Cloud-Based Benchmarking of Medical Image Analysis
Allan Hanbury · Henning Müller
Georg Langs
Editors

Cloud-Based Benchmarking of Medical Image Analysis

SpringerOpen
The VISCERAL project\(^1\) organized Benchmarks for analysis and retrieval of 3D medical images (CT and MRI) at a large scale. VISCERAL used an innovative cloud-based evaluation approach, where the image data were stored centrally on a cloud infrastructure, while participants placed their programs in virtual machines on the cloud. This way of doing evaluation will become increasingly important as evaluation of algorithms on increasingly large and potentially sensitive data that cannot be distributed will be done.

This book presents the points of view of both the organizers of the VISCERAL Benchmarks and the participants in these Benchmarks. The practical experience and knowledge gained in running such benchmarks in the new paradigm is presented by the organizers, while the participants report on their experiences with the evaluation paradigm from their point of view, as well as giving a description of the approaches submitted to the Benchmarks and the results obtained.

This book is divided into five parts. Part I presents the cloud-based benchmarking and Evaluation-as-a-Service paradigm that the VISCERAL Benchmarks used. Part II focusses on the datasets of medical images annotated with ground truth created in VISCERAL that continue to be available for research use, covering also the practical aspects of getting permission to use medical data and manually annotating 3D medical images efficiently and effectively. The VISCERAL Benchmarks are described in Part III, including a presentation and analysis of metrics used in the evaluation of medical image analysis and search. Finally, Parts IV and V present reports of some of the participants in the VISCERAL Benchmarks, with Part IV devoted to the Anatomy Benchmarks, which focused on segmentation and detection, and Part V devoted to the Retrieval Benchmark.

This book has two main audiences: Medical Imaging Researchers will be most interested in the actual segmentation, detection and retrieval results obtained for the tasks defined for the VISCERAL Benchmarks, as well as in the resources (annotated medical images and open source code) generated in the VISCERAL project,

\(^1\)http://visceral.eu
while *eScience and Computational Science Reproducibility Advocates* will gain from the experience described in using the Evaluation-as-a-Service paradigm for evaluation and benchmarking on huge amounts of data.

Vienna, Austria  
Sierre, Switzerland  
Vienna, Austria  
September 2016

Allan Hanbury  
Henning Müller  
Georg Langs
The work leading to the results presented in this book has received funding from the European Union Seventh Framework Programme (FP7/2007–2013) under Grant Agreement No. 318068 (VISCERAL).

The cloud infrastructure for the benchmarks was and continues to be supported by Microsoft Research on the Microsoft Azure Cloud.

We thank the reviewers of the VISCERAL project for their useful suggestions and advice on the project reviews. We also thank the VISCERAL EC Project Officer, Martina Eydner, for her support in efficiently handling the administrative aspects of the project.

We thank the many participants in the VISCERAL Benchmarks, especially those that participated in multiple Benchmarks. This enabled a very useful resource to be created for the medical imaging research community. We also thank all contributors to this book and the reviewers of the chapters (Marc-André Weber, Oscar Jimenez del Toro, Orcun Goksel, Adrien Depeursinge, Markus Krenn, Yashin Dicente, Johannes Hofmanninger, Peter Roth, Martin Urschler, Wolfgang Birkfellner, Antonio Foncubierta Rodriguez).

1 http://visceral.eu
# Contents

## Part I  Evaluation-as-a-Service

1  VISCERAL: Evaluation-as-a-Service for Medical Imaging ........ 3  
   Allan Hanbury and Henning Müller

2  Using the Cloud as a Platform for Evaluation  
   and Data Preparation ..................................... 15  
   Ivan Eggel, Roger Schaer and Henning Müller

## Part II  VISCERAL Datasets

3  Ethical and Privacy Aspects of Using Medical Image Data ....... 33  
   Katharina Grünberg, Andras Jakab, Georg Langs,  
   Tomàs Salas Fernandez, Marianne Winterstein, Marc-André Weber,  
   Markus Krenn and Oscar Jimenez-del-Toro

4  Annotating Medical Image Data ............................... 45  
   Katharina Grünberg, Oscar Jimenez-del-Toro, Andras Jakab,  
   Georg Langs, Tomàs Salas Fernandez, Marianne Winterstein,  
   Marc-André Weber and Markus Krenn

