VISCERAL Retrieval Benchmark

Guidelines for Participation v1.2

Document History

v1.0 - 20141110 - Initial version of document
v1.1 - 20141118 - Initial draft ready for release
v1.2 - 20141123 - Updated Section 3.2 to include the query scenarios

NOTE: This document covers the final submission for the VISCERAL Retrieval Benchmark, due on February 28, 2015. Submission instructions for those wishing to participate in the related Multimodal Retrieval in the Medical Domain 2015 Workshop (MRMD 2015) are on this page: http://www.visceral.eu/benchmarks/mrmd-2015/

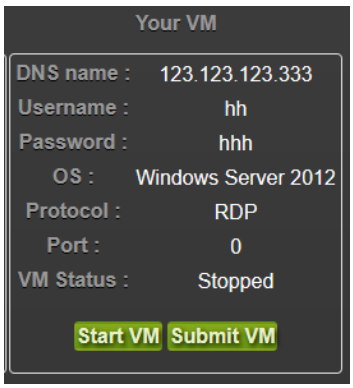
1. Introduction

1.1 Registration

The first step in participation is registration. This is done online on the following page: http://visceral.eu:8080/register/Registration.xhtml
During the registration process, participants will be required to sign and upload a participation agreement.
Once the participant is registered, logging into the registration system will reveal the *participant dashboard*.

1.2 Virtual Machine

The medical imaging data is stored on the Microsoft Azure Cloud. When participants register successfully, they will receive a virtual machine (VM) in the Microsoft Azure cloud (Windows or Linux VMs are available), provided and financed by VISCERAL, with the support of Microsoft Research. The information for accessing the VM will appear on the participant dashboard in the registration system (after a delay of up to a week).

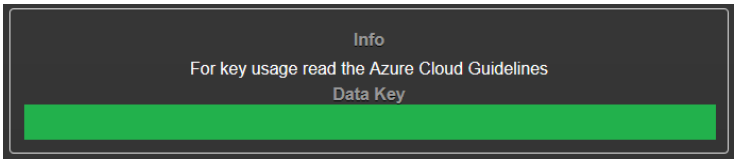


Note that each participating group should only register for the VISCERAL Benchmark once. After successful registration, the person completing the registration will get root access to the assigned VM, and will be able to create logins for colleagues.
Please shut down the VM if it will not be used for a longer time (using the “Stop VM” button on the participant dashboard in the registration system). It can again be started using the “Start VM” button.

Documents about the Benchmark, including information on using the VM and a list of volumes can be found by pressing View Benchmark Files.

1.3 Training Data

The training data can be accessed from the VM. The *Data Key* for accessing the data is provided on the participant dashboard in the registration system once the VM is assigned. Please only access the data on the cloud from within the assigned VM — accessing the data from outside the cloud results in additional costs for the organisers.



A list of all files in the training data set can be downloaded from the participant dashboard in the registration system by clicking on 

.

The image file URLs are constructed as:

`cURL+filename+saKey`

cURL: container URL,
`http://visceralstorage1.blob.core.windows.net/trainingset/`
filename: `PatientID_ModalityCounter_ModalityName_RegionName.nii.gz`
saKey: shared access key, e.g.
`?sr=c&si=readonly&sig=Z69O9Vz8TU0RxawtASpmpWZnT%2FhF2OgJOI7iEt60mis%3D`

1.3.1 Training Data downloads outside the cloud

For full training set downloads, an FTP access will be provided upon request. Please send an email for further information to oscar.jimenez@hevs.ch CC: ivan.eggel@hevs.ch

1.4 Further Information

Detailed information on the evaluation protocol and goals can be found in VISCERAL Deliverable 4.2[1].
Further information on the dataset and the file formats can be found in VISCERAL Deliverable 2.2.2[2].

All participants and organisers are automatically registered to the participants-retrieval_benchmark@visceral.eu mailing list, and can post on the mailing list. Use this list to communicate only among participants and the organisers, to ask questions, draw attention to problems or share hints and tips.

