of more adaptive and scale-aware FSL models.

1.4.2 Technical Contributions

From Chapter 3 to Chapter 6, we develop new approaches based on deep learning to address the research questions regarding the three themes. The key contributions in these approaches are listed below.

A novel metric learning method for FSL. Previous conventional metric spaces for FSL often fail to form clear decision boundaries when only a few training examples are available. In addition to the intrinsic data scarcity, FSL also faces challenges arising from high feature similarity between different categories and limited separability in low-dimensional embeddings. These limitations make metric-based methods difficult to obtain good performance in fine-grained classification tasks. To address this issues, we propose a novel prototype similarity framework for FSL tasks. To map these samples into a more discriminative embedding space, our framework applies a cosine sharpening transformation to enlarge inter-class margins during episodic training. Furthermore, an elastic regularization component is integrated to prevent overfitting due to limited support examples. The numerous experiments demonstrate the effectiveness of our framework in FSL tasks.

An effective attention-guided pyramid fusion method for FSL. FSL endures challenges due to the extremely limited amount of annotated data, making it difficult for models to form discriminative representations-especially when the visual characteristics that distinguish categories arise from different spatial scales. In fine-grained FSL scenarios, a single global (image-level) embedding can be insufficient: it tends to suppress subtle local cues, making class boundaries ambiguous [37]. To mitigate these issues, we propose an attention-guided feature pyramid network (AFPN) that explicitly leverages multi-level feature hierarchies to construct robust few-shot representations. AFPN employs a learnable attention mechanism to regulate the contribution of features originating from different depths of the backbone and to highlight structural patterns that are critical for class discrimination under few-shot supervision. A top-down fusion process then aggregates these heterogeneous representations into a coherent feature space with enhanced discriminability. The experimental results demonstrate AFPN significantly improves generalization to novel categories, and achieves a new state-of-the-art performance.

A new multi-scale feature fusion framework tailor-made for FSL. Unlike traditional recognition tasks where abundant labeled data enables models to easily obtain rich feature diversity, FSL tasks are conducted under severe data scarcity, making it difficult to learn discriminative and robust representations. Beyond the typical challenges caused by intra-class variability or fine-grained visual differences, few-shot models also face an additional limitation caused by insufficient feature diversity, making it particularly difficult for models to capture comprehensive visual patterns. To address these challenges, we propose an adaptive multi-scale feature fusion (AMFF) framework designed to systematically enrich and refine feature representations learned from only a handful of labeled examples. Specifically, we introduce a feature enrichment module (FEM) that leverages a multi-branch convolutional design to expand feature diversity by capturing complementary representations. Complementing this, a feature fusion module (FFM) adaptively recalibrates and aggregates the extracted multi-scale features with learned importance weights, enabling the network to emphasize scale-consistent and discriminative information. The experimental validation demonstrates that both components of our framework contribute significantly to its overall performance,

A new knowledge distillation method for FSL. Through an in-depth analysis of why conventional few-shot learning models struggle to generalize, we observe that feature extractors trained under limited supervision often fail to obtain robust representations and to transfer effectively to unseen classes. This issue becomes more serious when novel categories exhibit fine-grained variations that cannot be inferred from only a few labeled samples. Motivated by this limitation, we propose a dual selective knowledge transfer (DSKT) framework specifically designed to rethink the knowledge distillation process for few-shot scenarios. DSKT introduces a two-stage learning paradigm that explicitly separates the acquisition of broad visual knowledge from the task-specific adaptation process. Benefiting from its selective

1. INTRODUCTION

transfer strategy, DSKT effectively alleviates the limitations between preserving generalized visual understanding and adapting to new categories with very limited data. The numerous experiments demonstrate the effectiveness of our DSKT in FSL tasks.