



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).

English Summary

Learning from small data sets in machine learning is a crucial challenge, especially when dealing with data imbalances and anomaly detection. This thesis delves into the challenges and methodologies of learning from small datasets in machine learning, with a particular focus on addressing data imbalances and anomaly detection. It thoroughly explores various strategies for effective small dataset learning in ML, examining both existing approaches and introducing novel techniques. The research pivots around two key questions: firstly, it investigates current methods employed for learning from small datasets in machine learning, and secondly, it assesses the efficacy of batch normalization in enhancing model performance and utilizing salient image segmentation as an augmentation policy in self-supervised learning.

The thesis comprehensively reviews techniques for managing small datasets, including data selection and preprocessing, ensemble methods, transfer learning, regularization techniques, and synthetic data generation. A critical examination of batch normalization reveals its significant role in improving training time and testing errors for minority classes in highly imbalanced datasets. The study also demonstrates that utilizing salient image segmentation as an augmentation policy in self-supervised learning substantially improves representation learning. This improvement is particularly evident in the context of downstream tasks such as image segmentation, highlighting the effectiveness of this technique in enhancing model performance.

In summary, this study contributes to the field of machine learning by exploring

English Summary

strategies for learning from small datasets. It offers a detailed analysis of batch normalization, highlighting its potential in improving performance for minority classes in imbalanced datasets. Additionally, the study introduces salient image segmentation as an augmentation policy in self-supervised learning, showing its effectiveness in tasks like image segmentation. These findings provide a solid foundation for further research in small sample learning and present practical insights for machine learning practitioners working with limited data.