5  Datasets Created in VISCERAL ................................. 69  
   Markus Krenn, Katharina Grünberg, Oscar Jimenez-del-Toro,  
   András Jakab, Tomàs Salas Fernandez, Marianne Winterstein,  
   Marc-André Weber and Georg Langs

## Part III  VISCERAL Benchmarks

6  Evaluation Metrics for Medical Organ Segmentation  
   and Lesion Detection ........................................ 87  
   Abdel Aziz Taha and Allan Hanbury
7  VISCERAL Anatomy Benchmarks for Organ Segmentation and Landmark Localization: Tasks and Results .................. 107
Orcun Goksel and Antonio Foncubierta-Rodriguez

8  Retrieval of Medical Cases for Diagnostic Decisions: VISCERAL Retrieval Benchmark .............................. 127
Oscar Jimenez-del-Toro, Henning Müller, Antonio Foncubierta-Rodriguez, Georg Langs and Allan Hanbury

Part IV  VISCERAL Anatomy Participant Reports

9  Automatic Atlas-Free Multiorgan Segmentation of Contrast-Enhanced CT Scans ................................. 145
Assaf B. Spanier and Leo Joskowicz

10 Multiorgan Segmentation Using Coherent Propagating Level Set Method Guided by Hierarchical Shape Priors and Local Phase Information .................................................... 165
Chunliang Wang and Örjan Smedby

11 Automatic Multiorgan Segmentation Using Hierarchically Registered Probabilistic Atlases ............................. 185
Razmig Kéchichian, Sébastien Valette and Michel Desvignes

12 Multiatlas Segmentation Using Robust Feature-Based Registration .......................................................... 203
Frida Fejne, Matilda Landgren, Jennifer Alvén, Johannes Ulén, Johan Fredriksson, Viktor Larsson, Olof Enqvist and Fredrik Kahl

Part V  VISCERAL Retrieval Participant Reports

13 Combining Radiology Images and Clinical Metadata for Multimodal Medical Case-Based Retrieval .................... 221
Oscar Jimenez-del-Toro, Pol Cirujeda and Henning Müller

14 Text- and Content-Based Medical Image Retrieval in the VISCERAL Retrieval Benchmark ............................ 237
Fan Zhang, Yang Song, Weidong Cai, Adrien Depeursinge and Henning Müller

Index .................................................................................................................................................. 251
Contributors

Jennifer Alvén
Department of Signals and Systems, Chalmers University of Technology, Gothenburg, Sweden
e-mail: alven@chalmers.se

Abdel Aziz Taha
Institute of Software Technology and Interactive Systems, TU Wien, Vienna, Austria
e-mail: taha@ifs.tuwien.ac.at

Weidong Cai
Biomedical and Multimedia Information Technology (BMIT) Research Group, School of Information Technologies, University of Sydney, Sydney, NSW, Australia
e-mail: tom.cai@sydney.edu.au

Pol Cirujeda
Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona, Spain
e-mail: pol.cirujeda@upf.edu

Adrien Depeursinge
University of Applied Sciences Western Switzerland (HES-SO), Sierre, Switzerland
e-mail: adrien.depeursinge@hevs.ch

Michel Desvignes
GIPSA-Lab, CNRS UMR 5216, Grenoble-INP, Université Joseph Fourier, Saint Martin d’Hères, France
Université Stendhal, Saint Martin d’Hères, France
e-mail: michel.desvignes@gipsa-lab.grenoble-inp.fr
Ivan Eggel
Institute for Information Systems, University of Applied Sciences
Western Switzerland (HES–SO Valais), Sierre, Switzerland
e-mail: ivan.eggel@hevs.ch

Olof Enqvist
Department of Signals and Systems, Chalmers University of Technology,
Gothenburg, Sweden
e-mail: olof.enqvist@chalmers.se

Frida Fejne
Department of Signals and Systems, Chalmers University of Technology,
Gothenburg, Sweden
e-mail: fejne@chalmers.se

Antonio Foncubierta-Rodríguez
Computer Vision Laboratory, Swiss Federal Institute of Technology (ETH) Zurich,
Zurich, Switzerland
e-mail: antonio.foncubierta@vision.ee.ethz.ch

Johan Fredriksson
Centre for Mathematical Sciences, Lund University, Lund, Sweden
e-mail: johanf@maths.lth.se

Orcun Goksel
Computer Vision Laboratory, Swiss Federal Institute of Technology (ETH) Zurich,
Zurich, Switzerland
e-mail: ogoksel@ethz.ch

