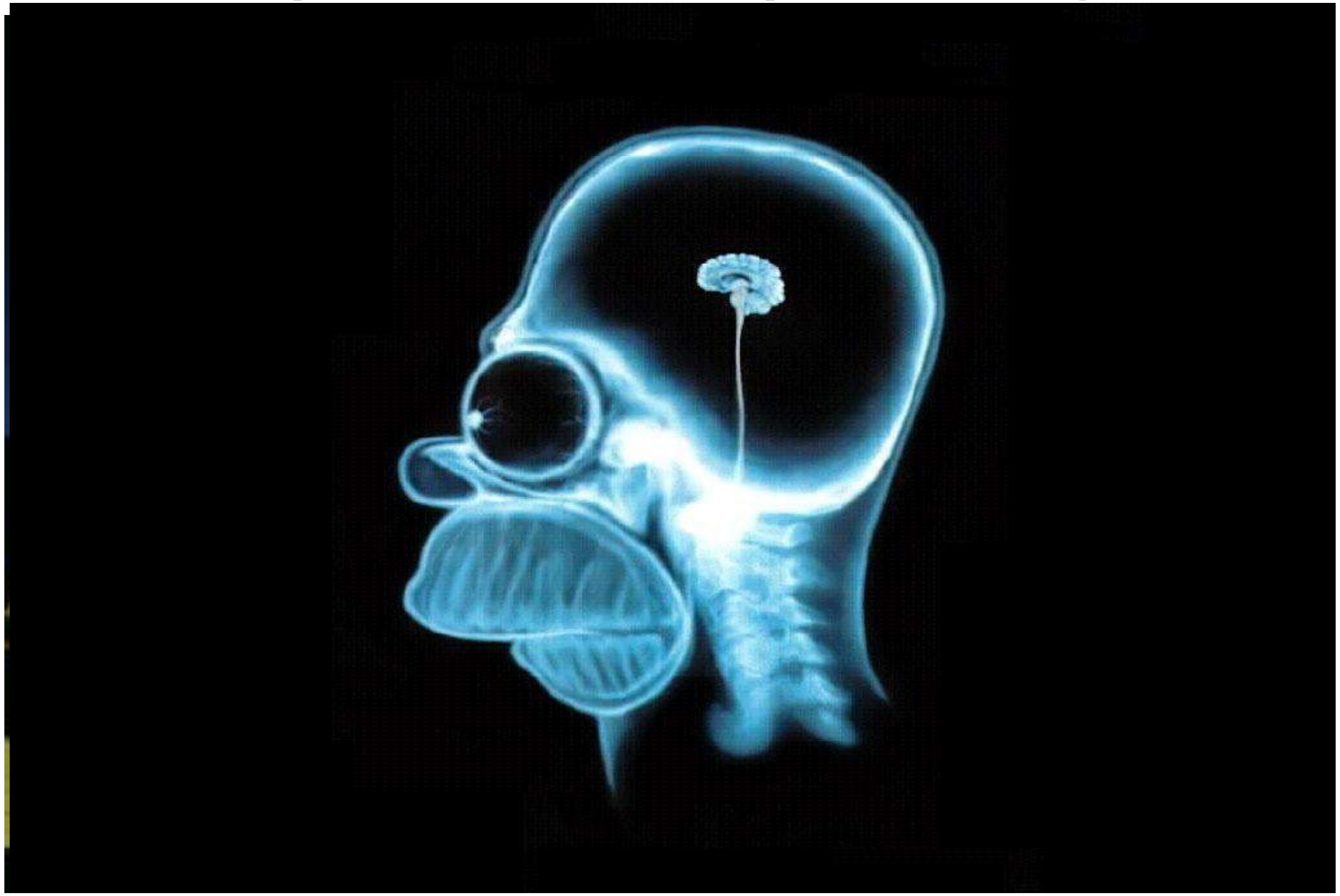
(Medical) image retrieval



Henning Müller
HES-SO

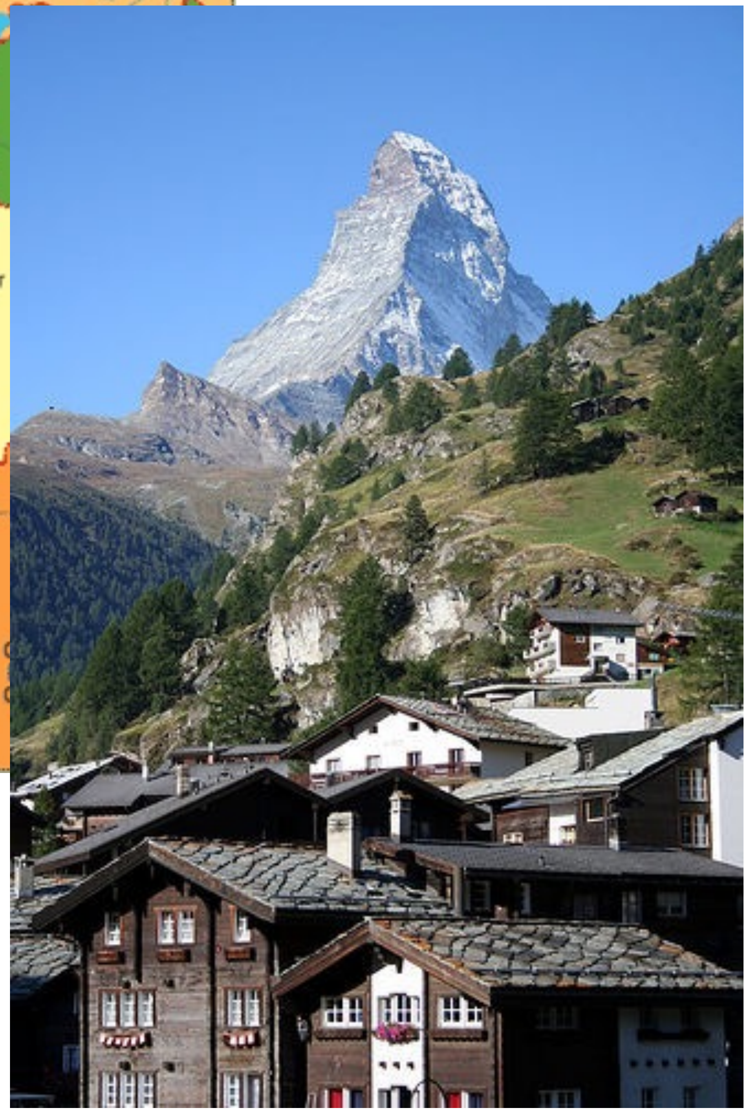
Visum Summer School 2021, Porto, Portugal

Who I am

- Medical informatics studies in Heidelberg, Germany (1992-1997)
  - Exchange with Daimler Benz research,
- PhD in CBIR, computer vision, Geneva, Switzerland (1998-2002)
  - Exchange with Monash University, Melbourne, AUS
- Professor in radiology and medical informatics at the University of Geneva (2014-)
- Professor in Computer Science at the HES-SO, Sierre, Switzerland (2007-)
  - Visiting faculty at Martinos Center (2015-2016)
- Member of the Swiss National Research Council

Where I am


N
Switzerland
GERMANY
FRANCE
SWITZERLAND
ITALY
LIECH.
Basell
Baden
Winterthur
Zürich
St.Gallen
Horgen
Biel
Lucerne
(Luzern)
Burghdorf
Neuchâtel
Bern
Steffisburg
Fribourg
Thun
Chur
Interlaken
Lausanne
Montreux
Sierre
Geneva
(Genéve)
Martigny
Champex
Locarno
8°
©maps.com



Where I am

Photo courtesy of Laurent Borella

A few questions

- Who works on medical data?
- Who has already worked on image retrieval?
- Any specific expectation?

Timing

- 90 minutes presentation
- 15 minutes break
- 45 minutes hands on (run by the Visum team)

Objectives of the presentation

- Give a historic view of how (medical) image retrieval evolved
  - With references and an outlook into the future
- Explain a few concrete applications of (medical) image retrieval
- Detail existing resources for medical imaging
  - For retrieval and other applications
  - From public or private repositories
  - From scientific challenges
  - Introducing Evaluation as a Service (EaaS)
- Motivate you to do research in this field!

A few general review papers

- Eakins, J. and Graham, M., 1999. Content-based image retrieval.
- Smeulders, A.W., Worring, M., Santini, S., Gupta, A. and Jain, R., 2000. Content-based image retrieval at the end of the early years. IEEE Transactions on pattern analysis and machine intelligence, 22(12), pp.1349-1380.
- Müller, H., Müller, W., Squire, D.M., Marchand-Maillet, S. and Pun, T., 2001. Performance evaluation in content-based image retrieval: overview and proposals. Pattern recognition letters, 22(5), pp.593-601.
- Zhou, X.S. and Huang, T.S., 2003. Relevance feedback in image retrieval: A comprehensive review. Multimedia systems, 8(6), pp.536-544.
- Datta, R., Joshi, D., Li, J. and Wang, J.Z., 2008. Image retrieval: Ideas, influences, and trends of the new age. ACM Computing Surveys (Csur), 40(2), pp.1-60.
- Wan, J., Wang, D., Hoi, S.C.H., Wu, P., Zhu, J., Zhang, Y. and Li, J., 2014, November. Deep learning for content-based image retrieval: A comprehensive study. In Proceedings of the 22nd ACM international conference on Multimedia (pp. 157-166).
- Yan, C., Gong, B., Wei, Y. and Gao, Y., 2020. Deep multi-view enhancement hashing for image retrieval. IEEE Transactions on Pattern Analysis and Machine Intelligence, 43(4), pp.1445-1451.

A few medical review articles

- Tang, L. H., Hanka, R., & Ip, H. H. (1999). A review of intelligent content-based indexing and browsing of medical images. Health Informatics Journal, 5(1), 40-49.
- Müller, H., Michoux, N., Bandon, D., & Geissbuhler, A. (2004). A review of content-based image retrieval systems in medical applications—clinical benefits and future directions. International journal of medical informatics, 73(1), 1-23.
- Akgül, C. B., Rubin, D. L., Napel, S., Beaulieu, C. F., Greenspan, H., & Acar, B. (2011). Content-based image retrieval in radiology: current status and future directions. Journal of digital imaging, 24(2), 208-222.
- Hwang, K. H., Lee, H., & Choi, D. (2012). Medical image retrieval: past and present. Healthcare informatics research, 18(1), 3-9.
- Kumar, A., Kim, J., Cai, W., Fulham, M., & Feng, D. (2013). Content-based medical image retrieval: a survey of applications to multidimensional and multimodality data. Journal of digital imaging, 26(6), 1025-1039.
- Müller, H., & Unay, D. (2017). Retrieval from and understanding of large-scale multi-modal medical datasets: a review. IEEE transactions on multimedia, 19(9), 2093-2104.
- Li, Z., Zhang, X., Müller, H., & Zhang, S. (2018). Large-scale retrieval for medical image analytics: A comprehensive review. *Medical image analysis*, *43*, 66-84.

Information retrieval

- Source Wikipedia:
  - **Information retrieval** (**IR**) is the process of obtaining information system resources that are relevant to an information need from a collection of those resources. Searches can be based on full-text or other content-based indexing. Information retrieval is the science of searching for information in a document, searching for documents themselves, and also searching for the metadata that describes data, and for databases of texts, images or sounds.

Retrieval vs. classification

- Retrieval is to fulfil an information need with a relevance definition (no fixed classes, browsing)
  - Subjective, user- and time-dependent
- Classification means that a fixed number of classes exist (exclusive or non-exclusive)
  - Can be linked to object recognition
- Localization means to find the region in an image (with an object, a lesion, …)
- Detection means to find whether a pattern is present or not
- Recognition, …

Visual Question Answering (VQA)

- Specific questions regarding an existing image
  - In this case on medical images
- Relies on the availability of images and training data with questions and answers to learn models that generalize to new questions
  - Much more complex than textual QA
- 2018-2021 had a task on medical VQA in ImageCLEF
  - Interest was strong for a challenging task
    - 8-15 participants each year

Challenges with CBIR

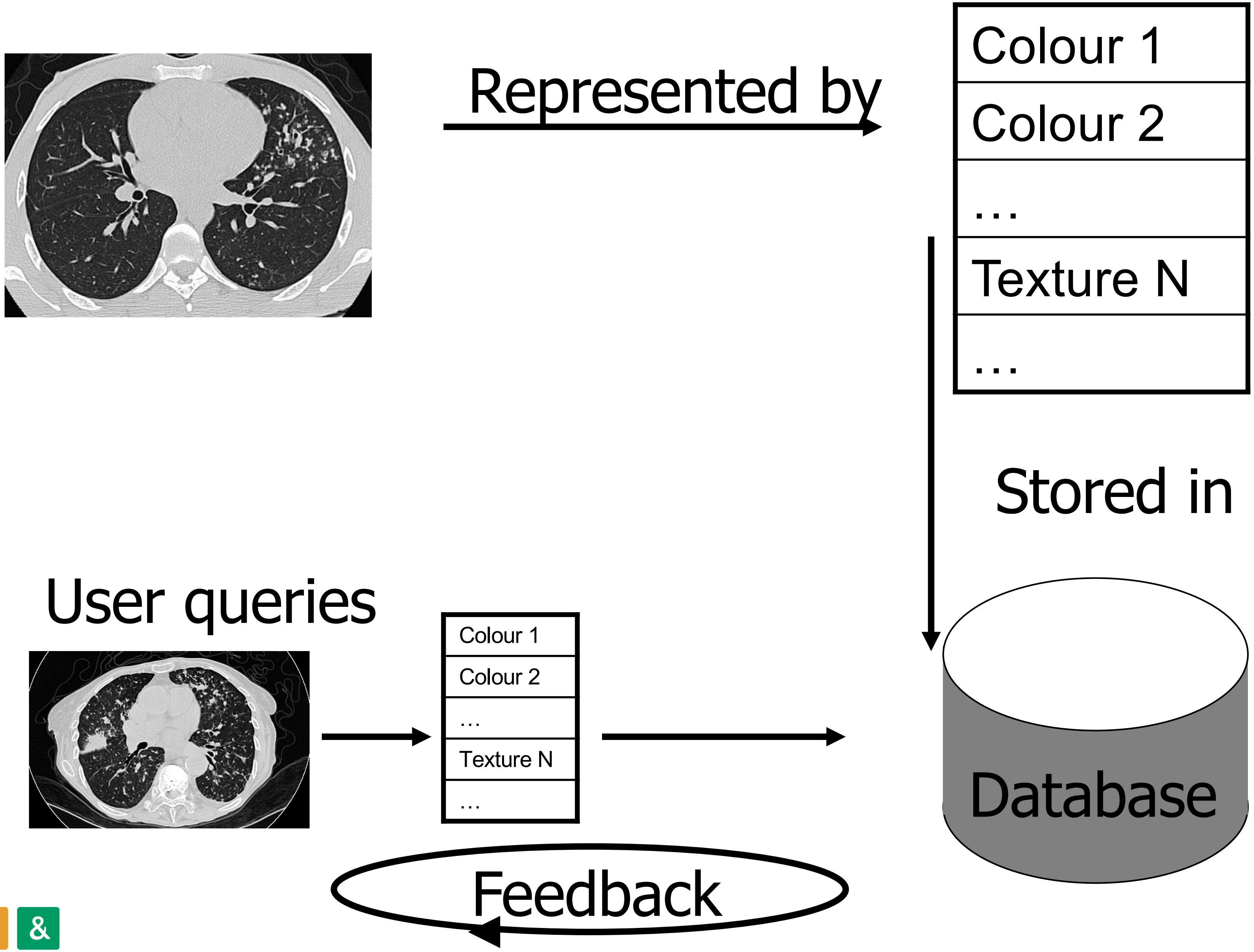
- Page zero problem
  - For visual retrieval you first need a starting image
- Sensory gap
  - When creating an image only a limited part of reality is captured (limited resolution, occlusion, …)
- Semantic gap
  - There is a mismatch between the semantic information that a user searches with and the low-level features most commonly used (colors, texture)
- Many other gaps have been described

What is in the image and what is it about?

How can we search (unit)?

- Search with an entire case (possibly several images, text, structured data)
  - Closest to what clinicians might require
- Search with full image example (or volume in case of tomographic data)
  - The actual region of interest is usually really small
- Search with an image region (ROI)
  - Find what an unusual pattern may be linked to
- Search inside a large image
  - For example in histopathology images that are extremely large

Simplified system overview

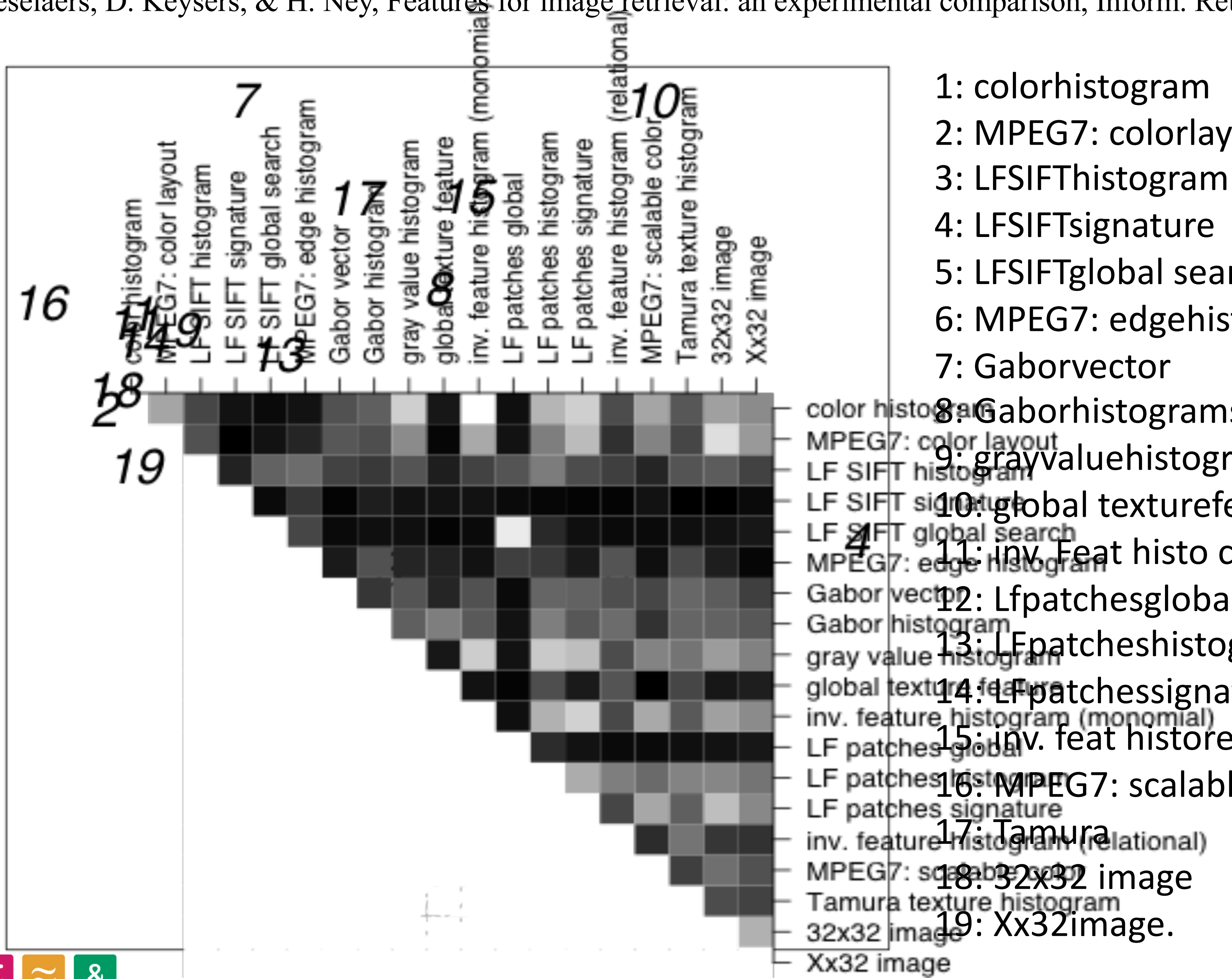


Types of visual features

- Handcrafted vs. partially learned vs. fully learned
  - Deep learning vs. traditional approaches
- Classifications of visual features
  - Low level vs. mid level vs. semantic/high-level
    - Higher levels via feature modeling (visual words) or latent semantic techniques, sometimes matching words and pictures
- Type of information that is modeled
  - Shape vs. grey level/color vs. texture
- Local vs. global features
    - Local based on segmentation or partitioning
- 2D vs. 3D vs. nD (3D +time, protocols)

