VISCERAL: Towards Large Data in Medical Imaging — Challenges and Directions

Georg Langs[1], Allan Hanbury[2], Bjoern Menze[3], Henning Müller[4]

[1] Computational Image Analysis and Radiology Lab Department of Radiology, Medical University of Vienna
[2] Institute for Software Technology and Interactive Systems Faculty of Computer Science, Vienna University of Technology
[3] Computer Vision Lab, ETH Zurich
[4] University of Applied Sciences Western Switzerland, Sierre (HES-SO)

Abstract. The increasing amount of medical imaging data acquired in clinical practice holds a tremendous body of diagnostically relevant information. Only a small portion of these data are accessible during clinical routine or research due to the complexity, richness, high dimensionality and size of the data. There is consensus in the community that leaps in this regard are hampered by the lack of large bodies of data shared across research groups and an associated definition of joint challenges on which development can focus. In this paper we describe the objectives of the project VISCERAL. It will provide the means to jump–start this process by providing access to unprecedented amounts of real world imaging data annotated through experts and by using a community effort to generate a large corpus of automatically generated standard annotations. To this end, VISCERAL will conduct two competitions that tackle large scale medical image data analysis in the fields of anatomy detection, and content–based image retrieval, in this case the retrieval of similar medical cases using visual data and textual radiology reports.

Keywords: Medical imaging, Large scale data, Localization, Retrieval

1 Introduction

The amount and complexity of information present in medical imaging data on a hospital scale is enormous. Part of this information has immediate diagnostic relevance, part becomes relevant only when studied in the context of a large cohort (e.g., when studying subtle characteristics of diseases such as mild cognitive impairment), and part might not only be relevant at the time of acquisition, but also when used as reference during later radiological assessment. In the context of both *computer aided diagnosis (CAD)*, and medical research, a large research community has emerged that tackles the extraction and quantification of task specific relevant information from medical imaging data [1]. Traditional problems in this domain are the segmentation of anatomical structures, the detection of pathology, or the measurement of specific markers that correlate with disease or treatment outcome.

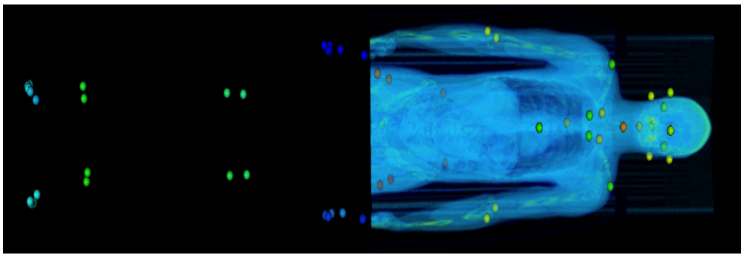

Fig. 1: Can we learn more than predefined targets from medical imaging data ?

The natural variability in these data and the corresponding difficulty in identifying anomalities that may be only subtle deviations from a healthy cohort results in a large body of work focusing on specific diseases. Typically approaches, such as CAD rely on a well controlled set of examples and corresponding expert annotations [2, 3].

This *paradigm* has several limitations. First, it typically focuses on replicating expert judgements learned during the training phase. More importantly, it does not scale to amounts of data necessary to understand diseases with subtle and complex effects that would require a much larger set of annotated examples in order to represent the difference between control– and patient group well. Both limitations together result in a severe hampering of efforts to identify novel markers, that are not part of the annotated training corpus or existing clinical routine.

1.1 Towards Large Data

Scaling to large data is necessary for computational approaches dealing with a wide range of diseases. It is a prerequisite to understand subtle characteristics of populations and to computationally identify novel markers. Among the challenges we face, and for which we lack adequate methodology, are:

- Making information in large data accessible during clinical routine (e.g., radiological assessment).
- Learning models of diseases and corresponding imaging markers with no or only limited supervision.
- Fast and efficient collection of large amounts of data relevant for a specific question.
- Leveraging the amount of data, and the relatively confined domain of the human body in an optimal way, when analyzing medical imaging data, and when identifying sublte markers, and relationships.
- Using both radiology report- and image information simultaneously, when searching for findings in images, which were part of clinical routine.