Katharina Grünberg
University of Heidelberg, Heidelberg, Germany
e-mail: katharina.gruenberg@med.uni-heidelberg.de

Allan Hanbury
TU Wien, Institute of Software Technology and Interactive Systems,
Vienna, Austria
e-mail: allan.hanbury@tuwien.ac.at

Andrés Jakab
Medical University of Vienna, Vienna, Austria
e-mail: andras.jakab@meduniwien.ac.at

Oscar Jimenez-del-Toro
Institute of Information Systems, University of Applied Sciences
Western Switzerland Sierre (HES-SO), Sierre, Switzerland
e-mail: oscar.jimenez@hevs.ch
Leo Joskowicz
The Rachel and Selim Benin School of Computer Science and Engineering,
The Hebrew University of Jerusalem, Jerusalem, Israel
e-mail: leo.josko@mail.huji.ac.il

Fredrik Kahl
Department of Signals and Systems, Chalmers University of Technology,
Gothenburg, Sweden
Centre for Mathematical Sciences, Lund University, Lund, Sweden
e-mail: fredrik.kahl@chalmers.se

Razmig Kéchichian
CREATIS, CNRS UMR5220, Inserm U1044, INSA-Lyon,
Université de Lyon, Lyon, France
Université Claude Bernard Lyon 1, Lyon, France
e-mail: razmig.kechichian@creatis.insa-lyon.fr

Markus Krenn
Medical University of Vienna, Vienna, Austria
e-mail: markus.krenn@meduniwien.ac.at

Matilda Landgren
Centre for Mathematical Sciences, Lund University, Lund, Sweden
e-mail: matilda@maths.lth.se

Georg Langs
Medical University of Vienna, Vienna, Austria
e-mail: georg.langs@meduniwien.ac.at

Viktor Larsson
Centre for Mathematical Sciences, Lund University, Lund, Sweden
e-mail: viktorl@maths.lth.se

Henning Müller
Institute for Information Systems, University of Applied Sciences
Western Switzerland (HES–SO Valais), Sierre, Switzerland
University Hospitals and University of Geneva, Geneva, Switzerland
e-mail: henning.mueller@hevs.ch

Tomàs Salas Fernandez
Agencia D’Informació, Avaluació I Qualitat En Salut, Catalonia, Spain
e-mail: tomas.salas@gencat.cat

Roger Schaer
Institute for Information Systems, University of Applied Sciences
Western Switzerland (HES–SO Valais), Sierre, Switzerland
e-mail: roger.schaer@hevs.ch
Örjan Smedby  
Center for Medical Image Science and Visualization (CMIV),  
Linköping University, Linköping, Sweden  
Department of Radiology and Department of Medical and Health Sciences,  
Linköping University, Linköping, Sweden  
School of Technology and Health (STH), KTH Royal Institute of Technology,  
Stockholm, Sweden  
e-mail: orjan.smedby@sth.kth.se

Yang Song  
Biomedical and Multimedia Information Technology (BMIT)  
Research Group, School of Information Technologies, University of Sydney,  
Sydney, NSW, Australia  
e-mail: yang.song@sydney.edu.au

Assaf B. Spanier  
The Rachel and Selim Benin School of Computer Science and Engineering,  
The Hebrew University of Jerusalem, Jerusalem, Israel  
e-mail: assaf.spanier@mail.huji.ac.il

Johannes Ulén  
Department of Signals and Systems, Chalmers University of Technology,  
Gothenburg, Sweden  
e-mail: ulen@maths.lth.se

Sébastien Valette  
CREATIS, CNRS UMR5220, Inserm U1044, INSA-Lyon,  
Université de Lyon, Lyon, France  
Université Claude Bernard Lyon 1, Lyon, France  
e-mail: sebastien.valette@creatis.insa-lyon.fr

Chunliang Wang  
Center for Medical Image Science and Visualization (CMIV),  
Linköping University, Linköping, Sweden  
Department of Radiology and Department of Medical and Health Sciences,  
Linköping University, Linköping, Sweden  
School of Technology and Health (STH), KTH Royal Institute of Technology,  
Stockholm, Sweden  
e-mail: chunliang.wang@liu.se

Marc-André Weber  
University of Heidelberg, Heidelberg, Germany  
e-mail: marcandre.weber@med.uni-heidelberg.de

