



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).

Nederlandse Samenvatting

We leven in een tijdperk van kunstmatige intelligentie waarin de vraag naar snelle aanpassing aan nieuwe visuele concepten exponentieel groeit. Het is cruciaal om intelligente visuele herkenningssystemen te ontwikkelen die efficiënt kunnen leren van een beperkt aantal voorbeelden. Een bijzonder belangrijke eigenschap, gezien de toenemende behoefte aan gepersonaliseerde visuele toepassingen, is *few-shot learning* (FSL). Een FSL-systeem is erop gericht nieuwe categorieën te herkennen op basis van een zeer klein aantal gelabelde voorbeelden, waardoor snelle implementatie mogelijk is in gespecialiseerde domeinen waar dataverzameling duur of tijdrovend is. Dit biedt een paradigma voor snelle, flexibele en schaalbare herkenning in realistische omgevingen.

Om FSL-systemen in staat te stellen de zogenaamde open-werelduitdagingen aan te gaan, onderzoeken we drie thema's die uitdagend maar tegelijkertijd praktisch zijn in toepassingsscenario's in de echte wereld.

Het eerste thema van deze scriptie is het aanpakken van de uitdagingen van metrische leertechnieken in FSL. Hoewel op metrische methoden gebaseerde benaderingen dominant zijn geworden, slagen conventionele gelijkenismetingen er vaak niet in om hun onderscheidend vermogen te behouden bij fijnmazige categorieën of significante domeinverschuivingen. Om deze beperkingen te ondervangen, stellen we een nieuw prototype gelijkenisnetwerk voor (ProtoSimNet), dat het cosinusverscherpingstransformatie (CST) en elastische regularisatie integreert. CST introduceert een parametrische verscherpingsfunctie die de hoekige beslissingsmarges in de cosinusgelijkenisruimte niet-lineair vergroot, terwijl de elastische regularisatiecomponent de stabiliteit van de embedding over verschillende leerepisodes behoudt. Deze gecombineerde aanpak vergroot effectief de verschillen tussen klassen en voorkomt tegelijkertijd overfitting aan voorbeelden met beperkte ondersteuning. ProtoSimNet behoudt relatieve gelijkenisrangschikkingen en verbetert tegelijkertijd het onderscheidend vermogen van het model aanzienlijk, met name in fijnmazige classificatiescenario's. Uitgebreide experimenten tonen aan dat ons ProtoSimNet nieuwe state-of-the-art prestaties behaalt op meerdere *few-shot* benchmarks.

Het tweede thema richt zich op het verrijken van kenmerken, waarbij de nadruk ligt op de behoefte aan rijkere en meer gestructureerde visuele representaties onder

beperkte supervisie. Traditionele kenmerkextractors hebben te lijden onder onvoldoende diversiteit en een slechte schaalaanpassing wanneer ze getraind worden met een beperkt aantal voorbeelden. Om deze problemen aan te pakken, ontwikkelen we twee frameworks: het attention-guided feature pyramid network (AFPN) en het adaptive multi-scale feature fusion (AMFF). AFPN verbetert de hiërarchische integratie van kenmerken door middel van parameterloze aandachtmechanismen en ruimtelijke aandacht over meerdere lagen, waardoor dynamische nadruk kan worden gelegd op de meest relevante semantische niveaus voor elke few-shot taak. AMFF verrijkt systematisch de kenmerkrepresentatie door middel van gestructureerde multi-branch verrijking en adaptieve fusie, waardoor de kenmerkdiversiteit die doorgaans niet beschikbaar is in few-shot situaties effectief wordt gesimuleerd. Deze frameworks tonen een opmerkelijke effectiviteit in het omgaan met schaalvariëaties, domeinverschuivingen en fijnmazige categorisatie, en presteren significant beter dan conventionele methoden voor kenmerkextractie.

Het derde thema onderzoekt kennisoverdracht om informatie die is opgedaan uit grote basiscategorieën te benutten voor snel leren van nieuwe klassen. Standaard distillatiemethoden missen selectiviteit en dragen vaak irrelevante of misleidende informatie over wanneer categorieën significant verschillen. Om dit probleem aan te pakken, stellen we een dual-selective knowledge transfer (DSKT)-framework voor dat dynamisch de meest nuttige kenniscomponenten selecteert uit zowel de leermodellen als de huidige taakkenmerken. Door selectief zowel supervisie-informatie op instantieniveau als op taakniveau te distilleren, behoudt DSKT de overdraagbare structuur, onderdrukt ruis en verbetert de generalisatie naar onbekende categorieën. Er wordt aangetoond dat onze voorgestelde methoden de nieuwe state-of-the-art prestaties behalen in FSL-taken.

We hebben grondige experimenten uitgevoerd om de effectiviteit van onze voorgestelde methoden te verifiëren aan de hand van deze drie thema's. De resultaten tonen significante verbeteringen ten opzichte van diverse basismodellen en state-of-the-art methoden. Deze dissertatie levert daarmee belangrijke nieuwe bijdragen, inzichten en bevindingen voor de onderzoeksgemeenschap en toekomstige toepassingen op het gebied van visuele interpretatie en FSL-systemen.