Learning from population-scale medical imaging data without labels

Master Project

Research Group: Digital Health & Machine Learning

Supervisor: Prof. Dr. Christoph Lippert

Abstract

In this project, we will implement and benchmark unsupervised and self-supervised deep learning methods, as well as deep generative models for 2D and 3D medical imaging data. Without the use of expert-curated labels, we will be training convolutional neural nets on millions of 2D X-ray images and 3D structural MRI scans from two population-based imaging efforts, the UK Biobank, the Vukuzazi study in South Africa, as well as clinical imaging data from the Mt. Sinai health system in New York City. While imaging data is abundant in these data sets, interpretation and annotation from medical experts is limited and represents a main bottleneck in the medical domain, where accurate expert annotations by radiologists are time-consuming and thus limited to small sample sizes. Therefore, our goal is to break this crucial sample size barrier using approaches that can learn from data without supervision.

Motivation and Background

Medical imaging plays a vital role in patient healthcare. It aids in disease prevention, early detection, diagnosis, and treatment. Physicians frequently order scans on which they perform various diagnostic tasks, such as detecting abnormalities and quantifying measurements. However, image interpretation by humans has limitations, due to subjectivity, variation across interpreters, fatigue, and, most importantly, non-scalability [1]. For instance, the UK Biobank [20] has released 3D whole body MRI scans for 30,000 participants, consisting of millions of individual image slices [21]. Vukuzazi is a community-based health screening study in Kwa-Zulu Natal, South Africa, a region suffering from an extremely high burden from HIV-AIDS and Tuberculosis. Since completion of the pilot in early 2018, over 10,000 subjects have been enrolled in the Vukuzazi study and undergone a routine chest X-ray to uncover signs of scaring in their lung fields. Mt. Sinai health system is the leading health system in New York City, amassing doctor-prescribed imaging data in routine clinical practice.

Diagnostic tasks performed by physicians include organ (or body part) segmentation (on 2D or 3D scans) as one of the most commonly used applications in medical imaging. Manual annotation/segmentation for patient scans is a non-trivial and time-consuming task, especially on 3D scans. In fact, with the growing sizes of imaging datasets, as mentioned earlier, expert annotation becomes nearly impossible without computerized assistance [23]. Even current semi-automatic software tools, which are being used by our collaborators (cardiologists and neuroscientists), fail to sufficiently reduce the time and effort required for annotation and measurement of these large data sets. For all these reasons, deep learning (DL) methods promise a solution. However, while large imaging datasets are being generated, expert annotations, which are required for training DL models, are still scarce in the medical domain due to the reasons mentioned above.

Thus, we plan to utilize large amounts of unlabeled data in this work by using unsupervised learning and generative models that can learn quantitative data representations without the use of label information.

Similarly, we will self-supervised learning, in which supervision signals are derived from the data itself. So, no manual annotation is required. Self-supervised methods exist in multiple application fields of deep learning, such as Word2Vec [2] and similar word embeddings in text mining. Word2Vec predicts a word from its surrounding words (and vice versa), thus utilizing the context to derive the semantic representation of the word. Context prediction, mainly spatial context, also have inspired self-supervised research in computer vision, such as in [3], in which a visual representation is trained using the pretext task of predicting the position of an image patch relative to another. Here, the model is phrased such that it has to understand the image content to solve this problem. Similar work has been done to solve Jigsaw Puzzles [4], and to color gray-scale images [5]. Other sources of supervision signals have been derived from videos. Videos can provide representations mostly through temporal continuity [6, 7, 8, 9, 12], and motion consistency [10, 11, 13]. In addition, other data modalities, such as text [24] or audio [25], have been used to learn representations for accompanying prediction tasks on images.

In the medical sector, self-supervised learning has found use cases in robotic surgery [14], in dense depth estimation in monocular endoscopy [15], in medical image registration [16], in body part recognition [17], in feature learning from 3D scans [18], and in studying disc degeneration using spinal MRIs [19]. However, while these attempts are a step forward for self-supervised learning in medical imaging, they still have some limitations. First, none of these works have reported evaluation experiments on sufficiently large datasets, sometimes without any comparisons to other methods. Second, most of the reported solutions hardly generalize, as they are highly engineered towards solving a specific application.

Objectives and Proposed Methodology

The long term goal of this research is to minimize the required effort by humans in annotating large medical imaging cohorts. As seen in the background section, little work has been done in improving supervision in the medical domain. Therefore, this work aims to fill this gap in literature and advance the research in self-supervised learning. We expect to achieve this through the following sub-objectives:

- Perform a comprehensive review of unsupervised / self-supervised approaches which have been applied in reducing annotations required by experts.
- Choose well-proven methods and evaluate them on our test datasets of large medical imaging corpora. This step is required to establish baseline numbers which we can compare with.
- Develop self-supervised methods, which can learn representations from unlabelled data, which can be solely images or maybe accompanied by other modalities. The methods we aim to create are not necessarily fully unsupervised in nature, they might be semi-supervised.
- Evaluate the developed methods against the chosen baselines on medical imaging cohorts.
- Visualize learned representations of imagine phenotypes for UK Biobank and Vukuzazi
- Quantify healthy variation in imaging phenotypes and identify subgroups in the data corresponding to common disease groups
- Statistically associate changes in learned representations of imaging phenotypes to genetic information in UK Biobank

In terms of medical tasks we plan to explore, we mention image segmentation on MRI scans as one of the tools used for disease diagnosis. Many disease phenotypes can be identified by measuring the respective body part's volume, e.g. heart or brain ventricles, and this requires segmenting the area to quantify the volumes. One may use the heart's beating motion captured in MRI scans to learn automatic segmentation labels, for instance. As another task, image classification is an effective tool for diagnosing diseases, such as Thoracic diseases from chest scans. Finally, object detection can also be applied to detect an abnormality, such as a mass, an organ, or any region of interest.