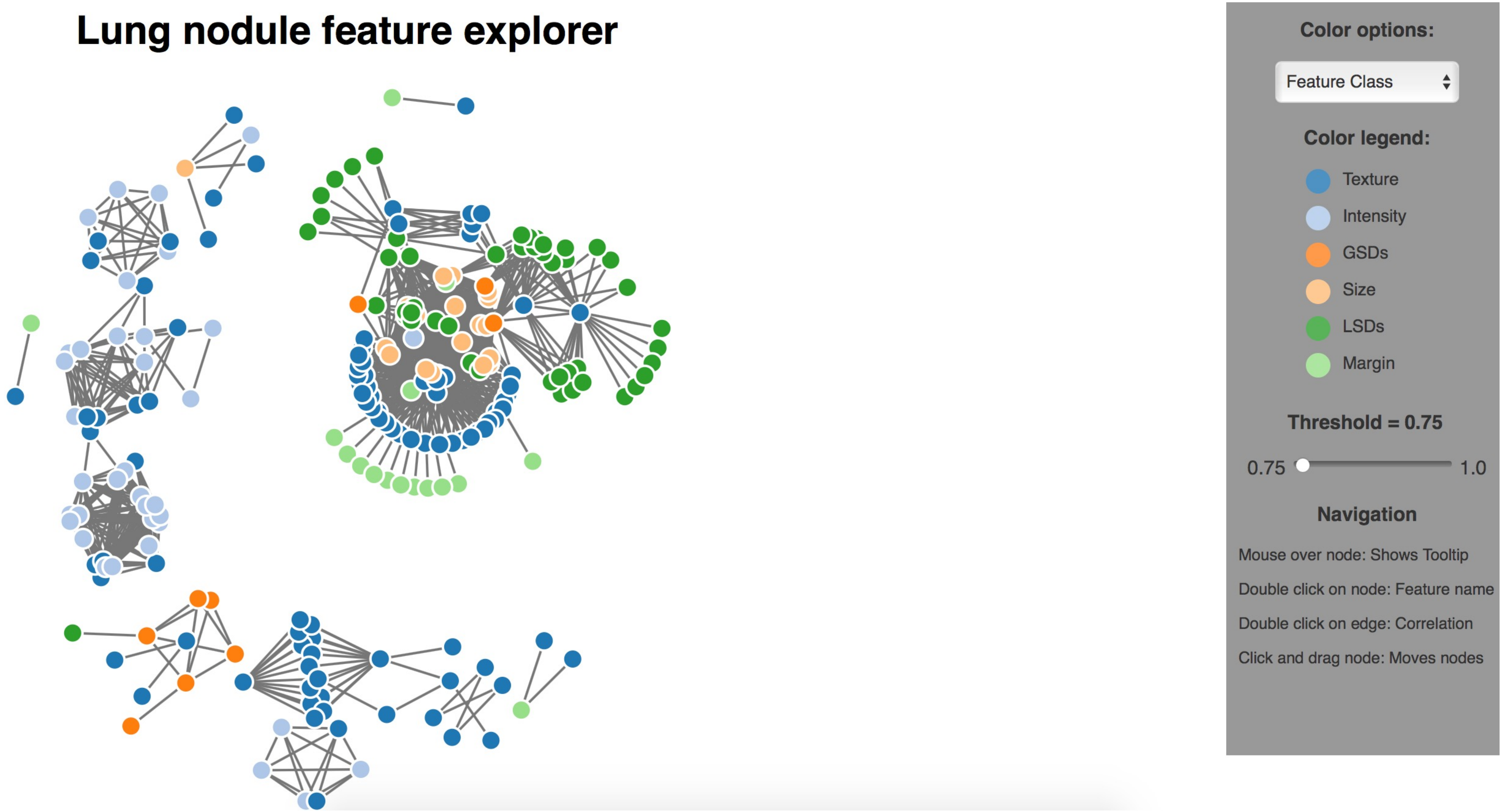
Correlation between Features

T. Deselaers, D. Keysers, & H. Ney, Features for image retrieval: an experimental comparison, Inform. Retrieval (2008) 11: 77



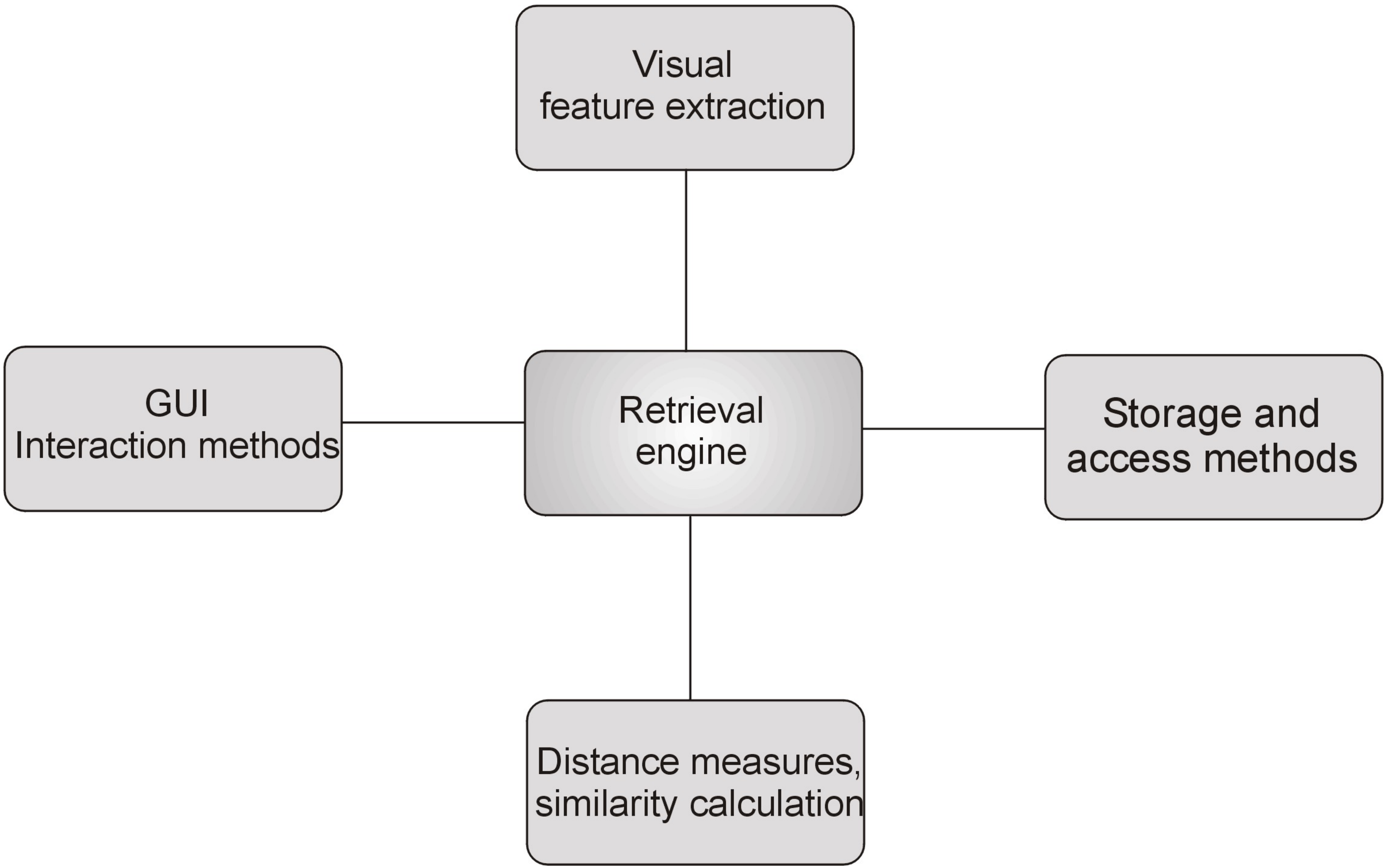
1: colorhistogram
2: MPEG7: colorlayout
3: LFSIFThistogram
4: LFSIFTsignature
5: LFSIFTglobal search
6: MPEG7: edgehistogram,
7: Gaborvector
8: Gaborhistograms
9: grayvaluehistogram
10: global texturefeature,
11: inv. Feat histo color
12: Lfpatchesglobal
13: LFpatcheshistogram
14: LFpatchessignature,
15: inv. feat historel
16: MPEG7: scalablecolor,
17: Tamura
18: 32x32 image
19: Xx32image.

Feature exploration



http://lung-nodule-feature-explorer.s3-website-us-east-1.amazonaws.com

Retrieval system components

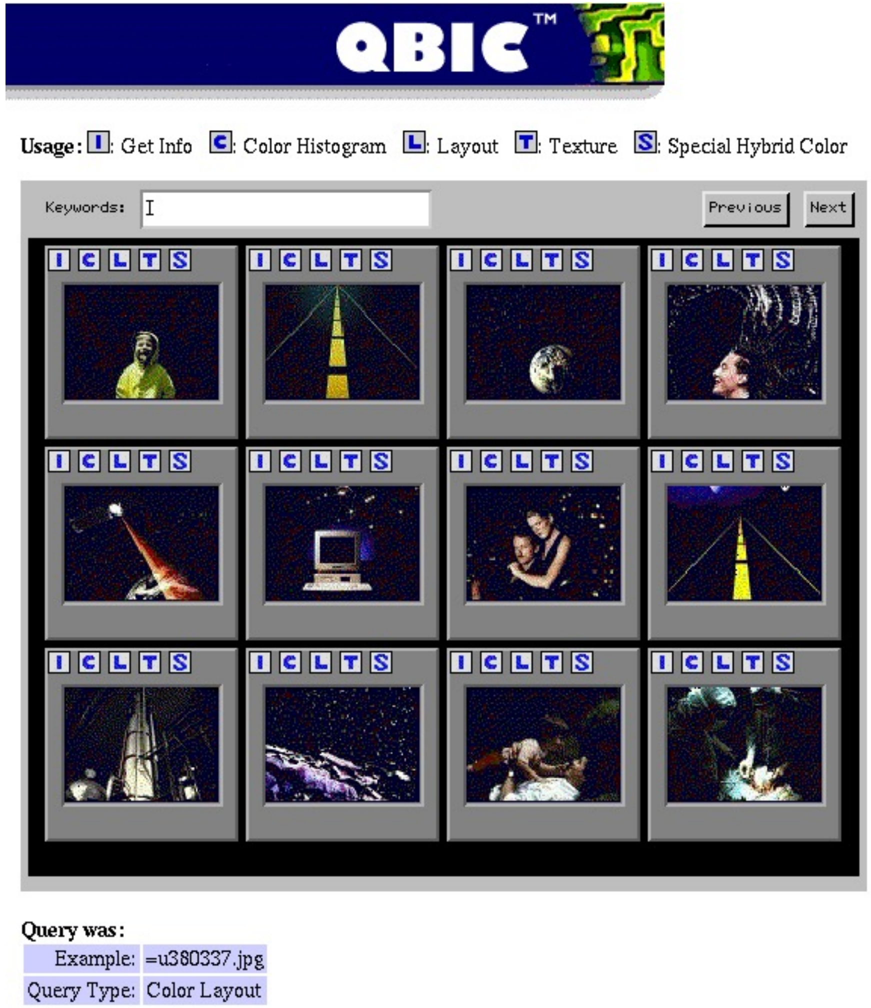


Perception of visual retrieval

Source: ct magazine

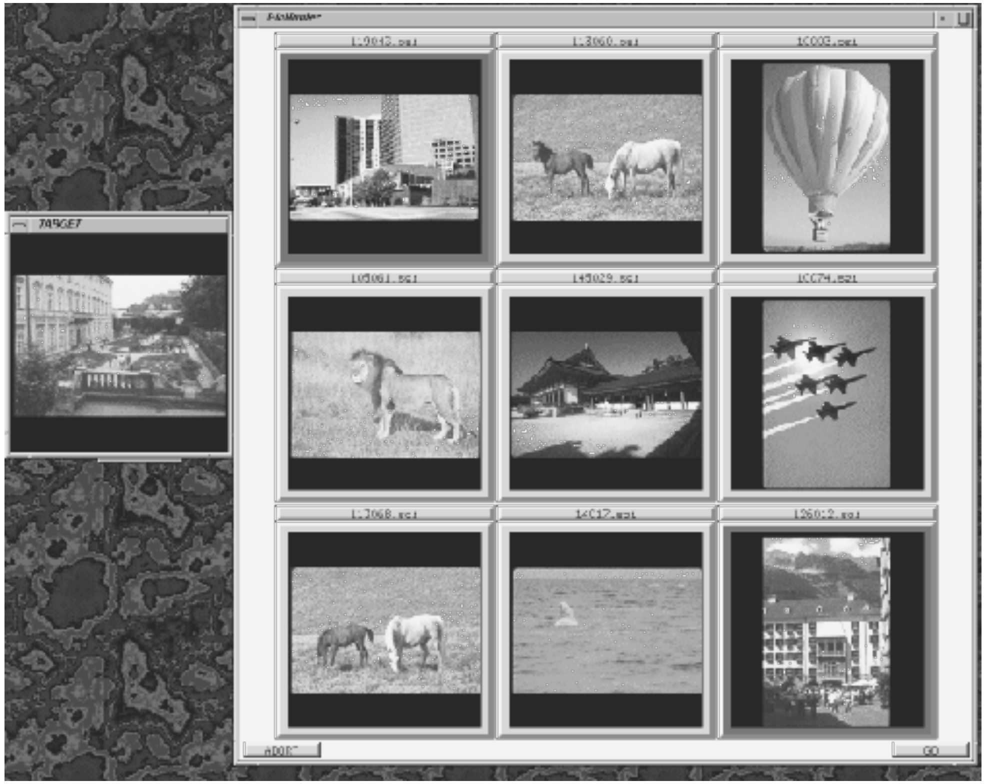
QBIC – Query by image content

- IBM, commercial product, 1993
- Add on for DB2
- Simple color, texture, layout features
- Very simple feedback

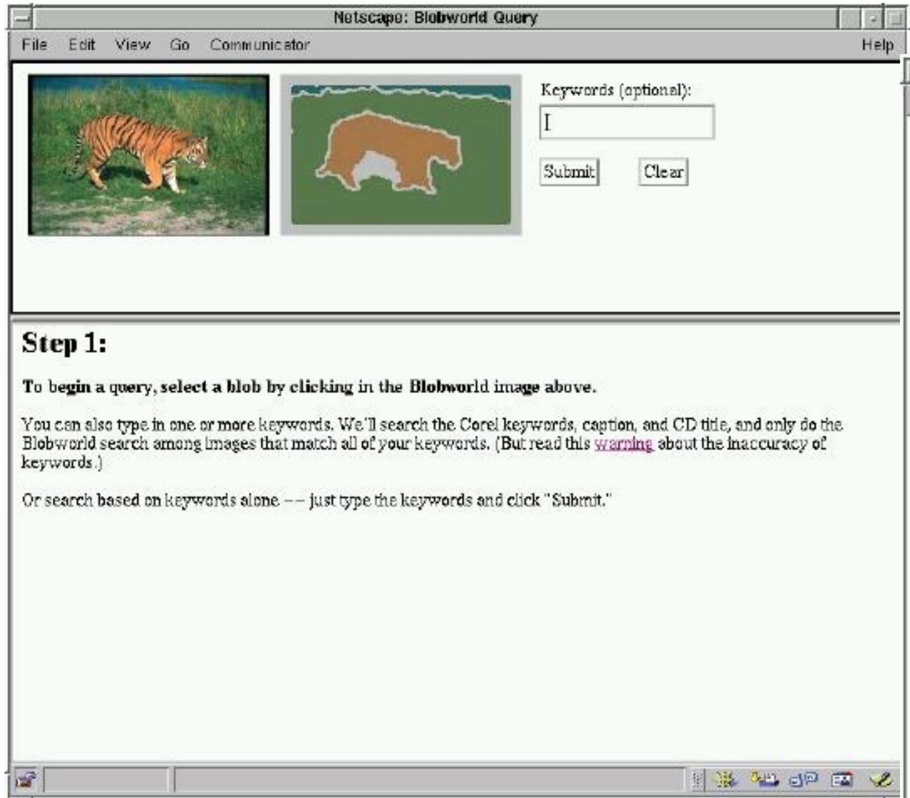


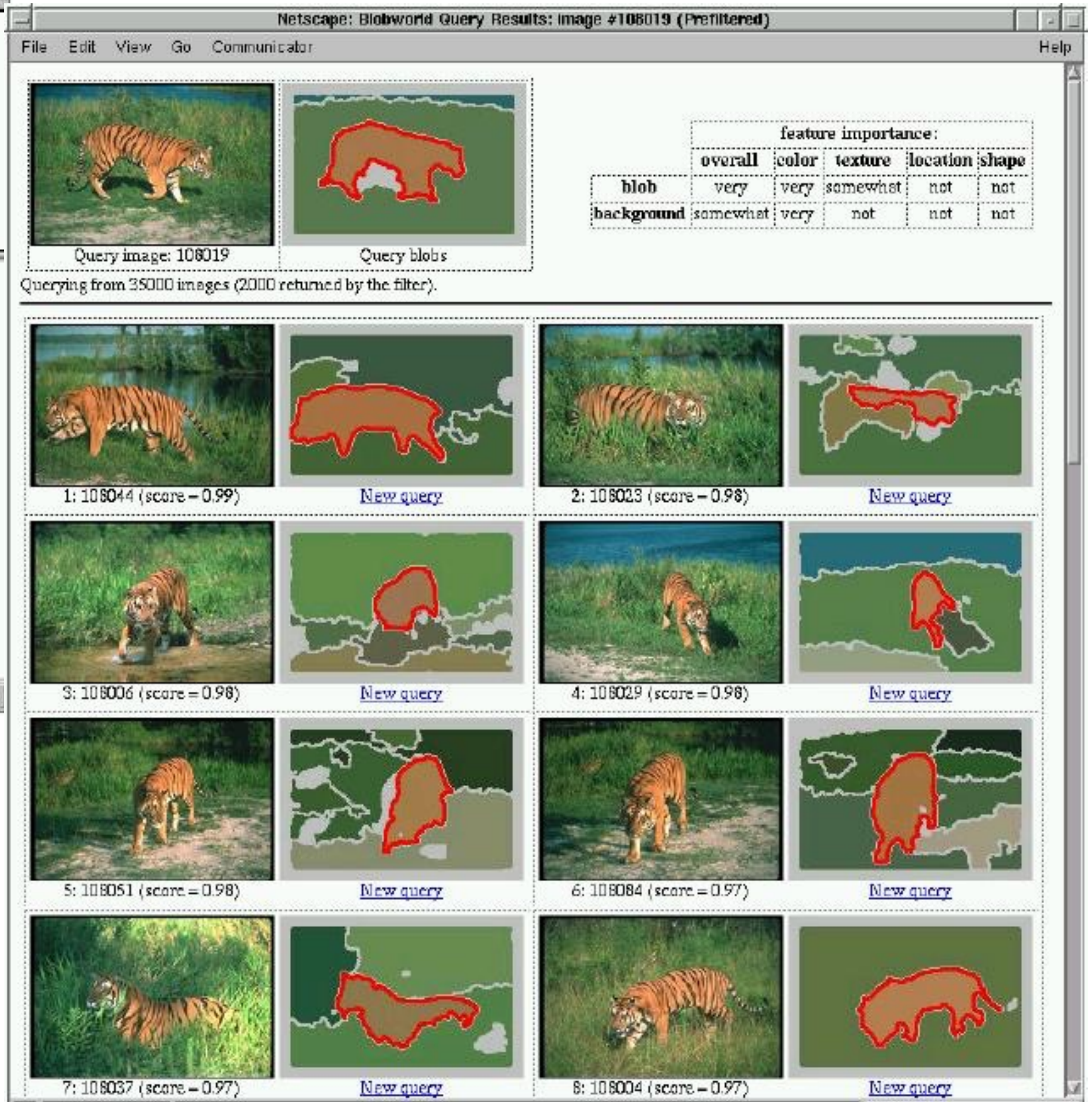
PicHunter, 1996

- Browsing tool for target image search
- Optimizes information gain from images being presented to the user (Bayesian method)
  - Other systems use filter image browsing