These are only a few of the challenges we face, and whose solution would be a crucial step towards harvesting the information present in medical images, that currently remains unused. To tackle these questions we need novel methodology.

A parallel development taking place in the computer vision community suggests that the step from small- to large data is accompanied by substantial changes in approaching tasks such as learning, representation, or retrieval [4]. It coincides with a shift from supervised learning methods to semi–supervised or even unsupervised training. An example is the aim to scale beyond hundreds of thousands of images and hundreds of categories in the context of the Internet. This has shown promising results in [5]. Unsupervised learning approaches for category models from large image data sets have been explored, too. Examples are [4, 6–8].

While information extraction from images in 2D is an active field of research by far the largest amount of visual data produced in hospitals are multi–slice series of images [2]. Benchmarks in medical imaging exist with ImageCLEF [9] but focusing rather on 2D and textual image retrieval, whereas the largest of data produced is currently 3D tomographic data. Medical imaging data are estimated to have reached 30% of the overall world wide data storage in 2010 [10].

In this paper we outline the main challenges we face when working with and analyzing large medical imaging data, and suggest two primary directions where advances are needed. We propose two corresponding challenges for discussion. They will be organized in the EU funded VISCERAL project to help focus the efforts in the community and to offer a means to compare methodology across research groups worldwide.

2 Open Questions in Medical Imaging and Directions Proposed

Recently, interest in alternatives and conceptual extensions to traditional CAD systems has emerged. Among others, an example of a particularly promising direction is the use of image retrieval that instead of providing a direct automated assessment allows for the efficient search for comparable and relevant cases. These cases are presented to aid the physician who is performing reading and interpretation of the case at hand. The visual content of medical images has been used for information retrieval for over 15 years [11] and has shown to improve quality of diagnosis [12]. Visual retrieval can extract rich information beyond the associated textual cues and is a promising direction to make use of the medical imaging databases in hospitals. Often, image retrieval is is mixed up with very simple tasks of classifying images into modality and anatomic region. Such a classification as preprocessing can not really be seen as a retrieval process and is rather the first step for the extraction of information usable to provide matching cases, anatomical regions or mine the data for specific pathologies.

Methodological challenges of interest include:

1. Scalability to large amounts of data. What is necessary to work on real data of a PACS and thus on very large and heterogeneous data sets, which have

never been analyzed at such a large scale for retrieval of medical visual data as of yet? The VISCERAL project will extend the medical image analysis towards very large data sets (hundreds of categories, and many thousands of data sets), which makes use of a new families of methods (unsupervised learning, modeling, and categorization) necessary. These methods have until now mainly been used in the computer vision community but only little for medical data.
2. Unsupervised and autonomous learning: the scale of the data (many TeraBytes) makes the fully autonomous and un–supervised building of models and categorization/indexing of the data crucial.
3. What is the right interface for injecting prior knowledge on anatomy and other structural elements into this algorithmic analysis?
4. Efficient annotation on pre–processed data instead of raw images to facilitate annotation. Making use of semi–supervised strategies instead of supervised learning on limited training bodies.
5. Generalization power to a large and diverse set of data, and the inclusion of a potentially growing set of training images during the learning phase.
6. Introducing image–based search and retrieval as an alternative to computational classification and traditional computer aided diagnosis that is concentrating on single specific phenomena.

3 Two Challenges to Focus Research Efforts

We propose two challenges, with the aim of spurning discussion, and refinement based on the response by the community. Challenges will be based on large amounts of partially annotated data.