A LinkedIn group has been set-up for discussion about the Benchmark. Ask questions and make comments on this group:
http://www.linkedin.com/groups/VISCERAL-Benchmark-Discussion-5089631

[1] VISCERAL Deliverable 4.2:
http://www.visceral.eu/assets/Uploads/Deliverables/VISCERAL-D4.2.pdf
[2] VISCERAL Deliverable 2.2.2:
http://www.visceral.eu/assets/Uploads/Deliverables/VISCERAL-D2.2.2.pdf

You can also follow VISCERAL on Twitter: @VisceralEU.

2. VISCERAL Retrieval dataset

The VISCERALretrieval data set consists 2311 volumes originated from three different modalities (CT,MRT1,MRT2). For a subset of these volumes we provide from the volume's radiology report extracted anatomy-pathology terms in form of csv files. The following table gives an overview of the dataset in which a participant should perform the retrieval task.

Modality	**Bodyregion**	**Volumes**	**available A-P Term Files**
CT	Ab	336	213
CT	Th	971	699
CT	ThAb	84	84
CT	unknown	211	211
CT	Wb	412	412
MRT1	Ab	167	114
MRT1	unknown	24	24
MRT2	Ab	68	18
MRT2	unknown	38	38
TOTAL		**2311**	**1813**

The volumes are available in NIFTI file format with the following naming convention:
<volumeID>_<modality>_<bodyregion>.nii.gz

Anatomy - pathology term files are named accordingly as
<volumeID>_<modality>_<bodyregion>.csv

An anatomy-pathology term file lists pathological terms that occur in the report of a volume together with its anatomy. Both entities are described textually and additionally with their corresponding Radlex ID (RID). Radlex is a unified language of radiology terms that can be used for standardized indexing and retrieval of radiology information resources. Detailed information about the ontology of Radlex can be found at http://rsna.org/RadLex.aspx.
Please note that a term file lists both, occurring and explicitly in the report negated pathologies. In the following we give an example to illustrate the naming convention and the file content specification.

```
VolumeID_Modality_Bodyregion.fcsv
i.e. 123456_MRT1_Ab.csv
```

```
# The first row depicts the header of an a-p term file.
```

```
Anatomy RID,Anatomy,Pathology RID,Pathology,Negated
#
# Each row entry in this file contains an occurring pathology and its anatomy and
states if the pathology is negated.
RID199,Ductus choledochus,RID4865,Ödem,1
RID58,Leber,RID3874,Raumforderung,0
```

This file would indicate there is **no "Ödem" in "ductus choledochus"** whereas there occurs **"Raumforderung" in the anatomy "Leber"**.

3. Content-based medical image retrieval

3.1 Task description

In this task, we evaluate the retrieval of relevant cases based on a query case. It serves the following scenario: a user is assessing a query case in a clinical setting, e.g., a CT volume, and is searching for cases that are relevant in this assessment. The algorithm has to find cases that are relevant in a large database of cases. For each topic (query case) there is:

- the patient 3D imaging data (CT, MRI)*
- 3D bounding box region of interest containing the radiological signs of the pathology*
- binary mask of the main organ affected*
- radiologic report extracted anatomy-pathology terms in form of csv files.

* Volumes in NIFTI file format: .nii.gz

The participants have to develop an algorithm that finds clinically-relevant (related) cases given a query case (imaging and text data), but not the final diagnosis.

3.2 Evaluation

For each topic, the algorithms should generate a ranked list of search results out of the VISCERAL Retrieval dataset (containing imaging data and text data). A set of 10 test query cases (topics) will be used to evaluate the result rankings of the algorithms.

The data base contains both imaging data and corresponding text data. There are two query scenarios:

- Image data and ROI available for the query case
- Image data, ROI and text data available for the query case.

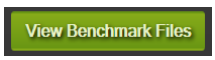
While the first case is of immediate clinical relevance, we expect also the second case to be valuable in evaluating specific retrieval algorithms. Therefore the use of text information is optional during retrieval evaluation.