Blobworld, 1997



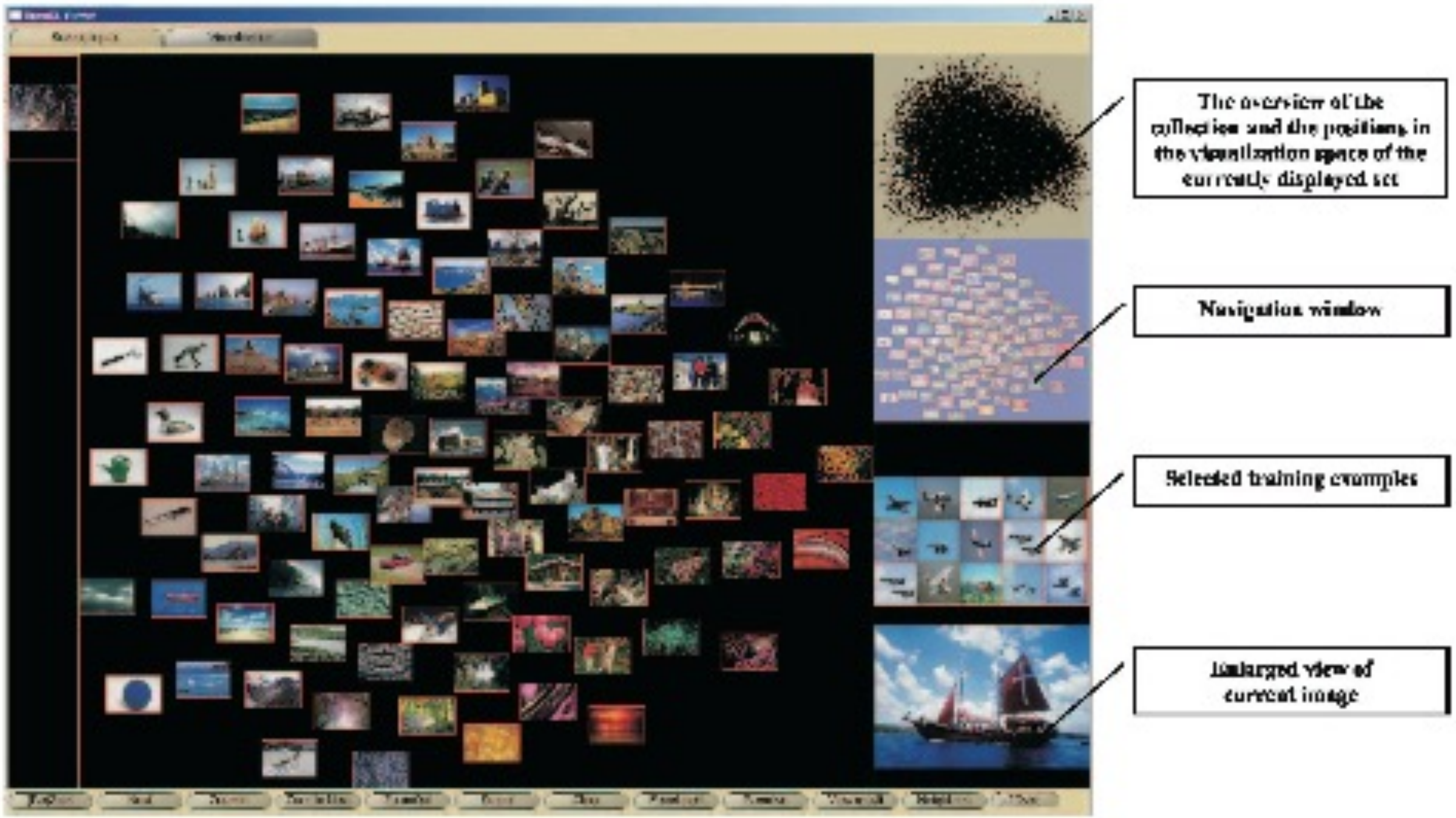


FIRE, 2005

- Open-source academic system
- Simple relevance feedback

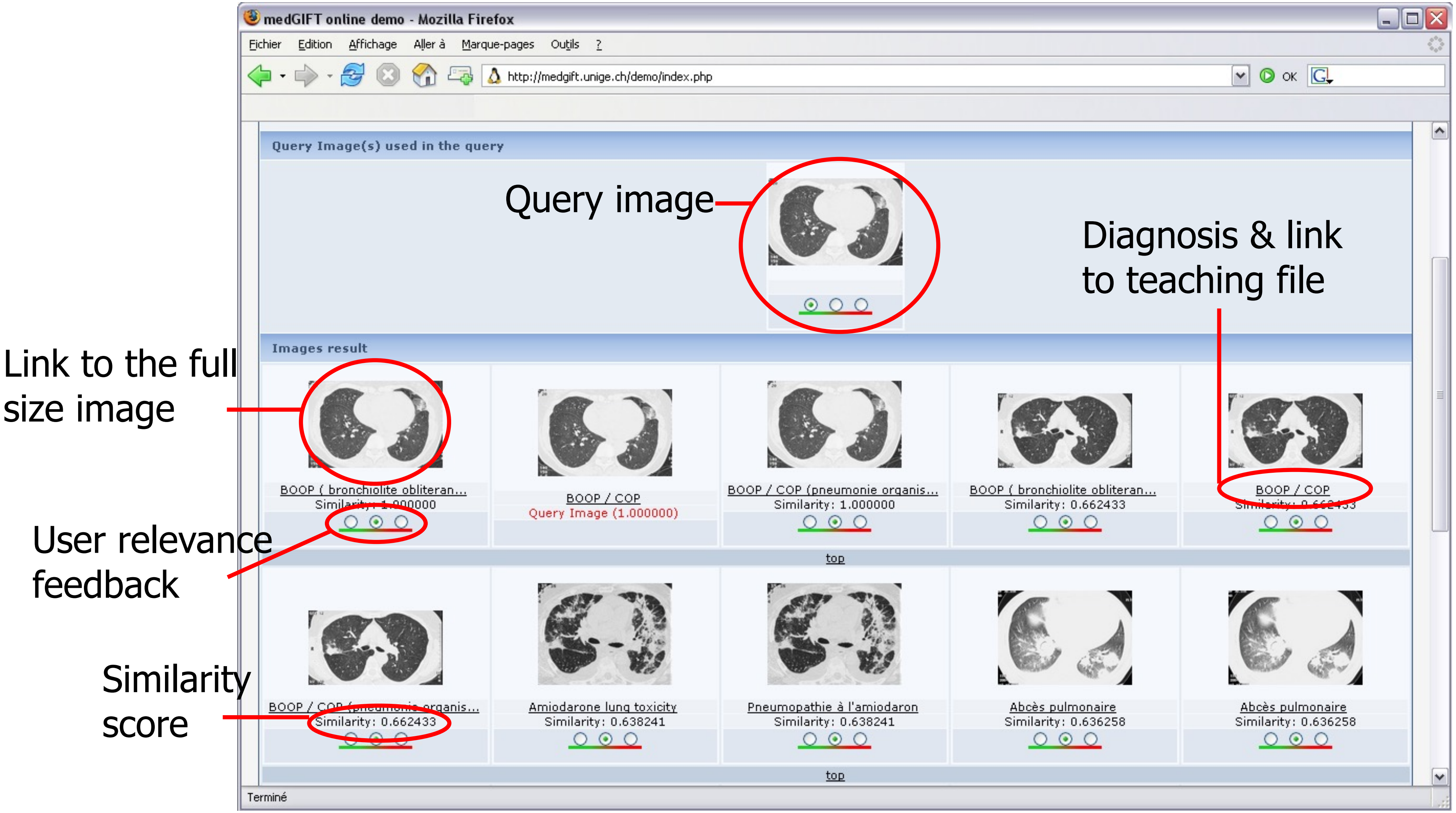
Browsing large collections

- Video Search and Retrieval Interface
  - Collection guide



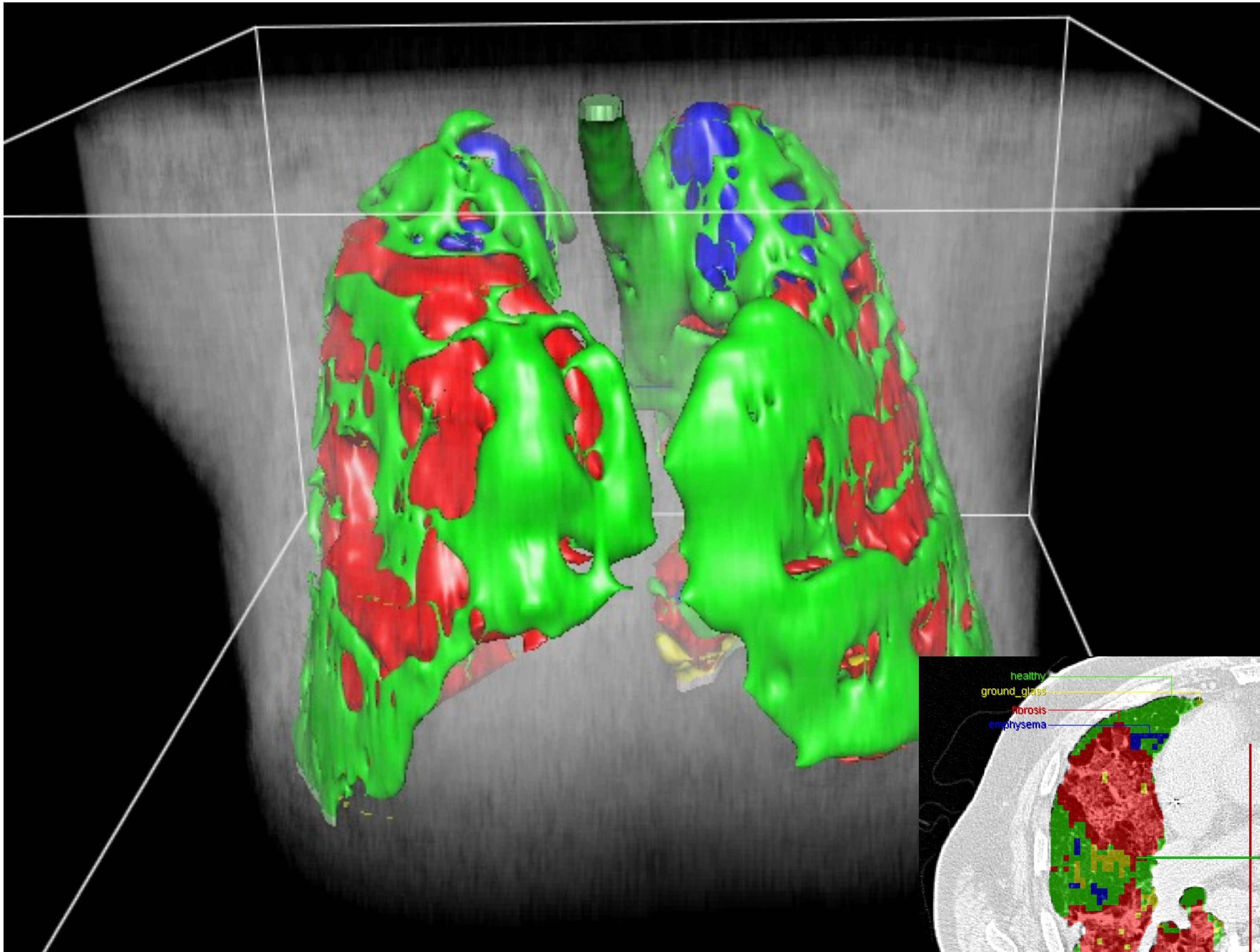
Nguyen et al. ACM Trans. MM 2008

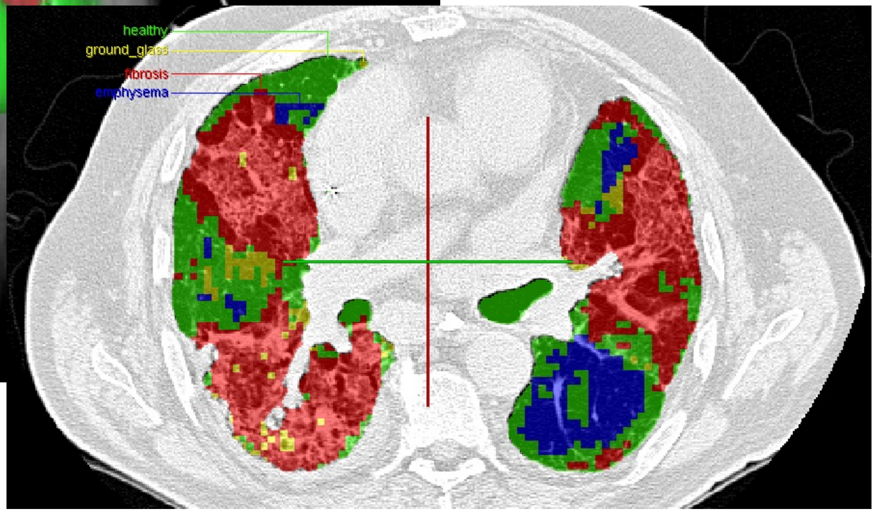
An early medical interface



Müller, H., Rosset, A., Vallée, J. P., & Geissbühler, A. (2003). Integrating content-based visual access methods into a medical case database. *Studies in health technology and informatics*, 480-485.

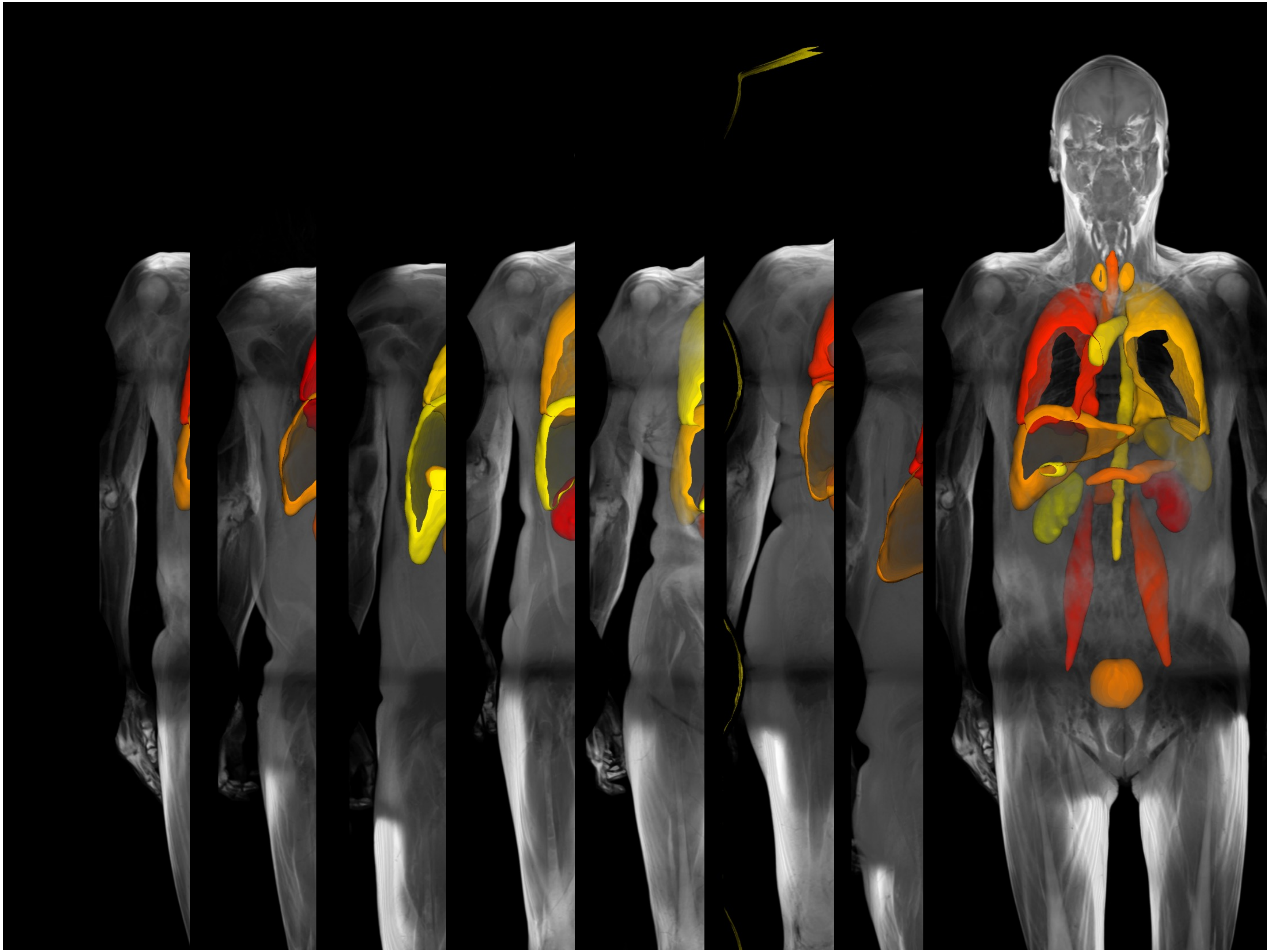
3D ROIs in lung CT



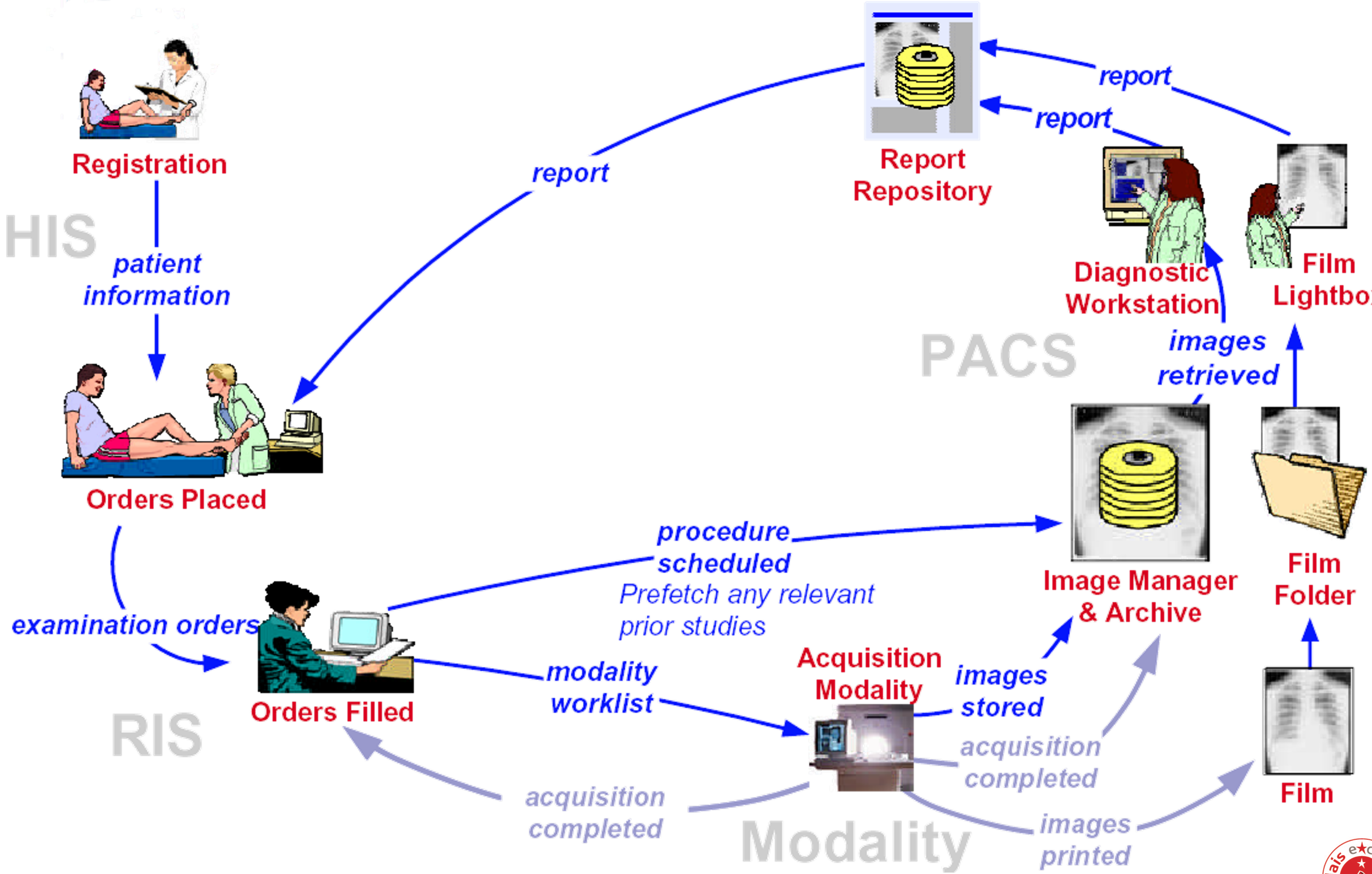


Specific for medical image retrieval

- Images are part of a medical case, so always have a clear/documented context and objective
  - Age or medication use influence the image content
  - Temporal component (trajectory of a patient)
  - Confidential data (so there are restrictions)
- Almost all images have a text attached to it
  - Usually a report detailing the findings, or a scientific text or teaching content (secondary use)
- Regions of interest (ROI) are often extremely small
  - Particularly in tomographic (3D) images
- Very standardized way of taking the image
  - But large anatomical variety, changing devices