3.1 Competition 1:

Anatomical structure identification, detection and segmentation competition on a full body scale for fully-automated processing in the clinic. Participants are provided with 3D image data (multimodal full body scans, volumes containing specific anatomical structures as encountered in clinical practice) together with training annotations on a subset of the data. For evaluation, test data consisting of 3D volumes will be processed by the participants algorithms. The objective is to identify, localize and segment the anatomical structures present in the data. We will evaluate both with regard to comprehensive identification and to subset localization, in order to be able to include algorithms developed for specific organs as a secondary task within the competition

3.2 Competition 2:

Similar case retrieval to allow tools and algorithms to be evaluated on real clinical tasks with a potentially larger impact. Given a query case consisting of either image, volume or potentially additional textual information the objective is to retrieve similar cases in terms of characteristics such as diagnosis or differential diagnosis. Challenges that have to be solved include incorrect or incomplete data (for example, data that was not entered into the record by a physician), and potentially very small regions relevant for the similarity computation (for example, a small fracture in an entire thorax CT).

3.3 Secondary impact

Besides the immediate impact on research in the two suggested direction the Visceral project has the potential to advance the state of the art in several related questions:

- Allow medical image analysis on a very large scale (the scale of a PACS in many regional hospitals) and compare the results across many research groups;
- Create ground truth based on partial manual annotation and then also based on the results of the participating systems, also allowing to scale from purely manual annotation to mixed approaches of manual and semi-automatic ground truthing;
- Develop an infrastructure for medical image data sharing and potentially also sharing of components of participants in the cloud;
- Compare techniques for quality and speed across many research groups based on the same data and tasks, allowing to identify the most promising approaches to several current research challenges;
- Potentially allow for the creation of approvals for image retrieval as diagnosis aid if a reference database and a proof of quality can be shown.

4 Corpora

In parallel to serving as a benchmark to compare methods across a large number of possible alternatives, the challenges will serve as the basis for two corpora that collect imaging- and text data together with expert annotations, and semi-automated annotations achieved during the competitions. Part of the data will be annotated by experts, to obtain a *Gold Corpus*. At the same time, every participant will be contributor to a *Silver Corpus*. The latter will be formed by deriving a consensus across the entries of the participants.

The generation of the silver corpus will require methods for consolidating annotations obtained from independent groups participating in the competition. In the competition, algorithms are to be trained or tuned on the gold corpus train data; test data will be the silver corpus data and the gold corpus test data. Averaging the labels over the silver corpus data will lead to a fused label that is better than the individual label estimate, if estimates are unbiased. We will measure bias of the labels on the gold corpus test data that will be hidden in the test set. This will require a tool that also returns general quality measure (bias, variance) for individual image volumes, as well as for individual contributors of labels sets (i.e., algorithms).

5 Infrastructure

To allow for the distribution of very large data sets a new type of infrastructure seems necessary. 10 TB of data can not easily be downloaded or shipped on hard disk to many participants. Could computing on the other hand is often proposed for dealing with large storage and computing capacities. VISCERAL will make these capacities available to the medical imaging community and give participants in the benchmark access to a virtual machine (Windows on Linux) in the cloud to access the training data and prepare the systems. Then, the execution of the tools on the text data will be started by the organizers and results can subsequently be compared based on the runs of the test data. This allows to compare tools not only based on classification or retrieval accuracy but also based on speed or efficiency criteria, which becomes increasingly important when dealing with very large data sets. This should also give equal opportunities to groups from poorer and richer countries as all resources are controlled. The fact that only access to training data is given to the participants means that non–anonymous data sets could be distributed and used in the same way potentially. Bringing the algorithms to the data and not the data to the algorithms seems the only feasible approach when looking at extremely large–scale challenges.

6 Conclusion

Medical image analysis has brought many new interesting and successful techniques. Over the past 30 years it has helped to develop novel tools to aid diagnosis, that have a substantial impact on diagnosis quality, and treatment success. The quickly increasing amount of data produced and digitally available poses new challenges. How can we make use of these data and how can we exploit knowledge being stored in past data.

At the same time, quickly rising costs in health care will make it necessary to use all available data in the best possible way to take case of new patients. Past data can help in this process, keeping the privacy of patients protected at the same time.