The algorithms will be evaluated in two groups corresponding to the information used for the retrieval **[image+ROI]** or **[image+ROI+text]**.

Medical experts will perform relevance assessment of the top ranked cases by each approach, to judge the quality of retrieval. Experts will assess the relevance of the ranked cases. The evaluation measures that will be considered are the precision of the top-ranked X cases. The precision for top-ranked 10 and 30 cases (P@10, P@30), mean uninterpolated average precision (MAP), the bpref measure, and in cases where the establishing of the number of relevant cases in the entire data set is feasible, the R-precision will be evaluated.

3.3 Retrieval test queries download

The test data (retrieval test queries) can be accessed from the VM just like the training data. The *Data Key* for accessing the data is also the same provided on the participant dashboard in the registration system. Please only access the data on the cloud from within the assigned VM — accessing the data from outside the cloud results in additional costs for the organisers.

A list of all files in the test data set can be downloaded from the participant dashboard in the registration system by clicking on 

.

4. Result file Submission

After the submission deadline, pure text files with the names

1. `retrievalSetup_runID_participantID.txt`
2. `resultRanking_runID_participantID.txt`

must be in the home directory of the *azureuser* user (for those using Linux VMs), or on the Windows Desktop of the *azureuser* user (for those using Windows VMs) of the virtual machine.

4.1 Specification: Retrieval setup

`retrievalSetup_runID_ParticipantID.txt`

```
#
# Information regarding the retrieval setup used for generating the result
# ranking
#

 1.   Image, text or both? ……
 2.   Was other training data used? ……
 3.   Run type: Automatic, Manual, Interactive? ……
 4.   Query language? ……
```

TREC_EVAL will be used during the evaluation phase for the result ranking. TREC_EVAL is the standard tool used by the TREC community for evaluating an ad hoc retrieval run, given the results file and a standard set of judged results. The participant submissions should compliant with the trec_eval format prior to submission. Result rankings that do not meet the required format will be rejected. Program download and more information about TRECEVAL can be found here: http://trec.nist.gov/trec_eval/. Do not hesitate to ask if you have questions regarding the trec_eval format.

trec_eval format

The format for submitting results is based on the trec_eval program (http://trec.nist.gov/trec_eval/). It requires the following parameters to be arranged in columns separated by a tab:

Ranking output	
Parameter	**Explanation**
-topic	Run number within the topic (test query): 1-10
-iter	Iteration number, for the VISCERAL Retrieval benchmark this number should always be 1
-volumeID	VolumeID from the retrieval training dataset without the extension (.nii.gz or .fcsv) and without any image path
-rank	Ranking result within the dataset for the corresponding query case (1-300)
-simScore	Similarity score assigned by the participants algorithm to perform the ranking
-participantID	ParticipantID should be the same in the entire file

topic	iter	volumeID	rank	simScore	runID

4.2 Specification: Ranking results

```
resultRanking_runID_ParticipantID.txt
```

```
1          1          123563_MRT1_Ab          1          0.567162     visc123
1          1          652431_CT_Wb            2          0.441542     visc123
……         ……         ……                      ……         ……           ……
……         ……         ……                      ……         ……           ……
1          1          764393_CT_undefined     1000       0.035022     visc123
2          1          833887_MRT2_Ab          1          0.94753      visc123
……         ……         ……                      ……         ……           ……
……         ……         ……                      ……         ……           ……
10         1          4553266_CT_Wb           995        0.014921     visc123
```

……: Rest of the results in the ranking. Notice there are no spaces between rows.

4.3 Key points for submitted runs

- For each run, a single file as shown in 4.2, containing the ranking for all topics should be submitted.
- The final submission should include up to 10 run files containing each all the 10 topics (shown in 4.2) and the retrieval setup file (shown in 4.1).
- The score should be in decreasing order (i.e. the image at the top of the list should have a higher score than images at the bottom of the list).
- Up to (but not necessarily) 300 images can be submitted per query case.
- Files should be pure text files and not be zipped or otherwise compressed.