Medical image workflow



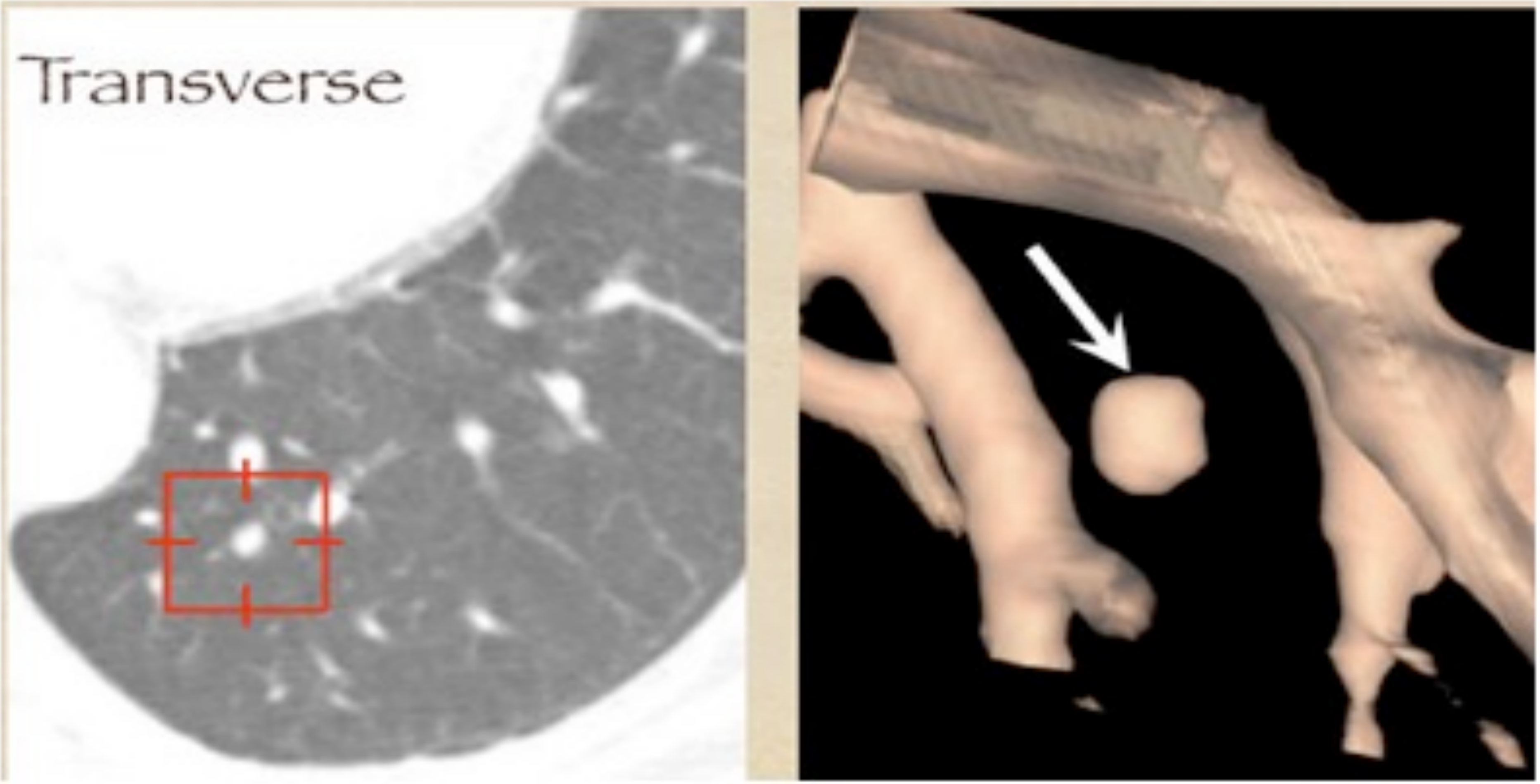
Reference: DICOM standard

Content vs. context: healthy patients

25 year old man:
- homogeneous tissue

88 year old man:
- lower mean density
- pre-fibrotic lesions

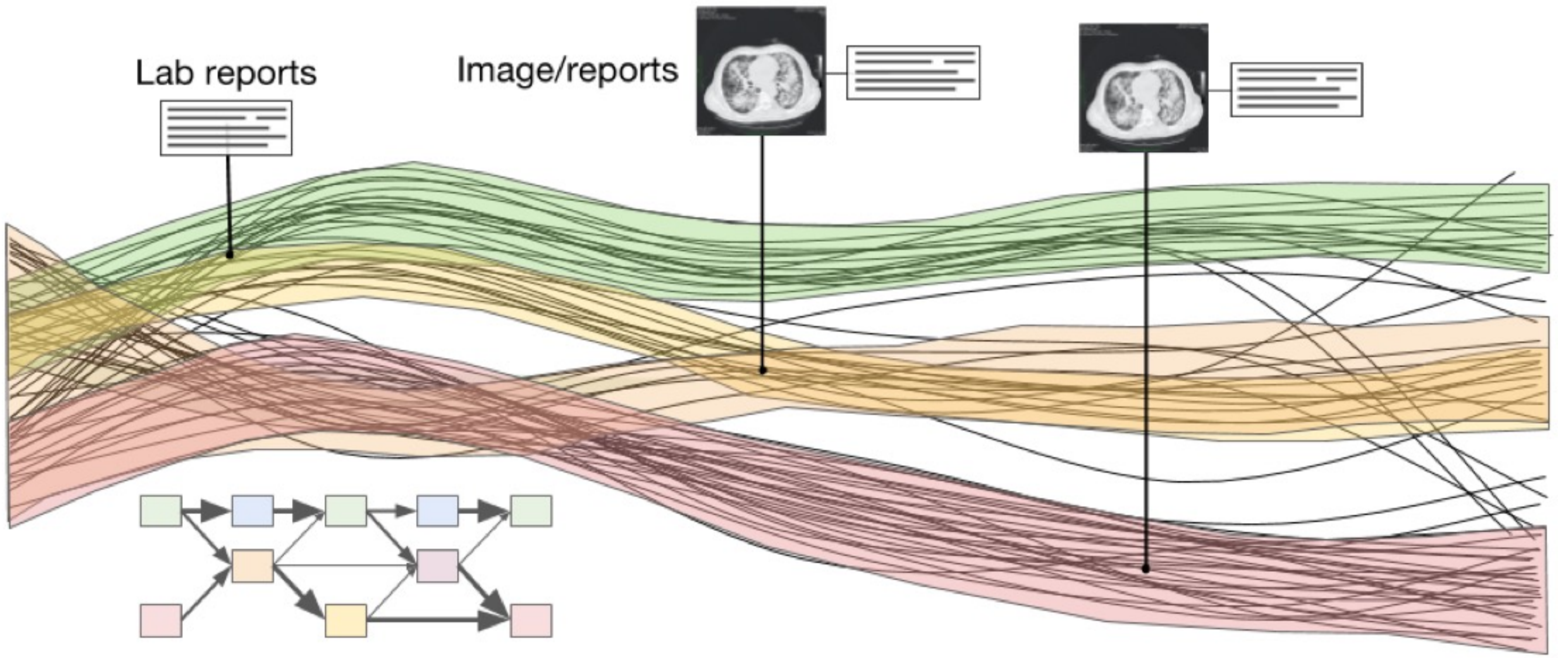
3D can be really important



CT finding (left) has the appearance of an adjacent vessel in transverse-section reconstruction and was not called by any of the four LIDC readers. After viewing transverse, coronal, sagittal, and volume-rendered reconstructions (right), all four university readers called the finding a lung nodule.

Trajectories of patients

- Changes often matter more than absolute values
  - Which direction is a patient coming from
  - Longitudinal data are needed but hard to obtain
- Makes computation even more complex …
  - Much possible noise in the data

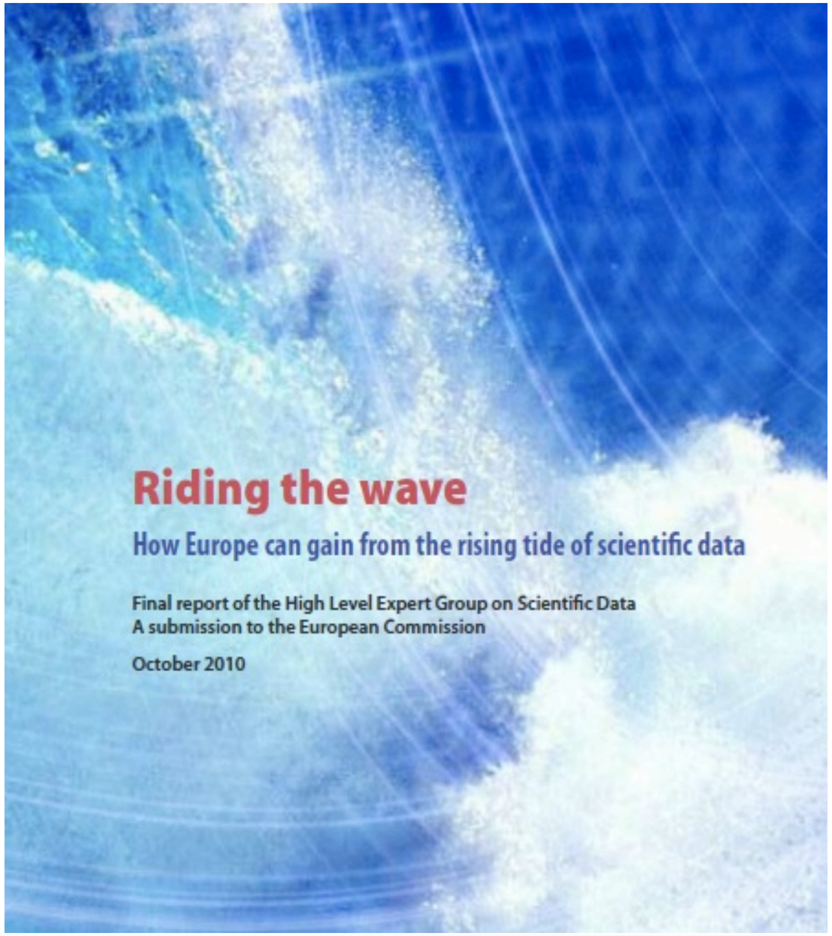


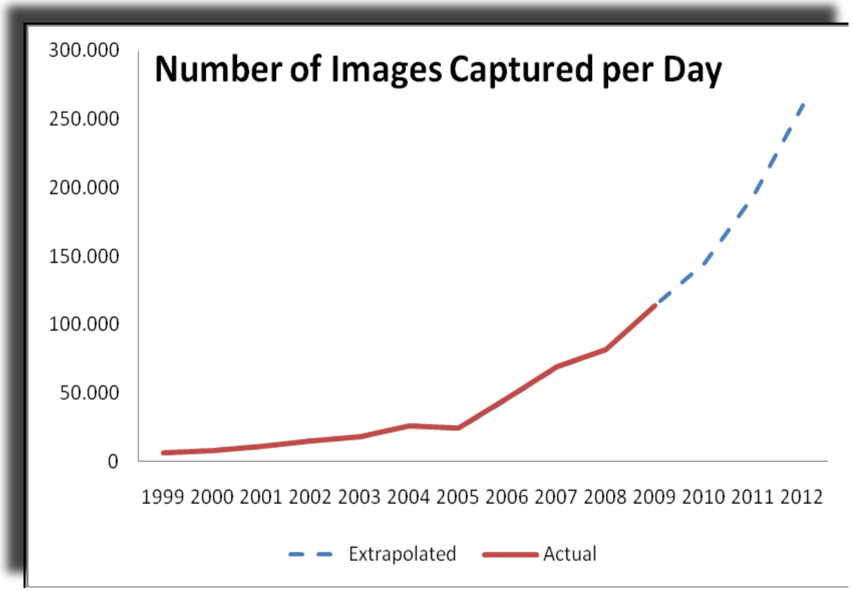
Fracture retrieval

- Database with >20’000 images
  – Before and after interventions, sometimes long term
- Assistance for treatment planning
  – Goal is to find similar cases
    - Based on several images (frontal, lateral), place of fracture, complexity of fracture but also patient data: age, weight, …
  – What is the best method?
    - Screw, plate, …
  – Local features required
    - Salient points, …

Motivations for medical CBIR

- Many images are produced
- Imaging data are very complex
  - And getting more complex
- Imaging is essential for diagnosis & treatment planning
- Images out of their context loose most of their sense
  - Clinical data are necessary
- Towards precision medicine
  - Using all data sources



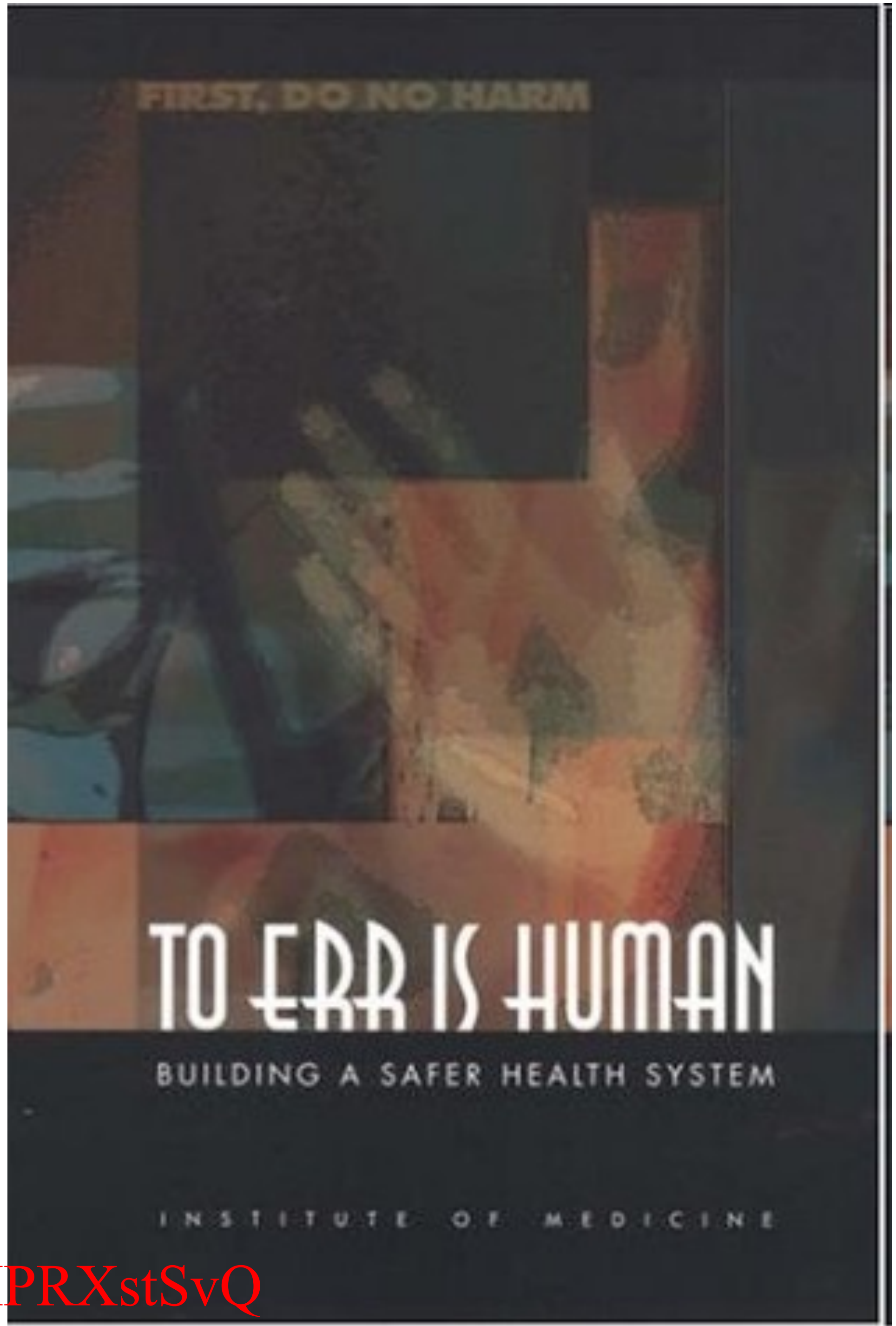


Applications of retrieval

- Clinical work
  - Find similar cases or get explications on a lesion (ROI) where little is known
    - Help less experienced persons, or as second opinion
    - Good integration, as much time pressure exists
- Research
  - Find cases for a study/clinical trial that correspond also to certain visual criteria (multimodal)
- Teaching
  - Usage of teaching files that exist in many places
  - Find images "visually similar to a pathology but with a different disease"

Why decision support?

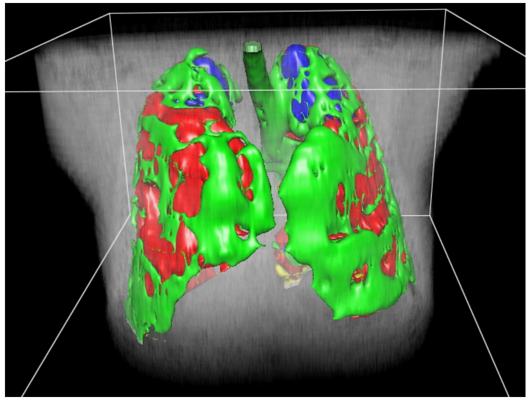
- Humans make mistakes
  - Overloaded, tired, …
- Machines do not generalize well to unknown domains
  - Rare situations
- Humans with decision support usually have best results



Geoff Hinton on radiology (2016)
https://www.youtube.com/watch?v=2HMPRXstSvQ

Imaging applications: CADx, CADe

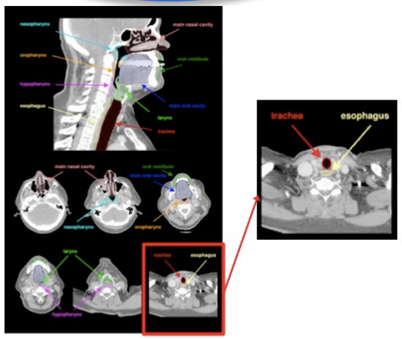
- Computer-Aided Diagnosis (CADx)
- Computer-Aided Detection (CADe)
  - Finding locations of lesions
- Computer-Aided Decision Support
- Many tools are in this area
  - Finding similar patients (retrieval)
  - Finding criteria for or against specific diseases (rules)
  - Prediction of findings such as tissue types
  - Predicting a diagnosis using machine learning



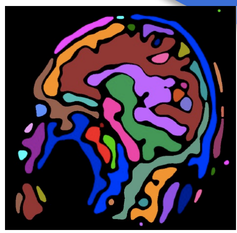
Steps in visual decision support

Resource creation

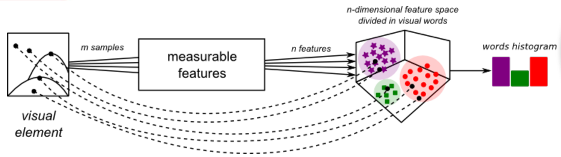
Image pre-treatment



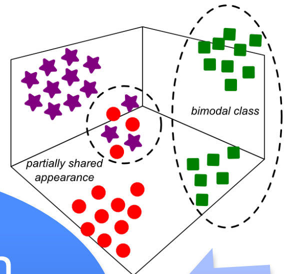
Visual feature extraction



Feature modeling

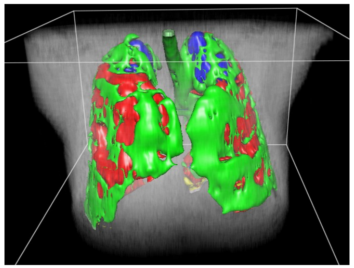


Multimodal fusion

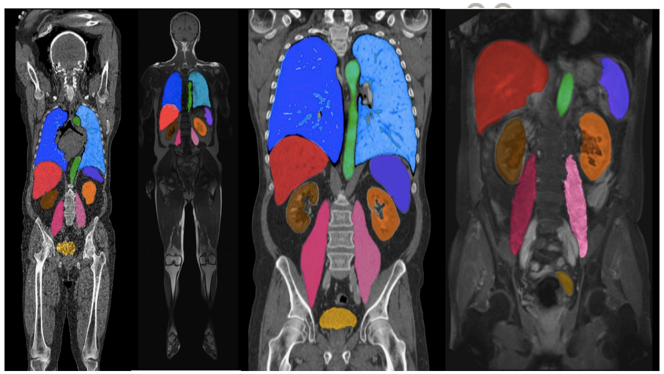


Classification, detection, retrieval

Visualization, results presentation



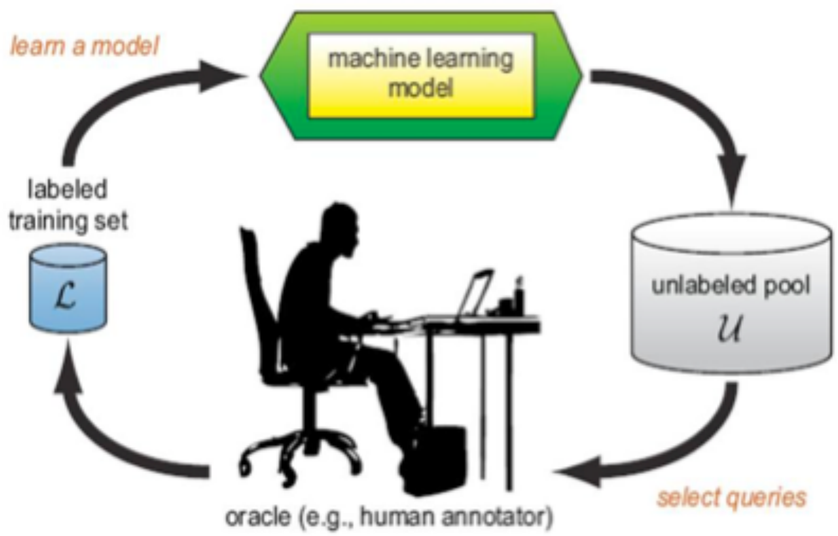
Annotating medical image regions



- Even “clear” annotations like entire organs are subjective
  – Even the same person at different moments annotates differently
  – Inter-annotator disagreement measures subjectivity
- Clear guidelines can create better annotations
  – Semi-automatic tools can harmonize but create a bias
- Automatic segmentations often try to model a human annotator closely
  – Data-driven vs. model-driven
  – Organs, lesions, landmarks, other structures

Active learning

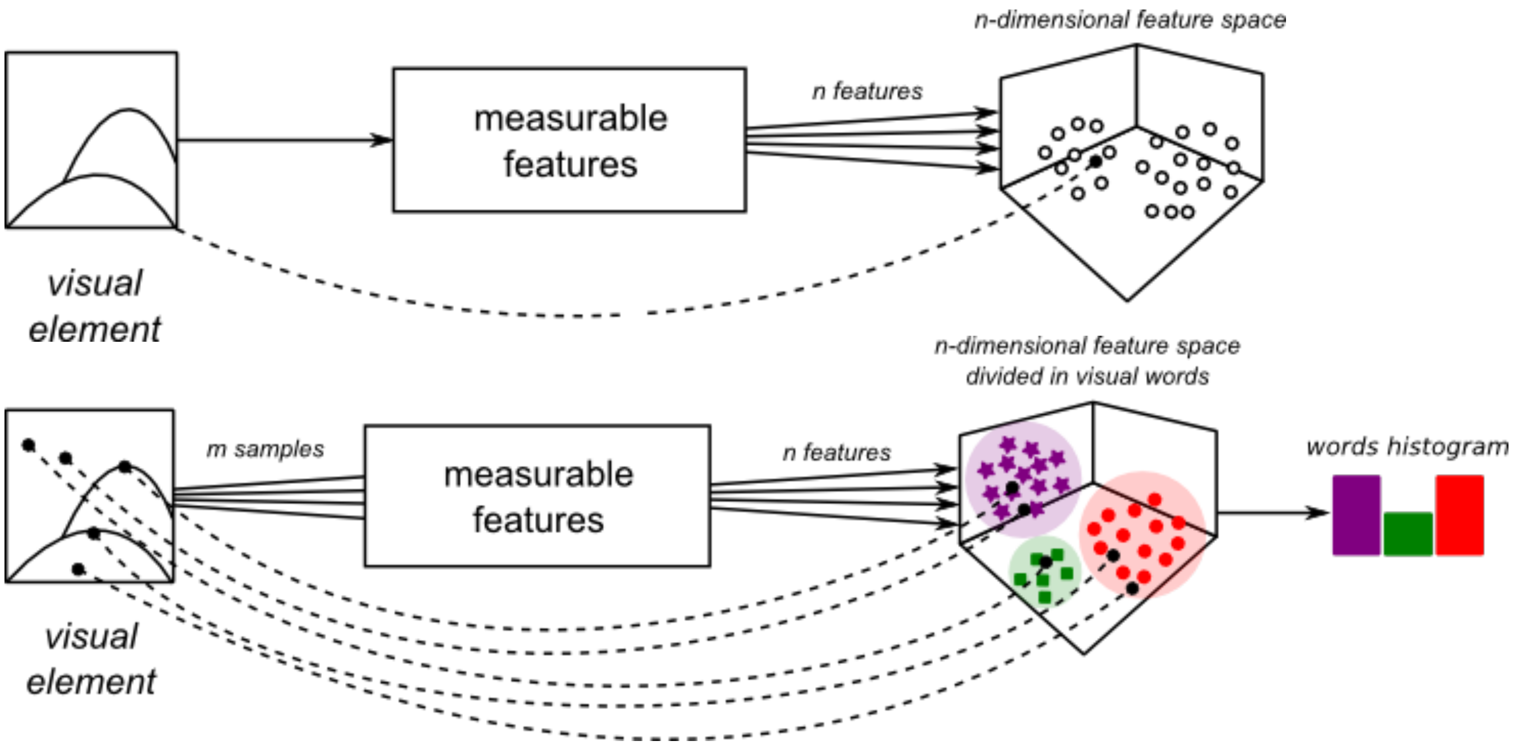
- Let the algorithm decide which non-labelled instances are best to be annotated
  - Usually iteratively to maximize information gain
- Interactive way to limit the amount of annotated data needed
- Visualization can make things easier



Pressforward.org

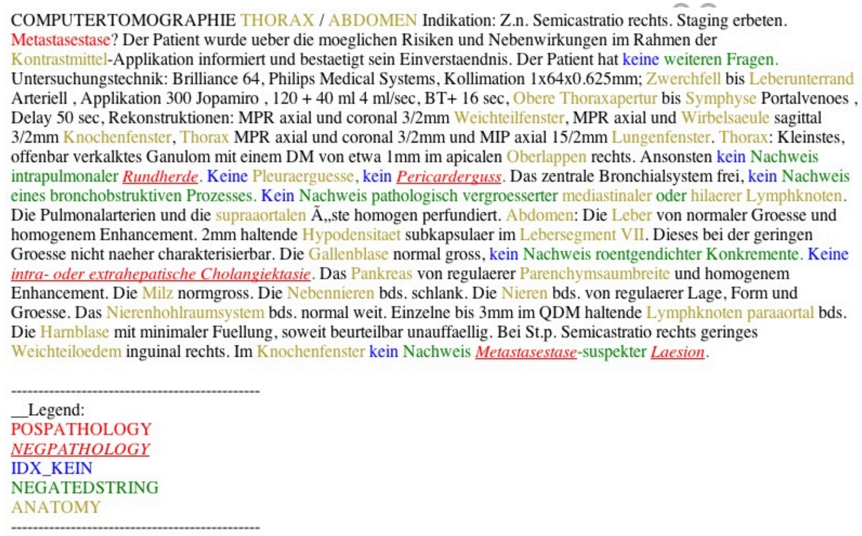
Visual feature modeling

- Visual words instead of raw visual features
  - Reducing the “curse of dimensionality”
- Use features based on the data actually present in a database



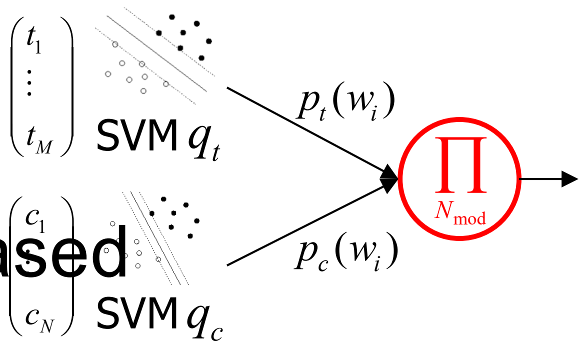
A Foncubierta, AG Seco de Herrera, H Müller, Medical Image Retrieval using a Bag of Meaningful Visual Words, ACM MM workshop on medical multimedia retrieval, Barcelona, Spain, 2013.

