



Universiteit
Leiden
The Netherlands

Learning from small samples

Kocaman, V.

Citation

Kocaman, V. (2024, February 20). *Learning from small samples*. Retrieved from <https://hdl.handle.net/1887/3719613>

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/1887/3719613>

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

Nederlandse Samenvatting

Kleine datasets vormen een uitdaging voor veel machine learning technieken, in het bijzonder bij ongebalanceerde datasets en anomalieën detectie. Dit proefschrift onderzoekt verschillende benaderingen voor het effectief leren van kleine datasets, het aanpakken van uitdagingen. Daarnaast biedt het een uitgebreide analyse van bestaande technieken in de literatuur. Het onderzoek richt zich specifiek op twee hoofdvragen: 1) wat zijn de huidige methoden voor het leren van taken gebaseerd op kleine datasets in machine learning, en 2) wat is de impact van batch normalisatie en opvallende beeldsegmentatie als augmentatie techniek bij self-supervised learning op de prestaties van het model.

Het proefschrift presenteert een gedetailleerd overzicht van technieken voor het omgaan met kleine datasets, zoals data selectie, data voorverwerking, ensemble-methoden, transfer learning, regularisatie technieken en synthetische data generatie. Vervolgens wordt de impact van batch normalisatie onderzocht, waarbij wordt aangetoond dat deze de trainingstijd en de kwaliteit van de voorspellingen voor minder voorkomende klassen in sterk ongebalanceerde datasets aanzienlijk verbetert. Het onderzoek gaat verder in op de rol van self-supervised learning als augmentatie techniek, waarbij wordt aangetoond dat het gebruik van opvallende beeldsegmentatie de geleerde representaties verbetert, vooral in downstream taken zoals beeldsegmentatie.

Samenvattend draagt deze studie bij aan het vakgebied van machine learning door strategieën te verkennen voor het leren van kleine datasets. Het biedt een gedetailleerde analyse van batch normalisatie, waarbij het potentieel wordt be-

Nederlandse Samenvatting

nadrukt om de prestaties voor minderheidsklassen in ongebalanceerde datasets te verbeteren. Daarnaast introduceert de studie opvallende beeldsegmentatie als augmentatiebeleid in zelfgestuurd leren, waarbij de effectiviteit ervan in taken zoals beeldsegmentatie wordt aangetoond. Deze bevindingen bieden een solide basis voor verder onderzoek in het leren van taken gebaseerd op kleine datasets en presenteren praktische inzichten voor machine learning professionals die werken met beperkte data.