In the Visceral project we propose two challenges to the medical imaging research community. Both challenges will use medical imaging data in a new scale, making in the order of 10 TB available for research. Both challenges concentrate on specific aspects in the image analysis chain. The first challenge concentrates on extracting anatomic regions and locations from the data, which is necessary in many contexts, such as for all further steps that compare tissue and abnormalities within the same anatomic regions of several patients. The second challenge focuses on the retrieval of similar cases: pathology–oriented retrieval. Algorithms that tackle the latter challenge can -although not traditional CAD - by viewed as a diagnosstic aid in fields such as evidence–based medicine where studies related to a specific patient need to be found and case–based reasoning where similar cases are compared to a patient being treated.

A crucial component of Visceral is the participation of the community at an early stage, when specifying and refining the tasks, and benchmakring measures, in order to truly support relevant research. and to allow VISCERAL to contribute to a leap forward in bringing more medical imaging to the clinical workflow.

References

1. Doi, K.: Current status and future potential of computer–aided diagnosis in medical imaging. British Journal of Radiology **78** (2005) 3–19
2. Depeursinge, A., Vargas, A., Platon, A., Geissbuhler, A., Poletti, P.A., Müller, H.: 3D case–based retrieval for interstitial lung diseases. In: MCBR–CDS 2009: Medical Content–based Retrieval for Clinical Decision Support. Lecture Notes in Computer Science (LNCS), Springer (February 2010) 39–48
3. Chen, W., Giger, M.L., Li, H., Bick, U., Newstead, G.M.: Volumetric texture analysis of breast lesions on contrast–enhanced magnetic resonance images. Magnetic Resonance in Medicine **58**(3) (September 2007) 562–571
4. Torralba, A., Fergus, R., Freeman, W.: 80 million tiny images: A large data set for nonparametric object and scene recognition. Pattern Analysis and Machine Intelligence, IEEE Transactions on **30**(11) (2008) 1958–1970
5. Oliva, A., Torralba, A.: Building the Gist of a Scene: The Role of Global Image Features in Recognition. Visual Perception (2006)
6. Fei-Fei, L., Fergus, R., Perona, P.: Learning generative visual models from few training examples: An incremental Bayesian approach tested on 101 object categories. Computer Vision and Image Understanding **106**(1) (2007) 59–70
7. Langs, G., Donner, R., Peloschek, P., Bischof, H.: Robust autonomous model learning from 2d and 3d data sets. Med Image Comput Comput Assist Interv Int Conf Med Image Comput Comput Assist Interv **10**(Pt 1) (2007) 968–76
8. Csurka, G., Dance, C., Fan, L., Willamowski, J., Bray, C.: Visual categorization with bags of keypoints. In: Workshop on Statistical Learning in Computer Vision, ECCV. (2004) 1–22
9. Kalpathy-Cramer, J., Müller, H., Bedrick, S., Eggel, I., Seco de Herrera, A.G., Tsikrika, T.: The CLEF 2011 medical image retrieval and classification tasks. In: Working Notes of CLEF 2011 (Cross Language Evaluation Forum). (September 2011)
10. : Riding the wave how europe can gain from the rising tide of scientific data. Final report of the High level Expert Group on Scientific Data. A submission to the European Comission (October 2010)
11. Lowe, H.J., Antipov, I., Hersh, W., Smith, C.A.: Towards knowledge–based retrieval of medical images. The role of semantic indexing, image content representation and knowledge–based retrieval. In: Proceedings of the Annual Symposium of the American Society for Medical Informatics (AMIA), Nashville, TN, USA (October 1998) 882–886
12. Aisen, A.M., Broderick, L.S., Winer-Muram, H., Brodley, C.E., Kak, A.C., Pavlopoulou, C., Dy, J., Shyu, C.R., Marchiori, A.: Automated storage and retrieval of thin–section CT images to assist diagnosis: System description and preliminary assessment. **228**(1) (July 2003) 265–270