Information fusion


COMPUTERTOMOGRAPHIE THORAX / ABDOMEN Indikation: Z.n. Semicastratio rechts. Staging erbeten. Metastasestase? Der Patient wurde ueber die moeglichen Risiken und Nebenwirkungen im Rahmen der Kontrastmittel-Applikation informiert und bestaetigt sein Einverstaendnis. Der Patient hat keine weiteren Fragen. Untersuchungstechnik: Brilliance 64, Philips Medical Systems, Kollimation 1x64x0.625mm; Zwerchfell bis Leberunterrand Arteriell , Applikation 300 Jopamiro , 120 + 40 ml 4 ml/sec, BT+ 16 sec, Obere Thoraxapertur bis Symphyse Portalvenoes , Delay 50 sec, Rekonstruktionen: MPR axial und coronal 3/2mm Weichteilfenster, MPR axial und Wirbelsaeule sagittal 3/2mm Knochenfenster, Thorax MPR axial und coronal 3/2mm und MIP axial 15/2mm Lungenfenster. Thorax: Kleinstes, offenbar verkalktes Ganulom mit einem DM von etwa 1mm im apicalen Oberlappen rechts. Ansonsten kein Nachweis intrapulmonaler *Rundherde*. Keine Pleuraerguesse, kein *Pericarderguss*. Das zentrale Bronchialsystem frei, kein Nachweis eines bronchobstruktiven Prozesses. Kein Nachweis pathologisch vergroesserter mediastinaler oder hilaerer Lymphknoten. Die Pulmonalarterien und die supraaortalen Ã„ste homogen perfundiert. Abdomen: Die Leber von normaler Groesse und homogenem Enhancement. 2mm haltende Hypodensitaet subkapsulaer im Lebersegment VII. Dieses bei der geringen Groesse nicht naeher charakterisierbar. Die Gallenblase normal gross, kein Nachweis roentgendichter Konkremente. Keine *intra- oder extrahepatische Cholangiektasie*. Das Pankreas von regulaerer Parenchymsaumbreite und homogenem Enhancement. Die Milz normgross. Die Nebennieren bds. schlank. Die Nieren bds. von regulaerer Lage, Form und Groesse. Das Nierenhohlraumsystem bds. normal weit. Einzelne bis 3mm im QDM haltende Lymphknoten paraaortal bds. Die Harnblase mit minimaler Fuellung, soweit beurteilbar unauffaellig. Bei St.p. Semicastratio rechts geringes Weichteiloedem inguinal rechts. Im Knochenfenster kein Nachweis *Metastasestase*-suspekter *Laesion*.

__Legend:
POSPATHOLOGY
NEGPATHOLOGY
IDX_KEIN
NEGATEDSTRING
ANATOMY

- Combine information from several sources (i. e. text or structured data with visual)
- Text data can be mapped to semantics to understand links
  - Also language-independent
- Early fusion
  - Combing all features using a single classifier in the end
- Late fusion
  - Using classifier output
  - Rank-based vs. score-based



Many fusion techniques

- Early fusion can work better but often requires dimensionality reduction
- Rank-based fusion (late)
  - Ranked list of items or classes are combined
  - If score distributions are not the same this can be better
- Score-based fusion (late)
  - Using scores of items for combined decision
    - Weighted linear combination
    - Sum of scores or simply minimum/maximum value
- Borda count, …


DOES IT HURT
WHEN I PRESS HERE?
PESSIN

Many different types of users

- Clinicians looking for the latest research
  - Half-life of medical information is estimated to be between 5 and 10 years Emanuel, E. (1975). A half-life of 5 years. *Canadian Medical Association Journal*, *112*(5), 572.
  - Impossibility to be up to date even in an extremely narrow domain Fraser, A. G., & Dunstan, F. D. (2010). On the impossibility of being expert. *BMJ*, *341*, c6815.
- Patients search for information on their disease or for friends/family
  - Hard to get information on rare diseases
- Pharma industry may search for clinical trials
  - And IPR questions
- Radiologists search in the image content (2D, 3D, 4D)
  - But the equipment is constantly changing

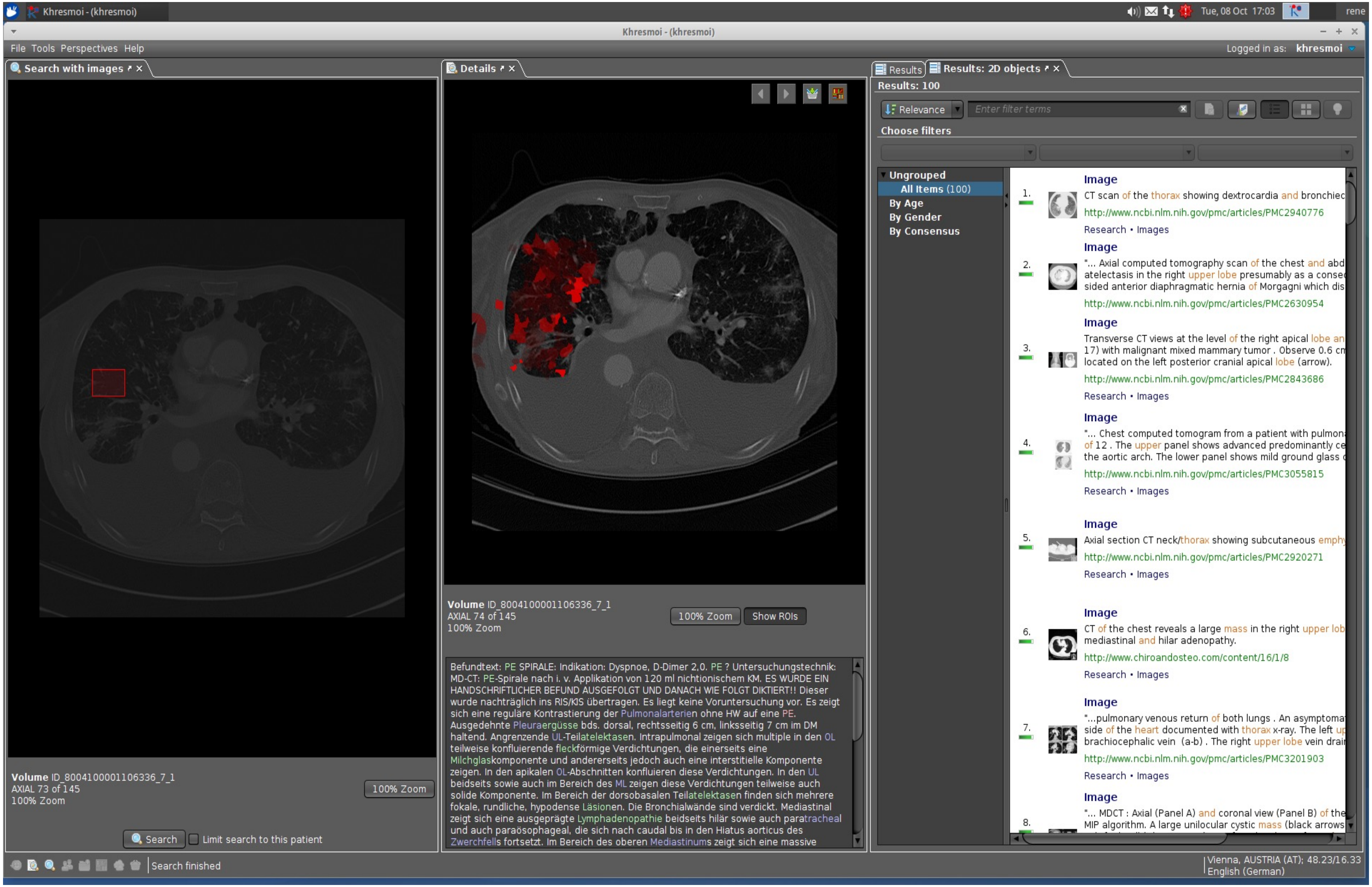
The informed patient

"I'M SORRY DOCTOR, BUT AGAIN I HAVE TO DISAGREE."

- Mixing multilingual data from many resources and semantic information for medical retrieval
  - LinkedLifeData.com

Allan Hanbury, Célia Boyer, Manfred Gschwandtner, Henning Müller, KHRESMOI: Towards a Multi-Lingual Search and Access System for Biomedical Information, Med-e-Tel, pages 412-416, Luxembourg, 2011.

Khresmoi4radiology interface

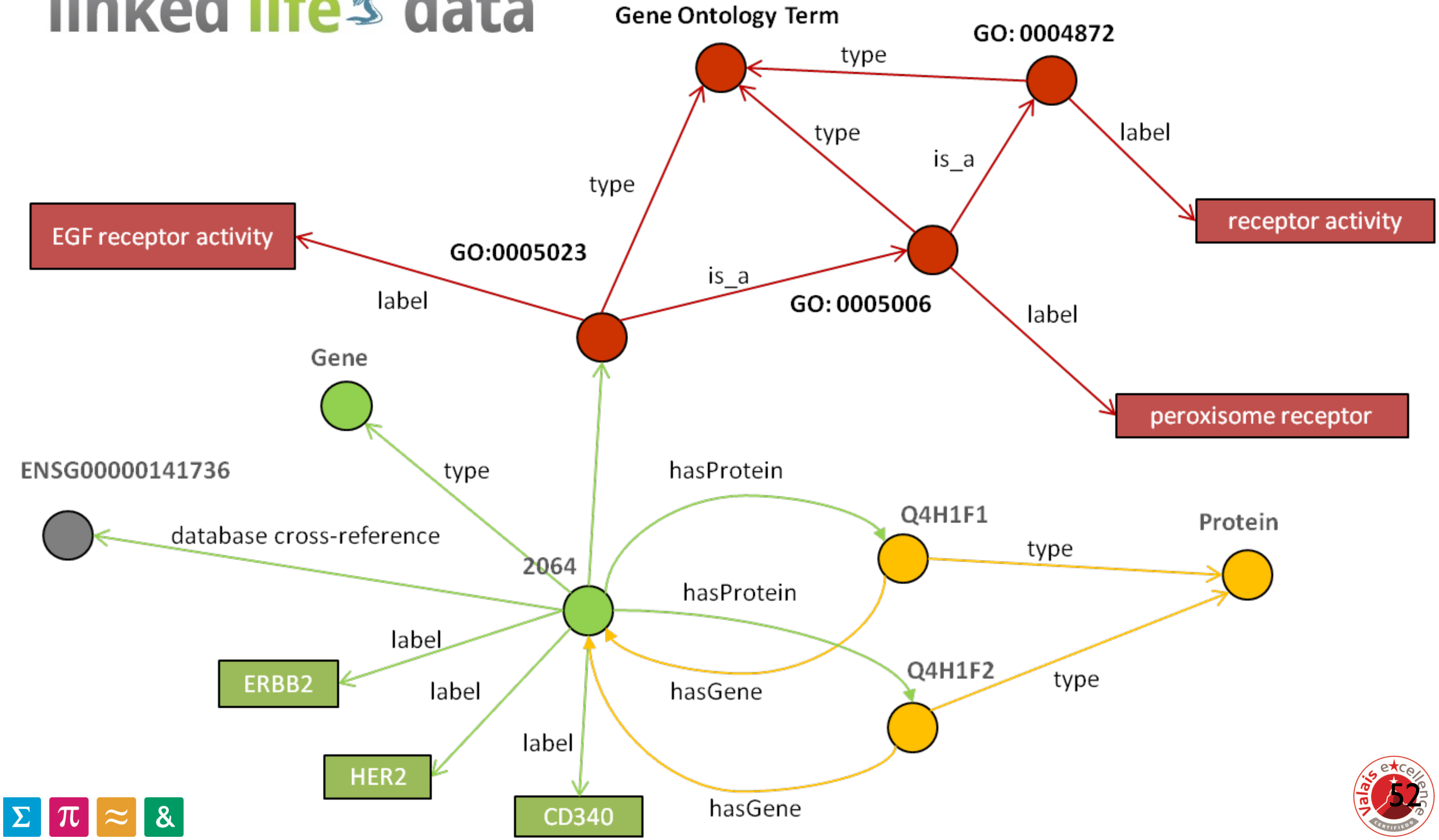


Example for semantic text analysis

COMPUTERTOMOGRAPHIE THORAX / ABDOMEN Indikation: Z.n. Semicastratio rechts. Staging erbeten. Metastasestase? Der Patient wurde ueber die moeglichen Risiken und Nebenwirkungen im Rahmen der Kontrastmittel-Applikation informiert und bestaetigt sein Einverstaendnis. Der Patient hat keine weiteren Fragen. Untersuchungstechnik: Brilliance 64, Philips Medical Systems, Kollimation 1x64x0.625mm; Zwerchfell bis Leberunterrand Arteriell , Applikation 300 Jopamiro , 120 + 40 ml 4 ml/sec, BT+ 16 sec, Obere Thoraxapertur bis Symphyse Portalvenoes , Delay 50 sec, Rekonstruktionen: MPR axial und coronal 3/2mm Weichteilfenster, MPR axial und Wirbelsaeule sagittal 3/2mm Knochenfenster, Thorax MPR axial und coronal 3/2mm und MIP axial 15/2mm Lungenfenster. Thorax: Kleinstes, offenbar verkalktes Ganulom mit einem DM von etwa 1mm im apicalen Oberlappen rechts. Ansonsten kein Nachweis intrapulmonaler *Rundherde*. Keine Pleuraerguesse, kein *Pericarderguss*. Das zentrale Bronchialsystem frei, kein Nachweis eines bronchobstruktiven Prozesses. Kein Nachweis pathologisch vergroesserter mediastinaler oder hilaerer Lymphknoten. Die Pulmonalarterien und die supraaortalen Ã„ste homogen perfundiert. Abdomen: Die Leber von normaler Groesse und homogenem Enhancement. 2mm haltende Hypodensitaet subkapsulaer im Lebersegment VII. Dieses bei der geringen Groesse nicht naeher charakterisierbar. Die Gallenblase normal gross, kein Nachweis roentgendichter Konkremente. Keine *intra- oder extrahepatische Cholangiektasie*. Das Pankreas von regulaerer Parenchymsaumbreite und homogenem Enhancement. Die Milz normgross. Die Nebennieren bds. schlank. Die Nieren bds. von regulaerer Lage, Form und Groesse. Das Nierenhohlraumsystem bds. normal weit. Einzelne bis 3mm im QDM haltende Lymphknoten paraaortal bds. Die Harnblase mit minimaler Fuellung, soweit beurteilbar unauffaellig. Bei St.p. Semicastratio rechts geringes Weichteiloedem inguinal rechts. Im Knochenfenster kein Nachweis *Metastasestase*-suspekter *Laesion*.

__Legend:
POSPATHOLOGY
NEGPATHOLOGY
IDX_KEIN
NEGATEDSTRING
ANATOMY

LinkedLifeData



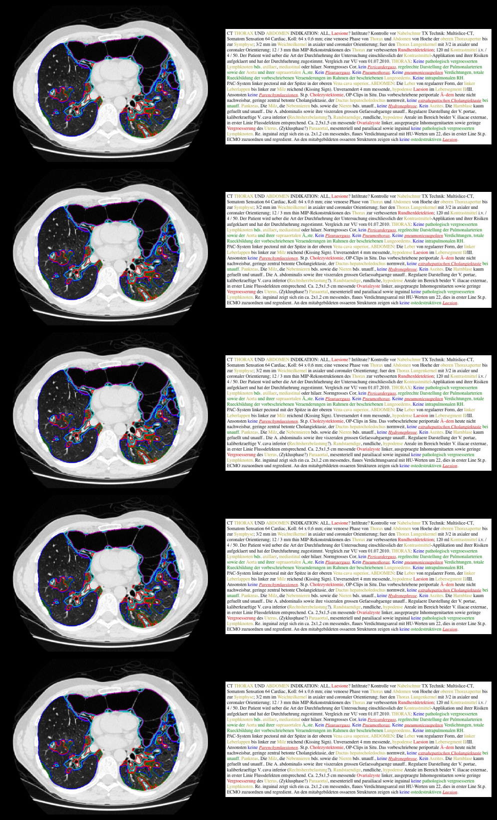
Many resources are available

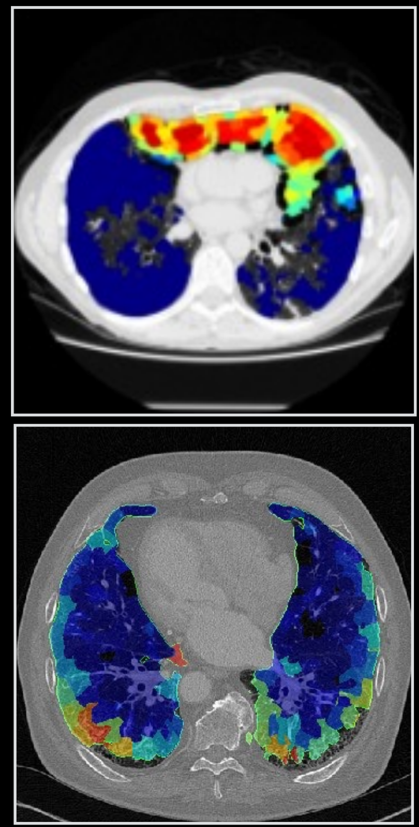
- NLM maintains many language resources
  - MeSH – Medical Subject Headings
  - UMLS – Unified Medical Language Syste
  - PMC and PubMed data sets
  - RadLex is maintained by the RSNA
- NLM maintains software and web interfaces
  - OpenI, PubMed.gov, PubMed Central
  - MedLine Plus – a resource for patients
  - NegBio, MetaMap
  - …

KHRESMOI: towards retrieval in clinical data

Hofmanninger CVPR 2015

- Combine imaging data and radiology report information
- Extract semantic information from radiologe reports
- Identify visual signatures of findings from this combined data