Marianne Winterstein  
University of Heidelberg, Heidelberg, Germany  
e-mail: marianne.winterstein@med.uni-heidelberg.de
Fan Zhang
Biomedical and Multimedia Information Technology (BMIT) Research Group,
School of Information Technologies, University of Sydney, Sydney,
NSW, Australia
e-mail: fzha8048@uni.sydney.edu.au
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>API</td>
<td>Application programming interface</td>
</tr>
<tr>
<td>BoVW</td>
<td>Bag of Visual Words</td>
</tr>
<tr>
<td>bpref</td>
<td>Binary preference</td>
</tr>
<tr>
<td>CAD</td>
<td>Computer-aided diagnosis</td>
</tr>
<tr>
<td>CECT</td>
<td>Contrast-enhanced CT</td>
</tr>
<tr>
<td>CLEF</td>
<td>Conference and Labs of the Evaluation Forum</td>
</tr>
<tr>
<td>CT</td>
<td>Computed tomography</td>
</tr>
<tr>
<td>Ctce</td>
<td>Contrast-enhanced computed tomography image</td>
</tr>
<tr>
<td>CVT</td>
<td>Centroidal Voronoi tessellation</td>
</tr>
<tr>
<td>DICOM</td>
<td>Digital Imaging and Communications in Medicine</td>
</tr>
<tr>
<td>EM</td>
<td>Expectation–maximization</td>
</tr>
<tr>
<td>EU</td>
<td>European Union</td>
</tr>
<tr>
<td>GM-MAP</td>
<td>Geometric mean average precision</td>
</tr>
<tr>
<td>HU</td>
<td>Hounsfield unit</td>
</tr>
<tr>
<td>IDF</td>
<td>Inverse document frequency</td>
</tr>
<tr>
<td>IRB</td>
<td>Internal review board</td>
</tr>
<tr>
<td>ISBI</td>
<td>International Symposium on Biomedical Imaging</td>
</tr>
<tr>
<td>k-NN</td>
<td>k-nearest neighbour</td>
</tr>
<tr>
<td>MAP</td>
<td>Mean average precision</td>
</tr>
<tr>
<td>MEC</td>
<td>Medical ethics committee</td>
</tr>
<tr>
<td>MR</td>
<td>Magnetic resonance</td>
</tr>
<tr>
<td>MRI</td>
<td>Magnetic resonance imaging</td>
</tr>
<tr>
<td>MRT1</td>
<td>Magnetic resonance T1-weighted image</td>
</tr>
<tr>
<td>MRT1cefs</td>
<td>Contrast-enhanced fat-saturated magnetic resonance T1-weighted image</td>
</tr>
<tr>
<td>MRT2</td>
<td>Magnetic resonance T2-weighted image</td>
</tr>
<tr>
<td>NIfTI</td>
<td>Neuroimaging Informatics Technology Initiative</td>
</tr>
<tr>
<td>NMI</td>
<td>Normalized mutual information</td>
</tr>
<tr>
<td>OS</td>
<td>Operating system</td>
</tr>
<tr>
<td>P10</td>
<td>Precision after 10 cases retrieved</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>P30</td>
<td>Precision after 30 cases retrieved</td>
</tr>
<tr>
<td>PACS</td>
<td>Picture archiving and communication systems</td>
</tr>
<tr>
<td>PCA</td>
<td>Principal component analysis</td>
</tr>
<tr>
<td>pLSA</td>
<td>Probabilistic Latent Semantic Analysis</td>
</tr>
<tr>
<td>QC</td>
<td>Quality control</td>
</tr>
<tr>
<td>RadLex</td>
<td>Radiology Lexicon</td>
</tr>
<tr>
<td>RANSAC</td>
<td>Random sample consensus</td>
</tr>
<tr>
<td>ROI</td>
<td>Region of interest</td>
</tr>
<tr>
<td>SIFT</td>
<td>Scale-invariant feature transform</td>
</tr>
<tr>
<td>SIMPLE</td>
<td>Selective and iterative method for performance level estimation</td>
</tr>
<tr>
<td>SURF</td>
<td>Speeded Up Robust Features</td>
</tr>
<tr>
<td>TF</td>
<td>Term frequency</td>
</tr>
<tr>
<td>TREC</td>
<td>Text Retrieval Conference</td>
</tr>
<tr>
<td>URL</td>
<td>Uniform resource locator</td>
</tr>
<tr>
<td>VISCERAL</td>
<td>Visual Concept Extraction Challenge in Radiology</td>
</tr>
<tr>
<td>VM</td>
<td>Virtual machine</td>
</tr>
</tbody>
</table>