



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).

English Summary

We are living in an era of artificial intelligence where the demand for rapid adaptation to new visual concepts grows exponentially. It is crucial to develop intelligent visual recognition systems that can learn efficiently from limited examples. A particularly important capability, given the expanding need for customized visual applications, is few-shot learning (FSL). A FSL system aims to recognize novel categories from very few labeled instances, enabling fast deployment in specialized domains where data collection is expensive or time-consuming, providing a paradigm for fast, flexible, and scalable recognition in real-world environments.

To enable FSL systems to meet the so-called open-world challenges, we explore three themes that are challenging yet practical in real-world application scenarios.

The first theme of this thesis is to address metric learning challenges in FSL. While metric-based approaches have become predominant, conventional similarity measurements often fail to maintain discriminative power when dealing with fine-grained categories or significant domain shifts. To migrate these limitations, we propose a novel prototype similarity network (ProtoSimNet), which integrates cosine sharpening transformation (CST) and elastic regularization. CST introduces a parametric sharpening function that nonlinearly expands angular decision margins in the cosine similarity space, while the elastic regularization component maintains embedding stability across diverse learning episodes. This combined approach effectively enlarges inter-class differences while preventing overfitting to limited support examples. ProtoSimNet preserves relative similarity rankings while significantly improving model’s discriminative ability, particularly in fine-grained classification scenarios. Extensive experiments demonstrate that our ProtoSimNet establishes new state-of-the-art performance across multiple few-shot benchmarks.

The second theme focuses on feature enrichment, emphasizing the need for richer and more structured visual representations under limited supervision. Traditional feature extractors suffer from insufficient diversity and poor scale adaptability when trained with limited samples. To address these issues, we develop two frameworks: the attention-guided feature pyramid network (AFPN) and the adaptive multi-scale feature fusion (AMFF). AFPN enhances hierarchical feature integration through parameter-free attention mechanisms and cross-layer spatial attention, enabling dynamic emphasis on the most relevant semantic levels for each few-shot task. AMFF

systematically enriches feature representation through structured multi-branch enrichment and adaptive fusion, effectively simulating the feature diversity typically unavailable in few-shot settings. These frameworks demonstrate remarkable effectiveness in handling scale variations, domain shifts, and fine-grained categorization, significantly outperforming conventional feature extraction methods.

The third theme explores knowledge transfer to leverage information learned from large base categories to guide fast learning on novel classes. Standard distillation methods lack selectivity and often transfer irrelevant or misleading information when categories differ significantly. To address this issue, we propose a dual-selective knowledge transfer (DSKT) framework that dynamically selects the most beneficial knowledge components from both teacher models and current task features. By selectively distilling both instance-level and task-level supervisory information, DSKT preserves transferable structure, suppresses noise, and enhances generalization to unseen categories. It is shown that our proposed methods achieve the new state-of-the-art performance in FSL tasks.

We conducted thorough experiments to verify the effectiveness of our proposed methods across these three themes. The results demonstrate significant improvements over various baselines and state-of-the-art methods. Thus this thesis provides important novel contributions, insights, and findings for the research community and future applications in the field of visual understanding and FSL systems.