Markers for disease patterns, learned based on images + reports

… but there are challenges

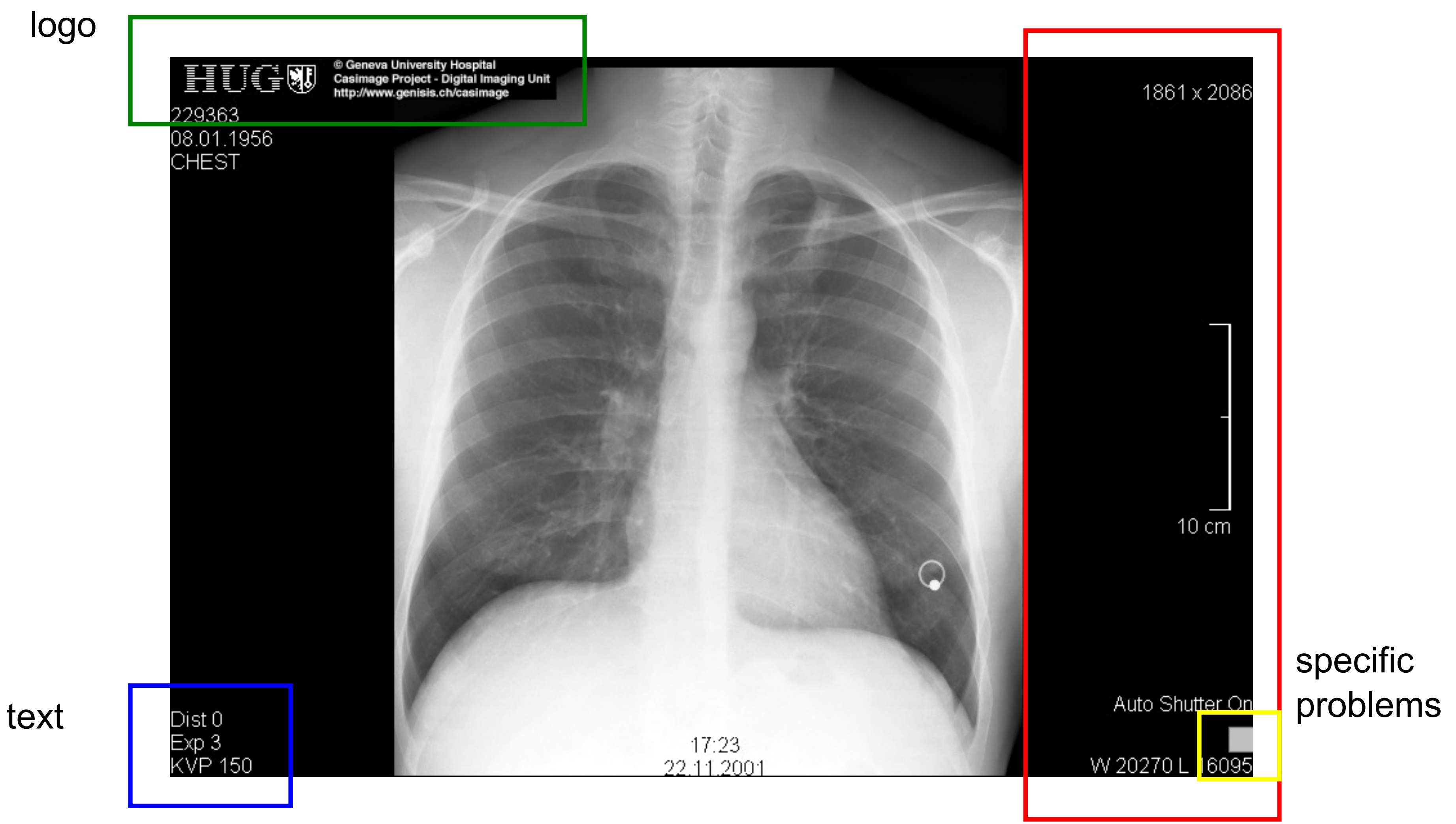
- Non-standard abbreviations
- Spelling mistakes
  - Quickly written
- Technical language
  - Latinized terms, synonyms
- Nested, complex phrases
- Negation …
  - Several levels (“little evidence of”)
  - Not clear what terms they refer to, double negations

A/B	acid-base ratio
ab	abdomen, abdominal abortion
Ab	antibody
AB	abortion, AB Blood Type
ABC	airway, breathing, circulation aspiration biopsy cytology
ABCD	airway, breathing, circulation, disability asymmetry, borders, color, diameter (feature ABCD rating (a staging system for prostate c
ABCs ABCDs ABCDEs	airway, breathing, circulation, etc. Refers to p recurrent.
ACA	acinic cell carcinoma Affordable Care Act
Abd	abdomen abdominal[abduction]
ABD	army battle dressing

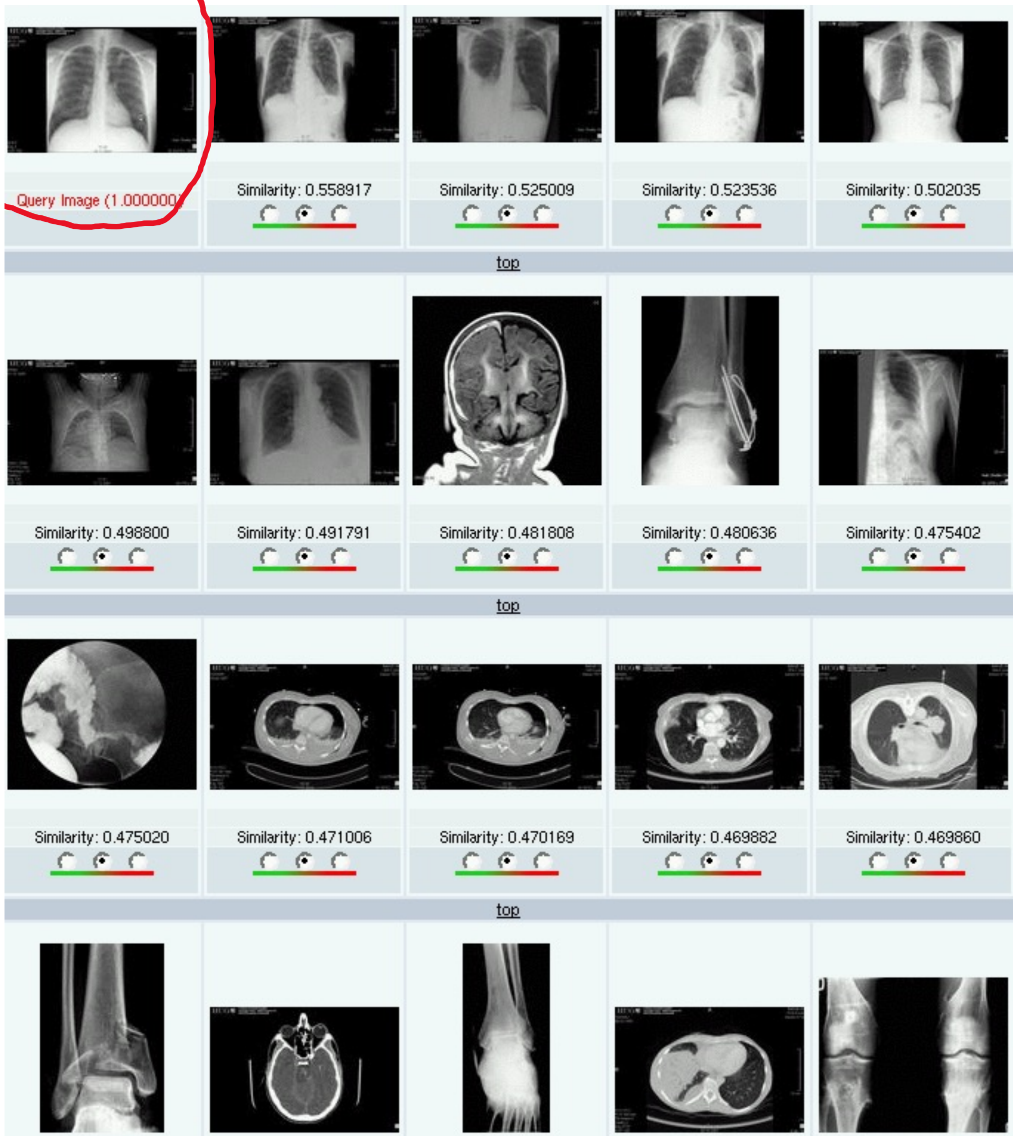
cardiac arrest *noun* heart attack

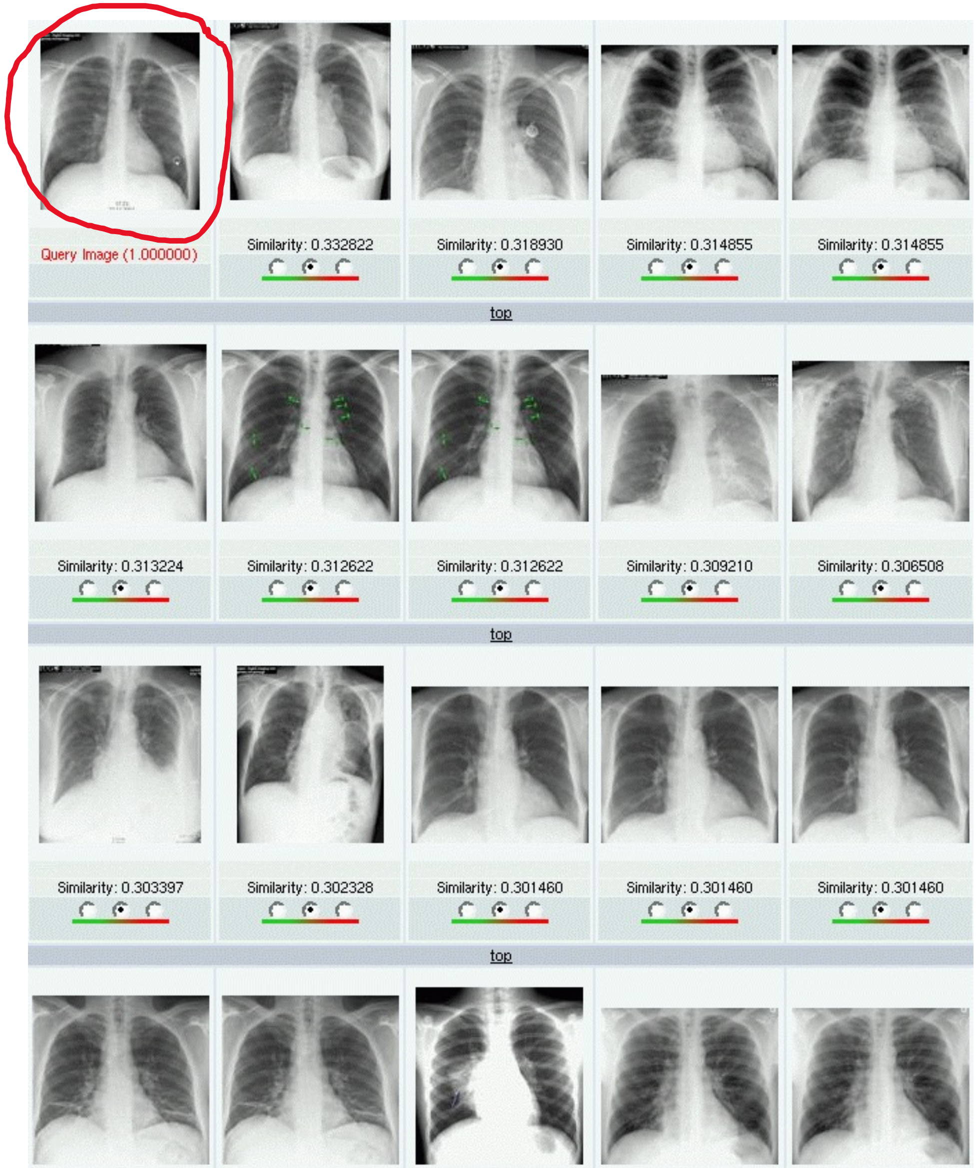
asystole	congestive heart failure	heart arrest	heart stoppage
cardiac infarction	coronary infarction	heart attack	myocardial infarction
cardiopulmonary arrest	coronary thrombosis	heart failure	tachycardia

Visual challenges



Large parts without information





Müller, H., Heuberger, J., & Geissbuhler, A. (2005). Logo and text removal for medical image retrieval. In *Bildverarbeitung für die Medizin 2005* (pp. 35-39). Springer, Berlin, Heidelberg.

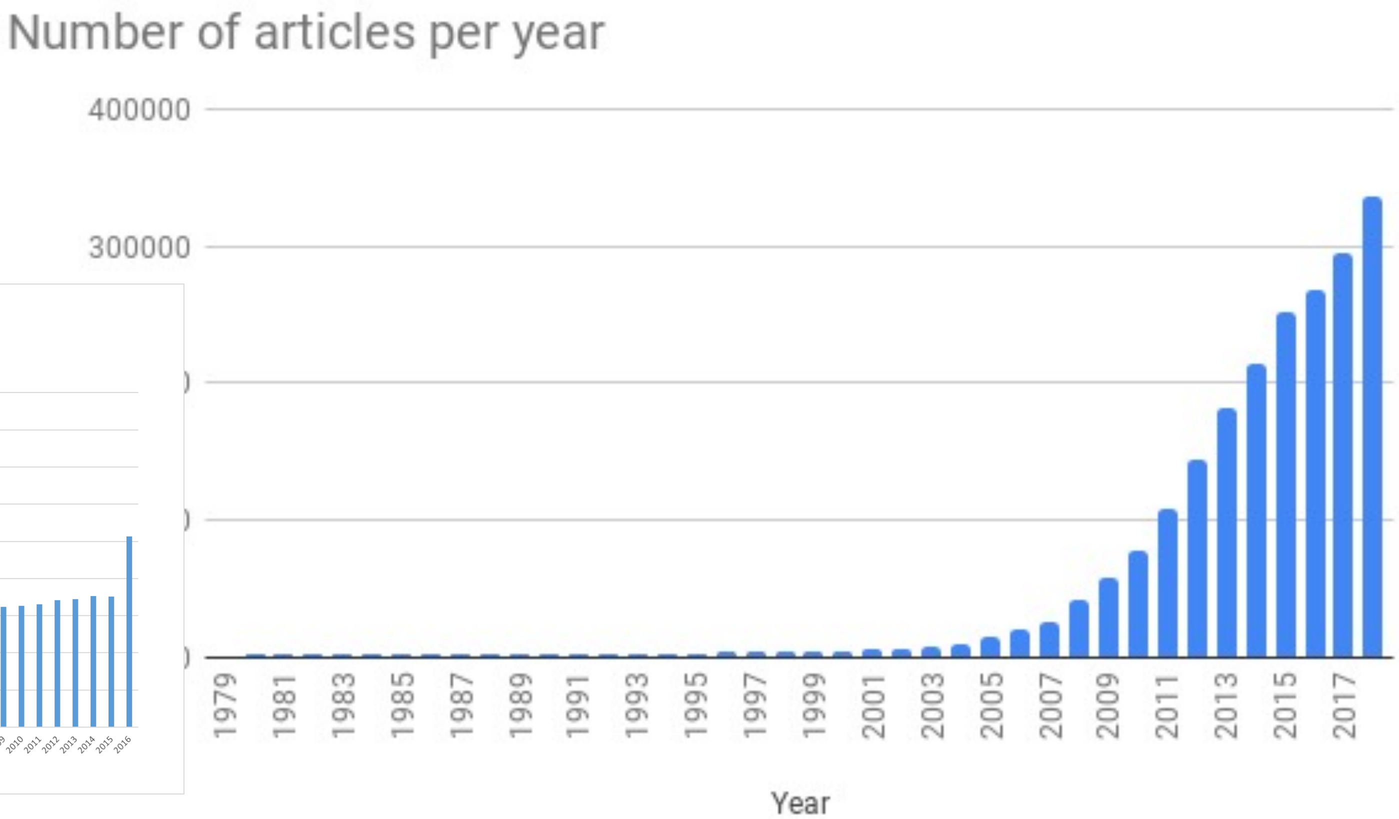
Medical image resources

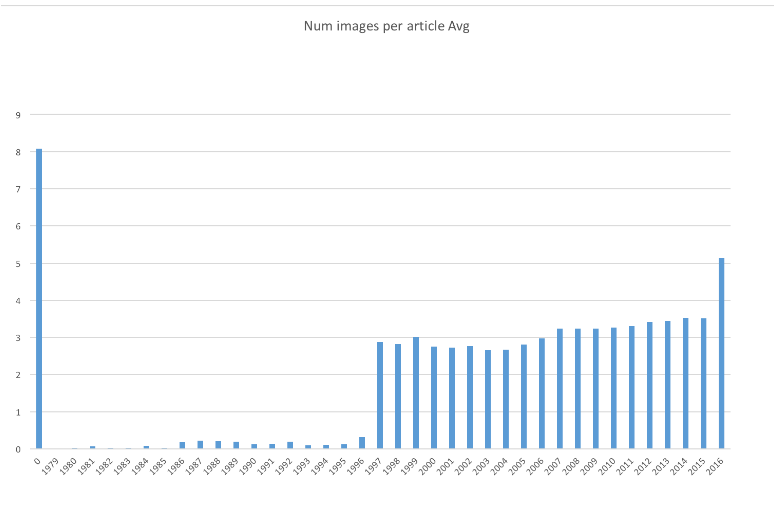
- The Cancer Imaging Archive (TCIA)
- The Cancer Genome Atlas (TCGA)
- Published data sets (Nature Scientific Data)
- Scientific challenges
  - ImageCLEF, Kaggle, Aicrowd, Codalab, Dream challenges, …
- PubMed Central (Medical open access literature)

- What is often missing are annotations …

PubMed Central

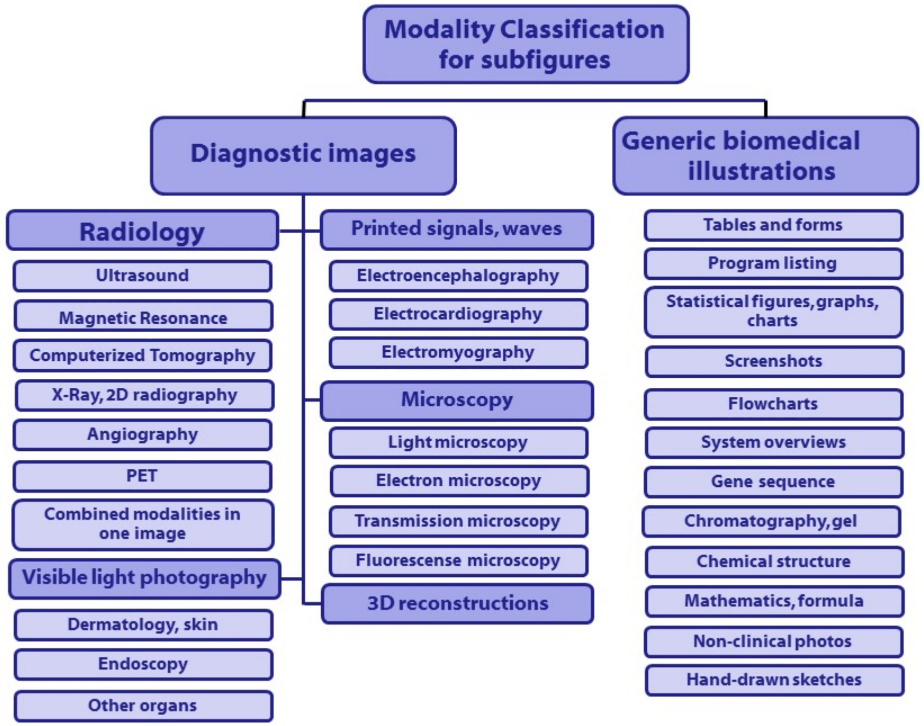
- Repository with the biomedical open access literature, including images as files, etc.
  - 3-4 images per article, increasing, compound figures (10/2020: 6.5 million articles)





Structuring the visual content

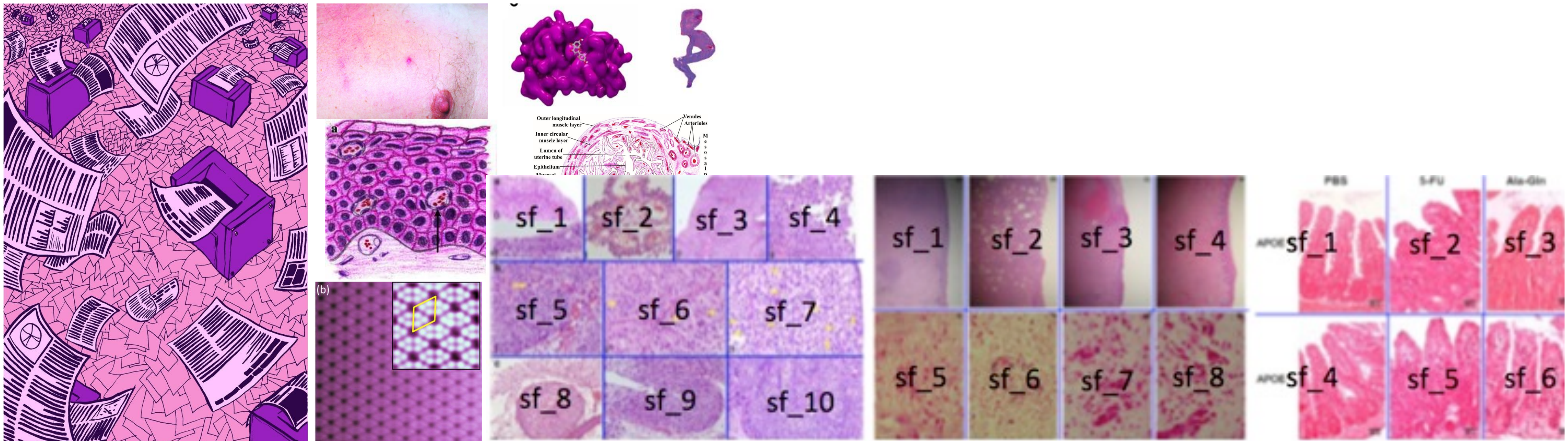
- Types of images make the literature images classifiable
  - Extremely large variety in most categories
    - Many sub-categories are possible
  - Categories with clinical relevance are most important
  - Allows removing noise
  - Compound figures are separately treated



Müller, H., Kalpathy-Cramer, J., Demner-Fushman, D., & Antani, S. (2012, February). Creating a classification of image types in the medical literature for visual categorization. In *Medical Imaging 2012:* (Vol. 8319).

