Doctoral Dissertation
Doctoral Program in Computer Enginering (34.th cycle)

Machine learning methods for the analysis and interpretation of images and other multi-dimensional data

Sina Famouri

* * * * * *

Supervisor
Prof. Fabrizio Lamberti

Doctoral Examination Committee:
Prof. Anke Meyer-Baese, Referee, Florida State University
Prof. Anna Vignati, Referee, University of Turin

Politecnico di Torino
Examination date TBD

I hereby declare that, the contents and organisation of this dissertation constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data.

. . . .

Sina Famouri

Turin, Examination date TBD

Abstract

Machine learning, and in particular deep learning-based models such as convolutional neural networks, have consistently demonstrated unprecedented performance in the analysis and interpretation of images and more generally of multidimensional complex data. In this context, fine-grained object recognition and classification, characterized by subtle differences between classes and large variations within classes, is attracting increasing attention not only in the general computer vision community, but also in specialized fields with high performance requirements, such as medical and industrial applications. Since collecting large annotated datasets is expensive and time consuming, especially in the above-mentioned domains, advances in unsupervised, semi-supervised, self-supervised, and transfer learning are key to substantial improvements. In particular, this research focuses on computational diagnosis in the medical domain, which combines the need for fine-grained analysis of subtle disease patterns with the hurdles of developing deep learning models that can efficiently process high-dimensional and multidimensional images.

Despite the contribution of deep learning in the medical field, the success of such models is hampered by data limitations. In terms of data volume, the digitization of medical records generates enormous amounts of data on a daily basis. However, apart from the administrative difficulties in obtaining the data, acquiring large and well-curated datasets is quite costly. Therefore, the typical size of datasets is rather small compared to natural RGB images. Hence, it may be difficult for deep neural networks to generalize beyond initial laboratory testing, limiting clinical application or even requiring further training in local clinical centers.

The goal of this thesis is to explore various ways to reduce deep learning over-reliance on large datasets to make it more data efficient without sacrificing robustness. Possible solutions to combat the lack of data in the medical domain include either reducing the cost of data annotation or, alternatively, reducing the amount of annotated data required for training. This research takes a step toward the following objectives: (1) providing methods that makes training robust against noise, (2) incorporating information from different views of the same organ, and (3) developing multi-task, cross-domain training pipelines that exploit self-supervision to lessen reliance on annotations.

For object detection in medical images, achieved results show that, if the reference standard is noisy, the model is not able to correctly quantify the agreement of reference standard and the networks predictions, which consequently affects the training in a negative way. As a result, a novel approach is devised to counteract this effect by changing the criteria used for matching the bounding boxes proposed by the network with the ground truth. This helps to relax the requirements on

annotated medical images and makes the training robust to noise, enabling the use of larger automatically curated datasets for training.

Regarding the use of information from different views of the same organ, a registration method has been proposed to align two views of the same organ in mammography images and incorporate this information into object recognition.

Finally, to address the problem of data scarcity, a dataset has been created as a harmonized collection of 17 publicly available datasets covering various medical imaging modalities and body parts. Subsequently, self-supervised pre-training has been applied to this large collection, and the learned representations have been evaluated by transfer learning to different tasks of computer-aided diagnosis. The aforementioned dataset and the model based on it are referred to as MedNet interchangeably in this context. Experimental results have shown that the extracted features are able to discriminate better than ImageNet, which is currently the *de facto* standard for transfer learning in the medical field. However, further investigation has shown that the final performance of the two pre-trained networks is similar after fine-tuning and that ImageNet slightly outperforms MedNet. Thus, this research has examined the complementary roles played not only by the source and target datasets, but also by the self-supervised and target tasks in determining the most effective strategy for training deep neural networks on medical datasets.

This Ph.D. thesis has been typeset by means of the TEX-system facilities. The typesetting engine was pdfLATEX. The document class was `toptesi`, by Claudio Beccari, with option `tipotesi=scudo`. This class is available in every up-to-date and complete TEX-system installation.