The applications for the above tasks are countless beyond the medical field, too. For instance, applications for image segmentation, object detection, and image classification are abundant in the enterprise world, using customer or public datasets. Thus, we expect improvements on self-supervised methods to affect these applications positively, by reducing the cost required for data annotation.

References

[1] H. Greenspan., B. van Ginneken., R. M. Summers. *Deep Learning in Medical Imaging: Overview and Future Promise of an Exciting New Technique.* Guest Editorial, IEEE Transactions on Medical Imaging, VOL. 35, NO. 5, MAY 2016.

[2] T. Mikolov., K. Chen., G. Corrado., J. Dean. *Efficient Estimation of Word Representations in Vector Space.* arXiv:1301.3781 [cs.CL]

[3] C. Doersch., A. Gupta., A.A. Efros. *Unsupervised Visual Representation Learning by Context Prediction.* arXiv:1505.05192 [cs.CV]

[4] M. Noroozi., P. Favaro. *Unsupervised Learning of Visual Representations by Solving Jigsaw Puzzles.* arXiv:1603.09246 [cs.CV]

[5] R. Zhang., P. Isola., A.A. Efros. *Colorful Image Colorization.* arXiv:1603.08511 [cs.CV]

[6] X. Wang., A. Gupta. *Unsupervised Learning of Visual Representations using Videos.* arXiv:1505.00687 [cs.CV]

[7] C. Vondrick., H. Pirsiavash., A. Torralba. *Anticipating Visual Representations from Unlabeled Video.* arXiv:1504.08023 [cs.CV]

[8] H. Mobahi., R. Collobert., J. Weston. *Deep Learning from Temporal Coherence in Video.* Proceedings of the 26th International Conference on Machine Learning, Montreal, Canada, 2009.

[9] D. Jayaraman., K. Grauman. *Slow and steady feature analysis: higher order temporal coherence in video.* arXiv:1506.04714 [cs.CV]

[10] J. Walker., A. Gupta., M. Hebert. *Dense Optical Flow Prediction from a Static Image* . arXiv:1505.00295 [cs.CV]

[11] S. Purushwalkam., A. Gupta. *Pose from Action: Unsupervised Learning of Pose Features based on Motion.* arXiv:1609.05420 [cs.CV]

[12] I. Misra., C.L. Zitnick., M. Hebert. *Shuffle and Learn: Unsupervised Learning using Temporal Order Verification.* arXiv:1603.08561 [cs.CV]

[13] D. Pathak., R. Girshick., P. Dollár., T. Darrell., B. Hariharan. *Learning Features by Watching Objects Move.* arXiv:1612.06370 [cs.CV]

[14] M. Ye., E. Johns., A. Handa., L. Zhang., P.Pratt., G.Z. Yang. *Self-Supervised Siamese Learning on Stereo Image Pairs for Depth Estimation in Robotic Surgery.* arXiv:1705.08260 [cs.CV]

[15] X. Liu., A. Sinha., M. Unberath., M. Ishii., G. Hager., R. H. Taylor., A. Reiter. *Self-supervised Learning for Dense Depth Estimation in Monocular Endoscopy.* arXiv:1806.09521 [cs.CV]

[16] H. Li., Y. Fan. *Non-rigid Image Registration using Self-Supervised Fully Convolutional Networks without Training Data.* Proc IEEE Int Symp Biomed Imaging. 2018 Apr; 2018: 10751078.

[17] P. Zhang., F. Wang., Y. Zheng. *Self supervised deep representation learning for fine-grained body part recognition.* 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017).

[18] M. Blendowski., H. Nickisch., M.P. Heinrich. *Self-Supervised Convolutional Feature Training for Medical Volume Scans.* 1st Conference on Medical Imaging with Deep Learning (MIDL 2018), Amsterdam, The Netherlands.

[19] A. Jamaludin., T. Kadir., A. Zisserman. *Self-Supervised Learning for Spinal MRIs.* arXiv:1708.00367 [cs.CV]

[20] UK Biobank
http://www.ukbiobank.ac.uk/

[21] *https://imaging.ukbiobank.ac.uk/*

[22] German National Cohort
https://nako.de/wp-content/uploads/2015/10/2016-03-18_NAKO-Fact-Sheet-EN.pdf

[23] Annotating Medical Image Data
https://link.springer.com/content/pdf/10.1007%2F978-3-319-49644-3_4.pdf

[24] Y. Patel., L. Gomez., M. Rusiñol., C. Jawahar., D. Karatzas. *Self-supervised learning of visual features through embedding images into text topic spaces*. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017)

[25] A. Owens., J. Wu., J.H. McDermott., W.T. Freeman, A. Torralba. *Ambient sound provides supervision for visual learning*. European Conference on Computer Vision. pp. 801816. Springer (2016)