Challenges in the data

  - >14,000,000 images in 1/2020
- Look-alikes and compound figures
  - Much strange content that needs to removed
- Compound figures can not easily be separated, as they contain aspects of several classes
  - Cutting them into subfigures makes content accessible



Meta data available for PMC

- Text of the figure caption
  - Relatively specific but often short
  - Hard for compound figures that contain many parts
- Full text of the article
  - Non-specific for single figures
  - Location of the figure is available
- Article title and author-generated key words
- Global MeSH terms (Manually attached)
  - Cover species and organs, as they are systematically annotated
- Not all is available for all articles (incomplete)

Tasks to make figures usable

- Removing very small images & strange aspect ratios
- Classify figures into figure types
  - Using image data and also text
  - Remove non-relevant images
- Detect and cut compound figures into their parts
  - Classify these into figure types again
- Filter human and animal tissue
- Filter our specific organs
- Check diseases or grading/staging images
  - Classes for machine learning

Müller, H., Andrearczyk, V., del Toro, O. J., Dhrangadhariya, A., Schaer, R., & Atzori, M. (2020, January). Studying Public Medical Images from the Open Access Literature and Social Networks for Model Training and Knowledge Extraction. In *International Conference on Multimedia Modeling*.

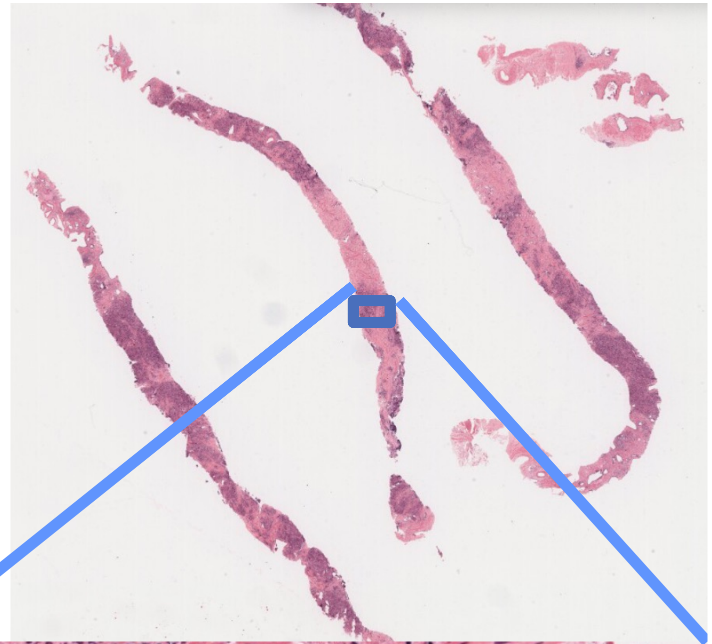
Advantages of literature images

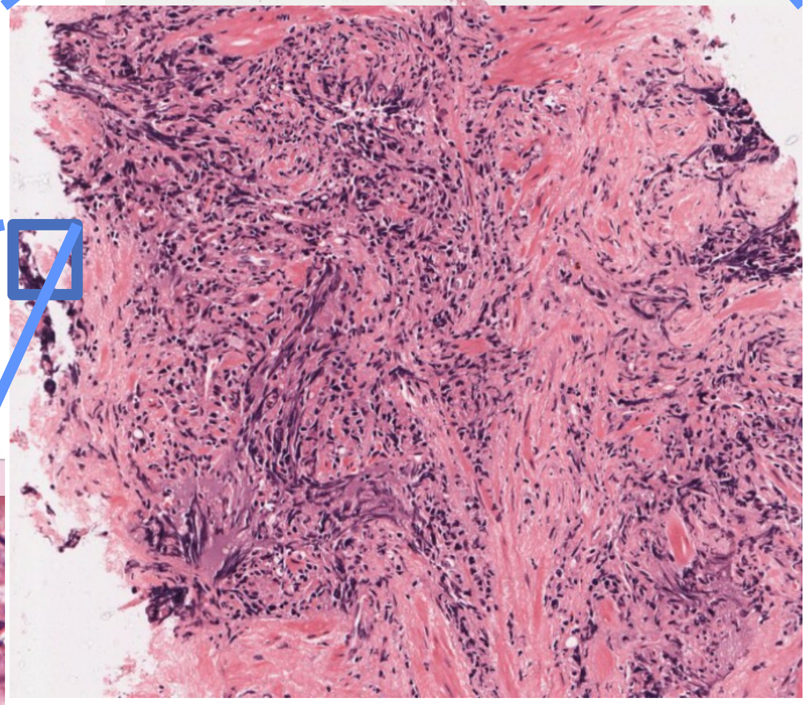
- Rare images (unusual, untypical) are generally used for articles and case descriptions
  - A few typical cases but mainly extreme cases to share the knowledge on them
  - Creates critical mass for rare diseases
- Images are from many laboratories and thus contain many image variations (staining, scanners)
  - Increase generalizability of models thanks to diversity
- Exponentially increasing content
  - Allows to improve models continuously based on latest imaging equipment
- No ethics constraints!

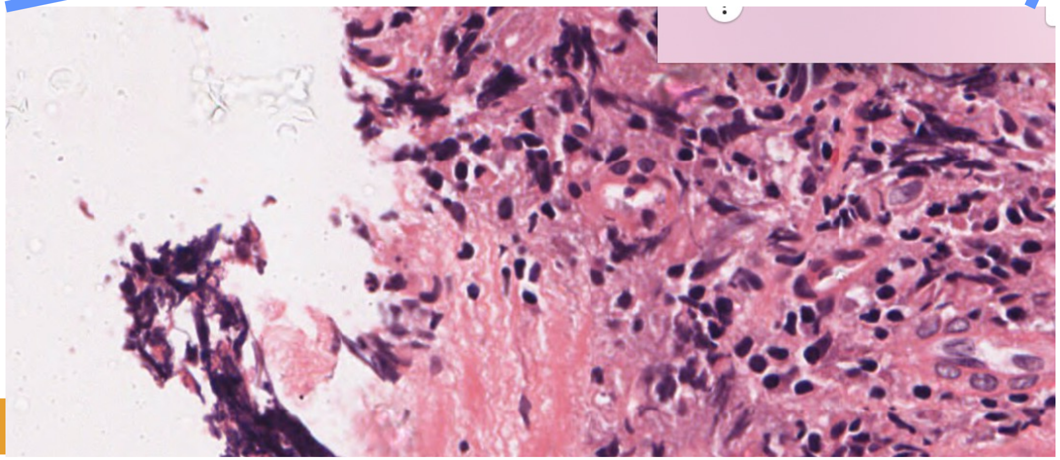
Dhrangadhariya, A., Jimenez-del-Toro, O., Andrearczyk, V., Atzori, M., & Müller, H. (2020, March). Exploiting biomedical literature to mine out a large multimodal dataset of rare cancer studies. In *Medical Imaging 2020,* (Vol. 11318).

Histopathology imaging

- Very large images
  - ~100,000x100,000 pixels
  - Up to 10 GB in size
- Cellular level
  - Pixel size ~250 nm
- Large variety

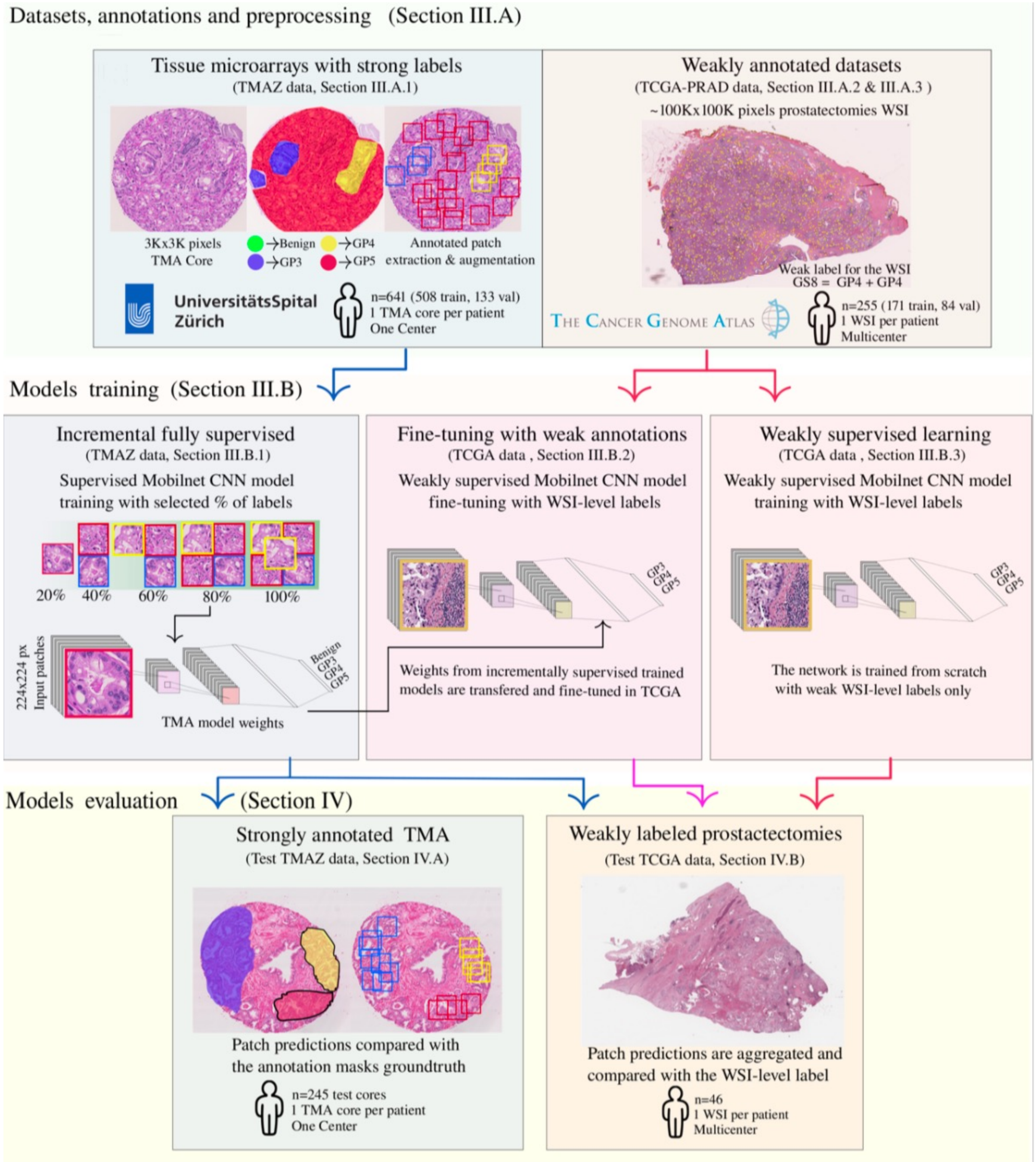






Learning from weak & strong labels

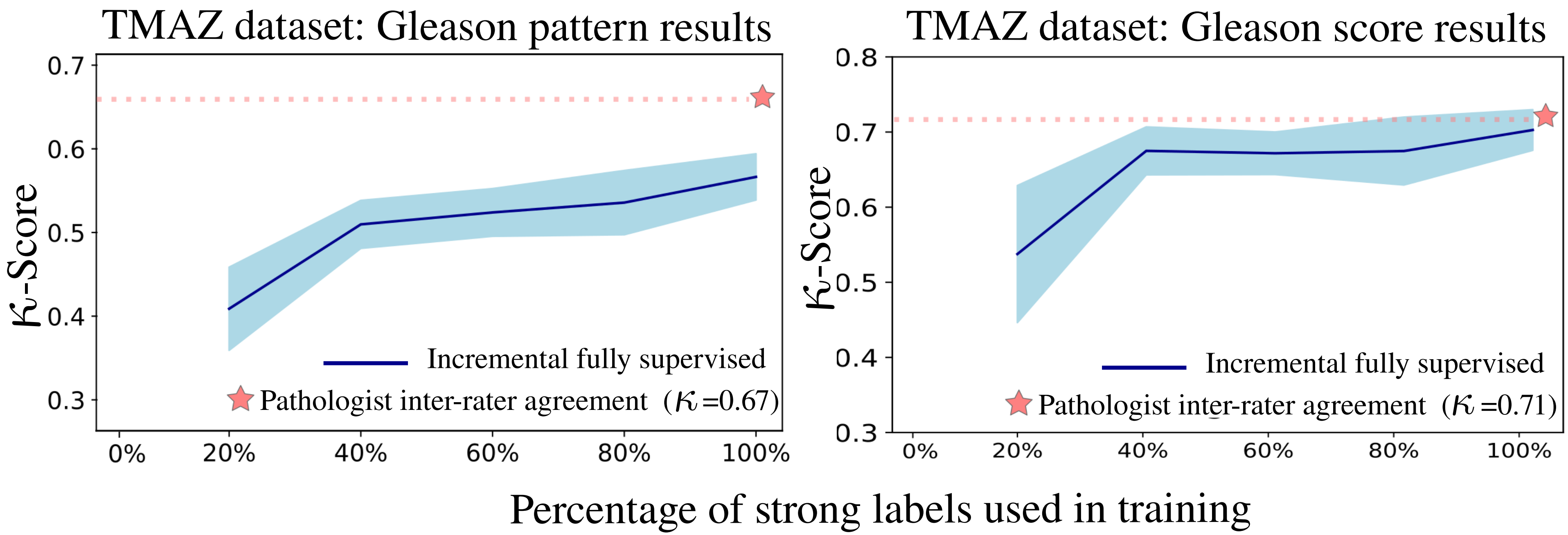
- In this case for Gleason pattern classification and Gleason scoring, prostate cancer
- Amount of data with weak and strong labels needs to be determined experimentally
  - Weak labels can be obtained from the pathology reports or figure captions with little efforts
  - Some data sets are distributed with strong labels
- Performance should be evaluated on unrelated datasets as well
  - Show generalizability of the learned models

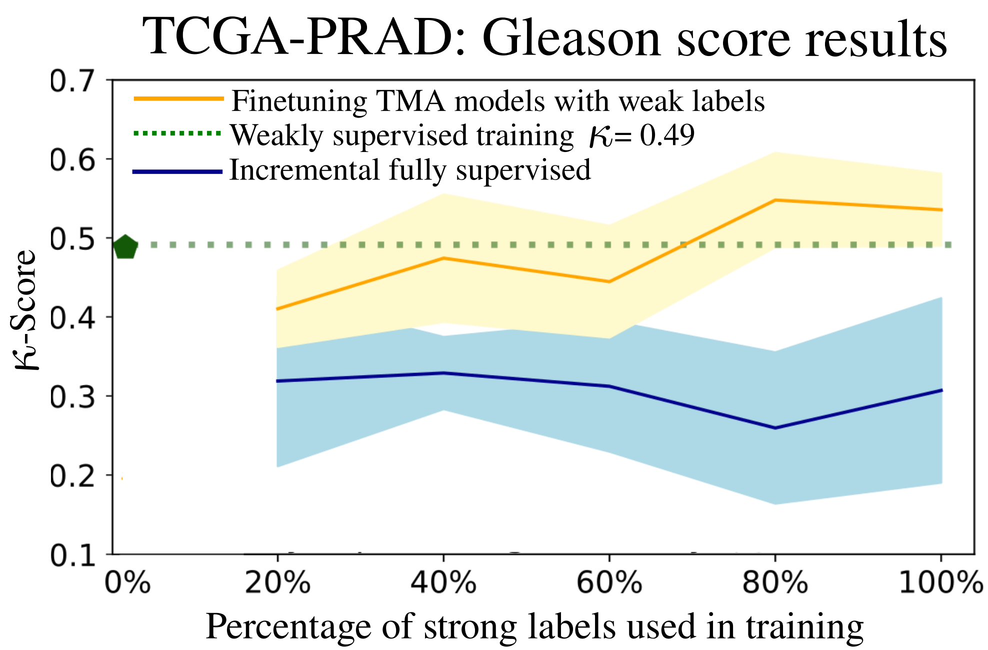


Hes·so VALAIS WALLIS
School of Management & Tourism

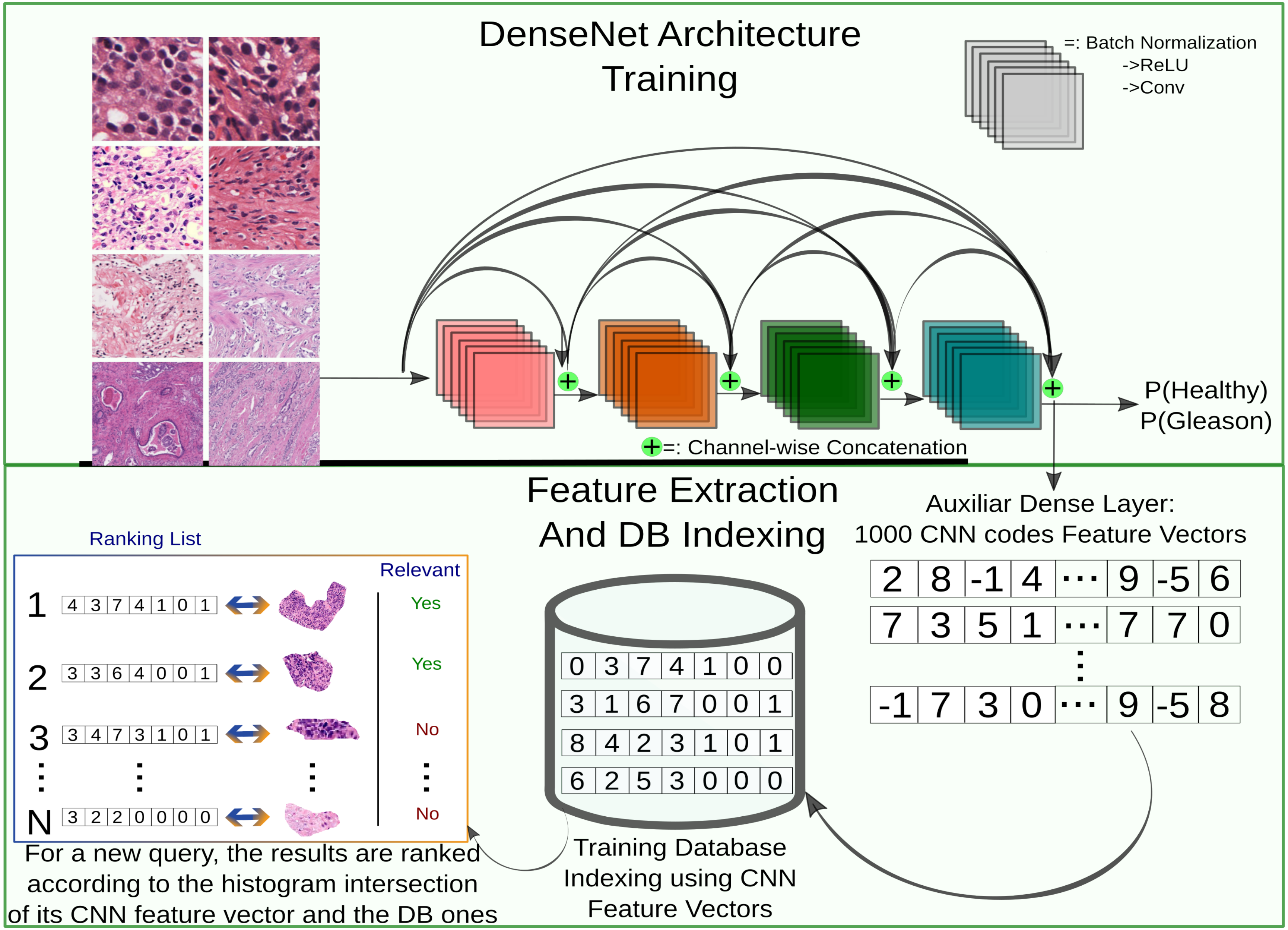
ExaMode

Performance





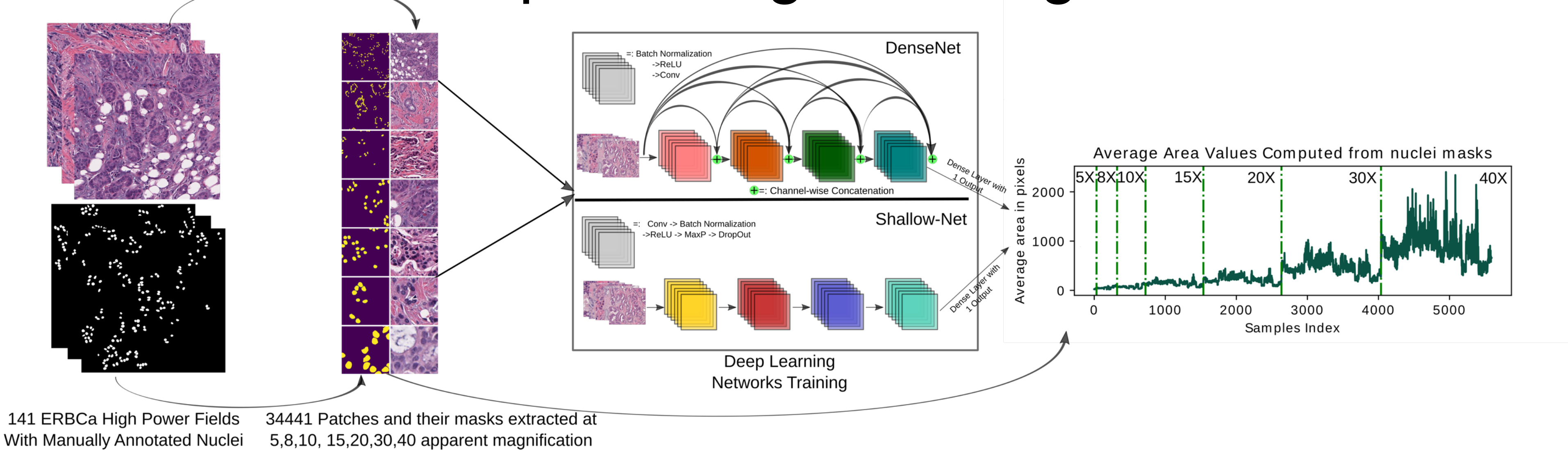
Deep features for retrieval



R. Schaer, S. Otalora, O. Jimenez-del-Toro, M. Atzori, H. Müller, Deep learning based retrieval system for gigapixel histopathology cases and open access literature, Pathology Informatics, 2019.

Magnification regression

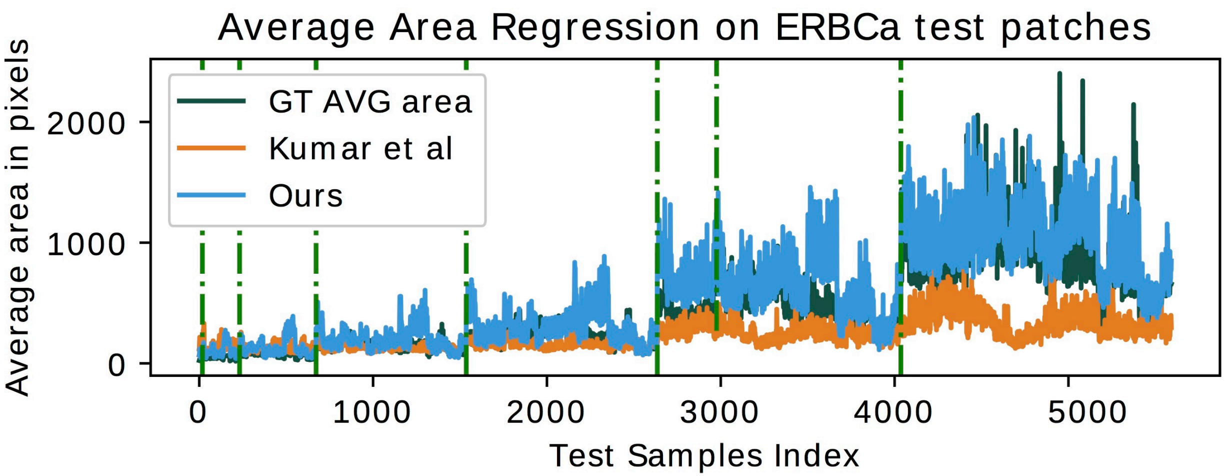
- Magnification (or pixel size) in literature images is not always known
- If we want to compare visual similarity the scale of the structures is essential
  - Unlike object recognition where scale is irrelevant
- Brut force deep learning vs. using nuclei size

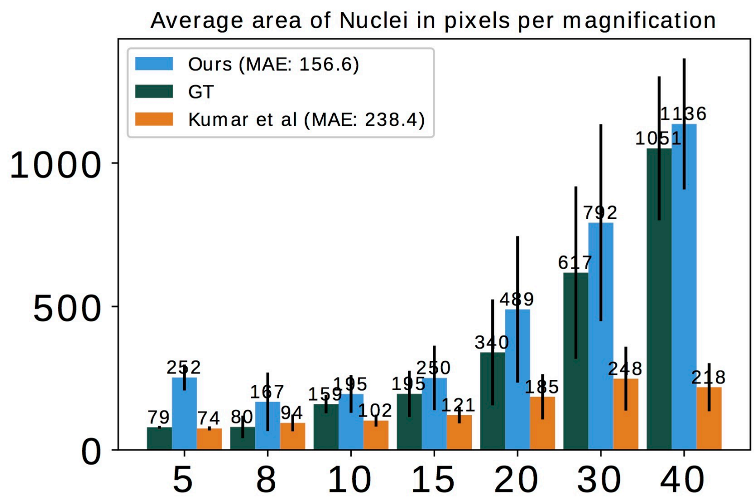


Sebastian Otalora, Vincent Andrearczyk, Manfredo Atzori, **Henning Müller**, Histopathology Nuclei Area Regression for Exploiting Open Access Content Using Deep Learning, *MICCAI workshop on Computational Pathology Imaging (COMPAY) 2018*, Granada, Spain, 2018.

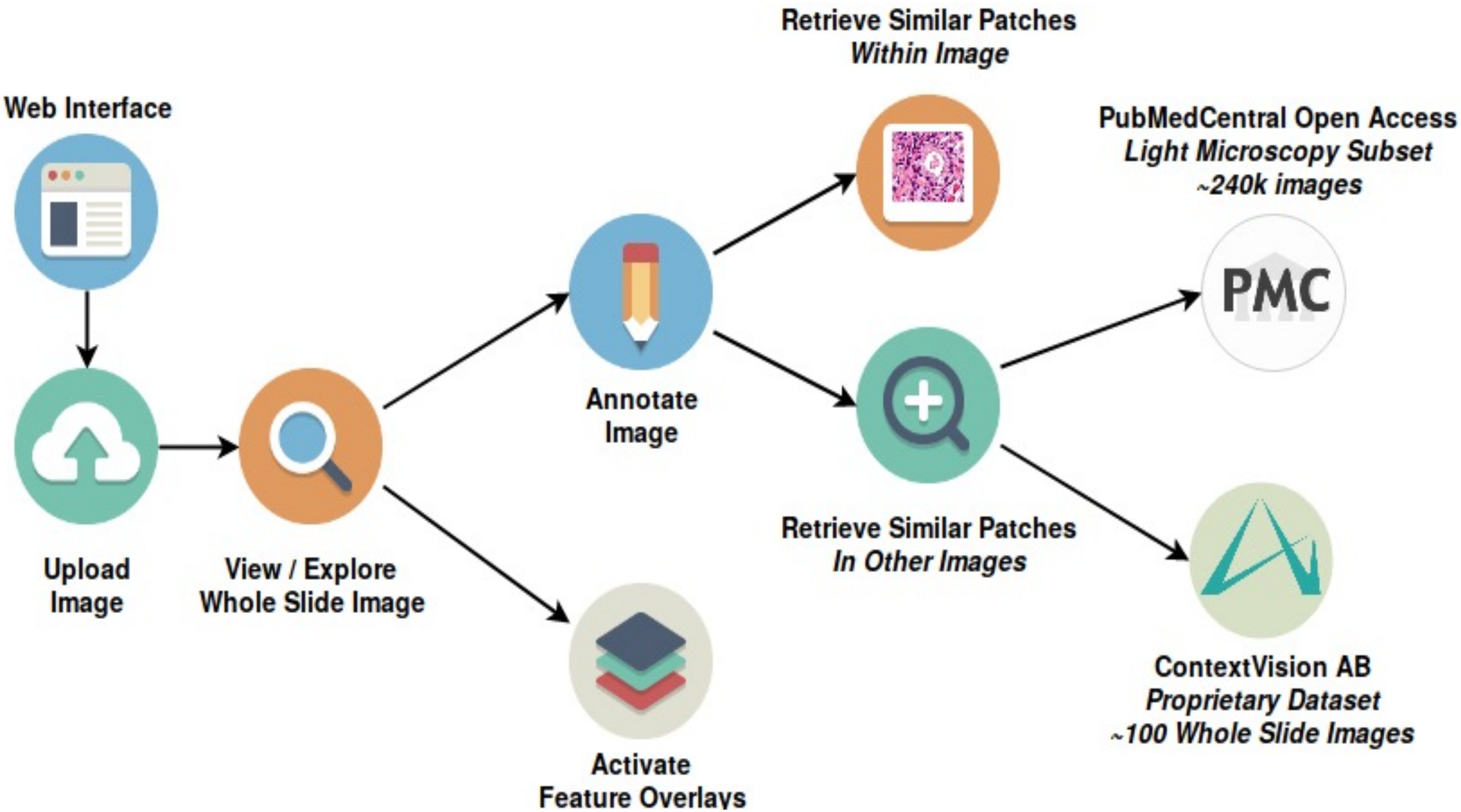
Results of scale regression

- Scale is consistently closer to real scales than other techniques with our technique
  - Not as good as others only for lower magnifications

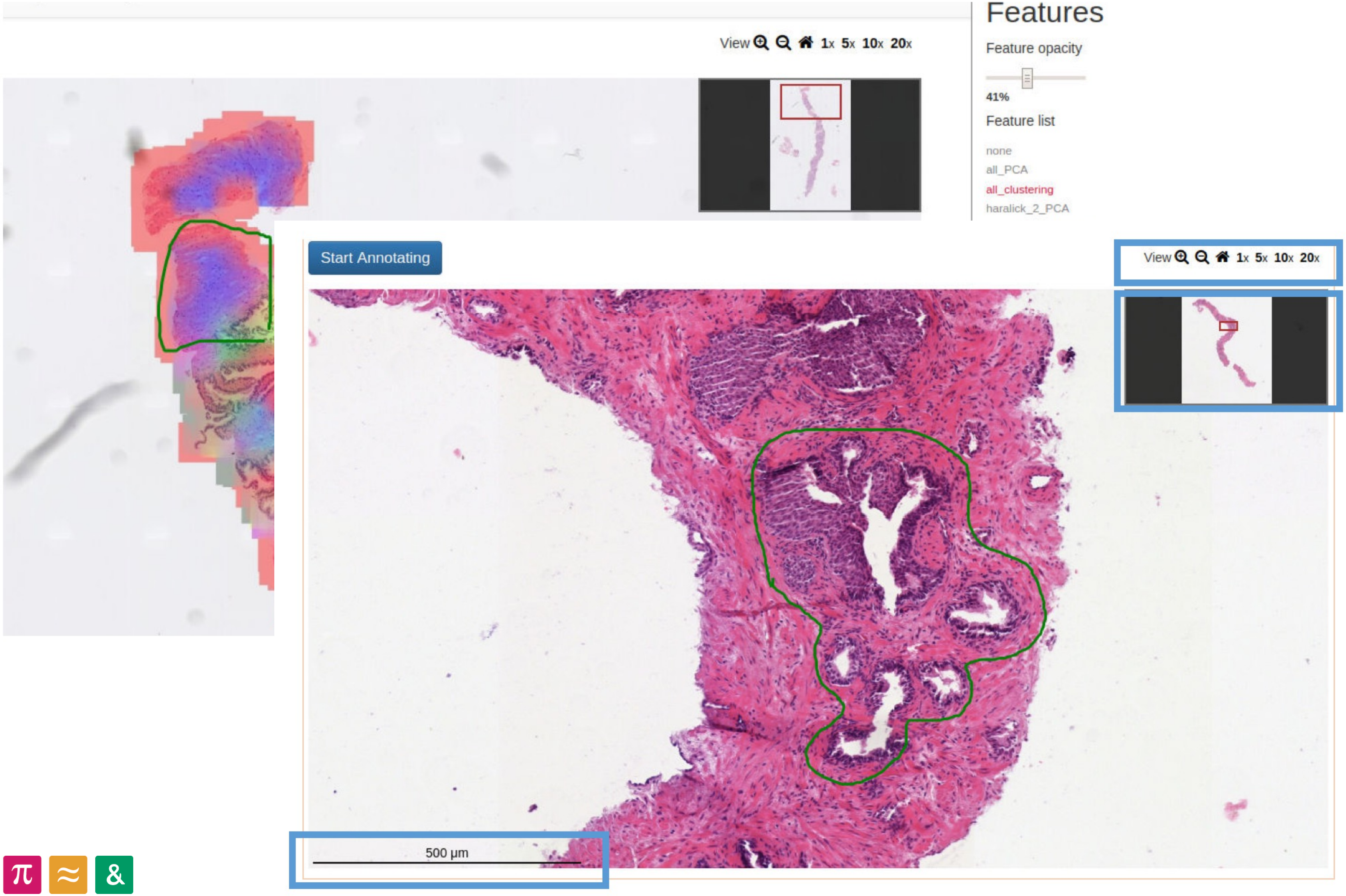




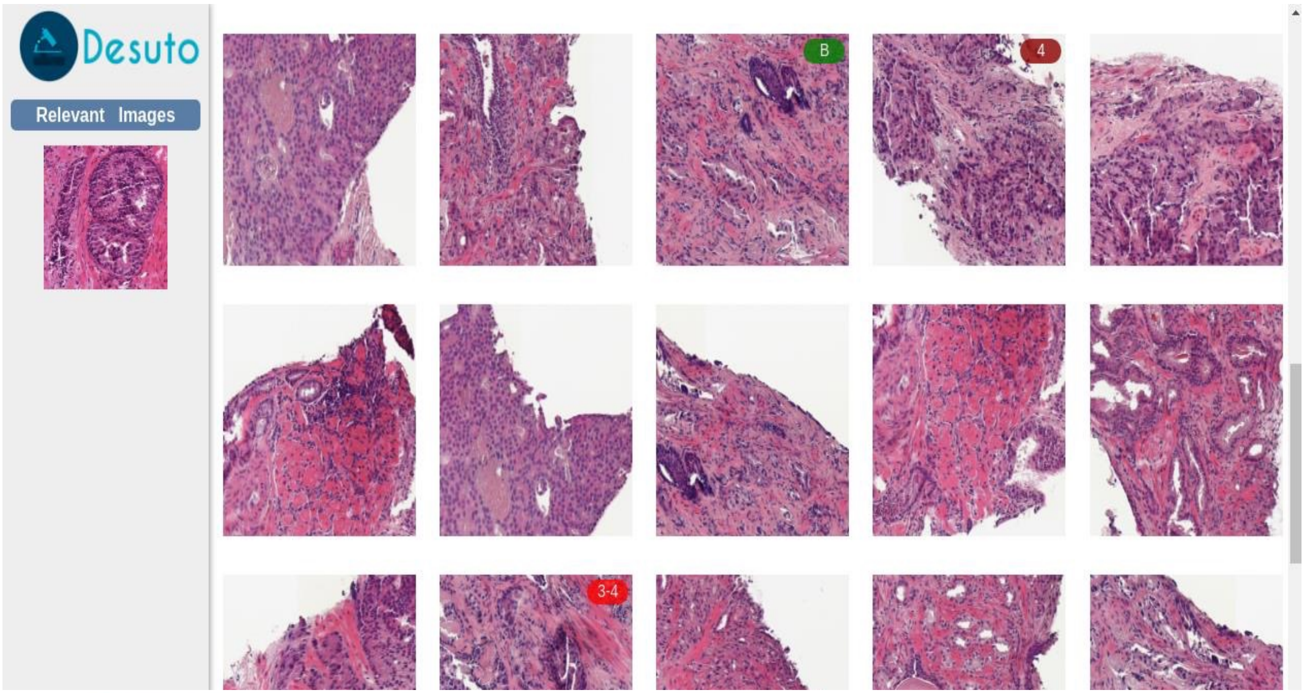
Histopathology retrieval

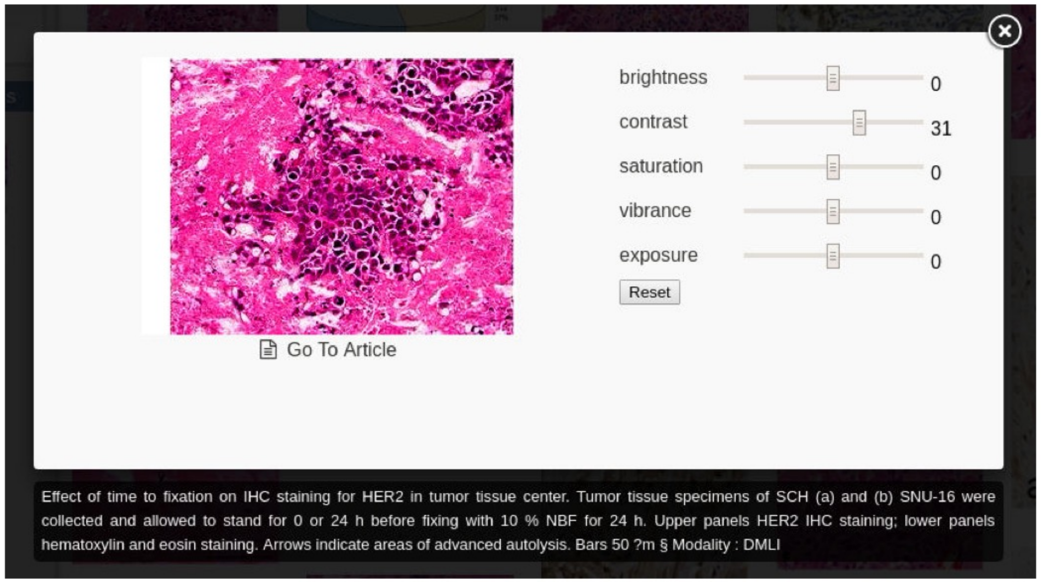


Overlays and annotations

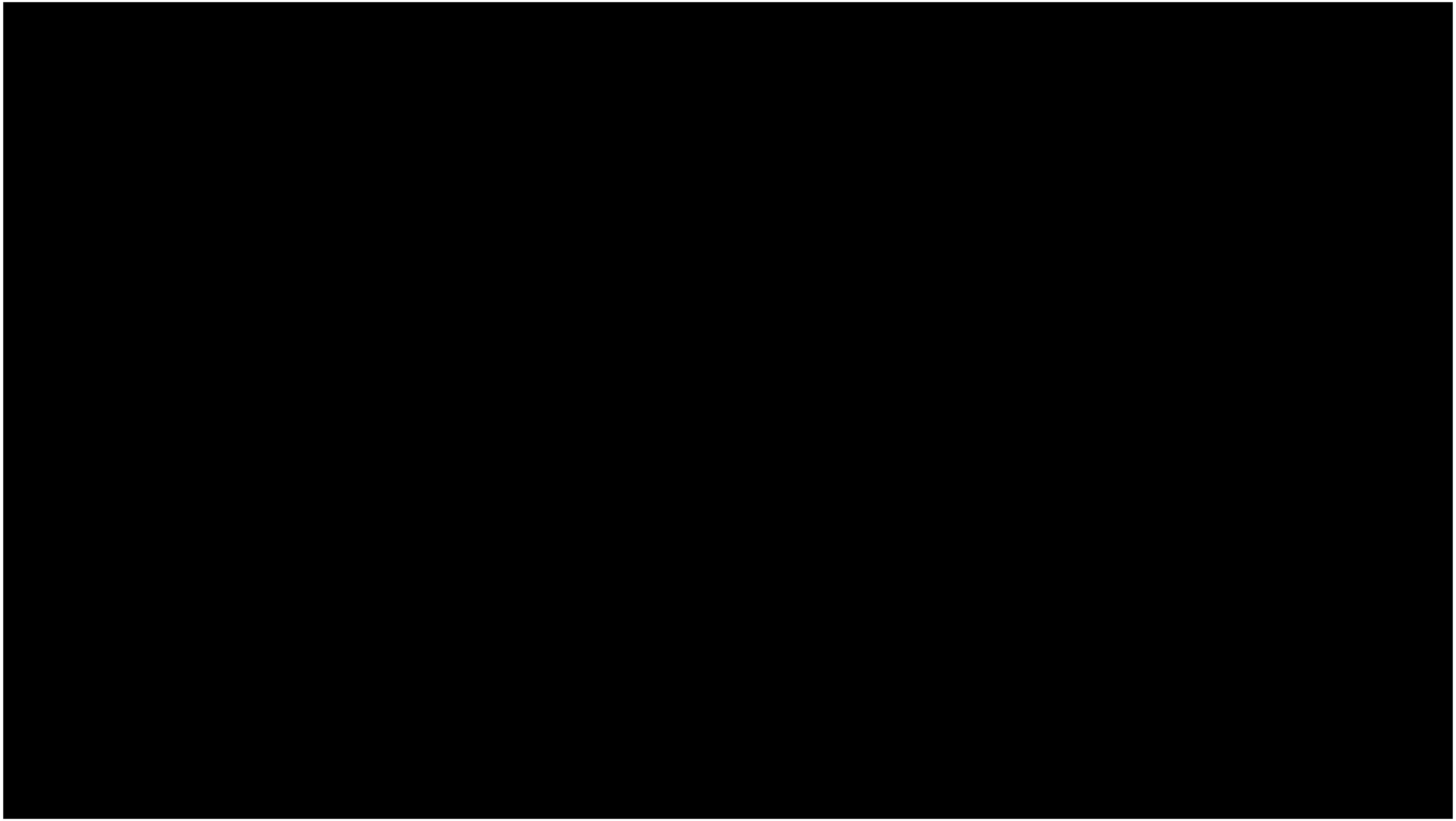


Visual retrieval



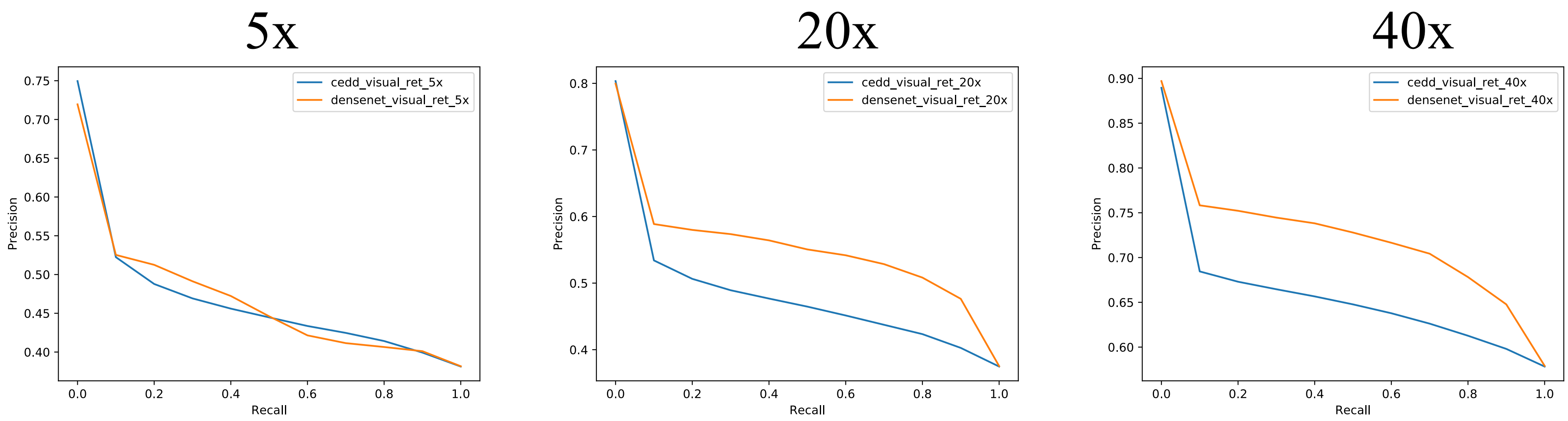


Video of the retrieval system

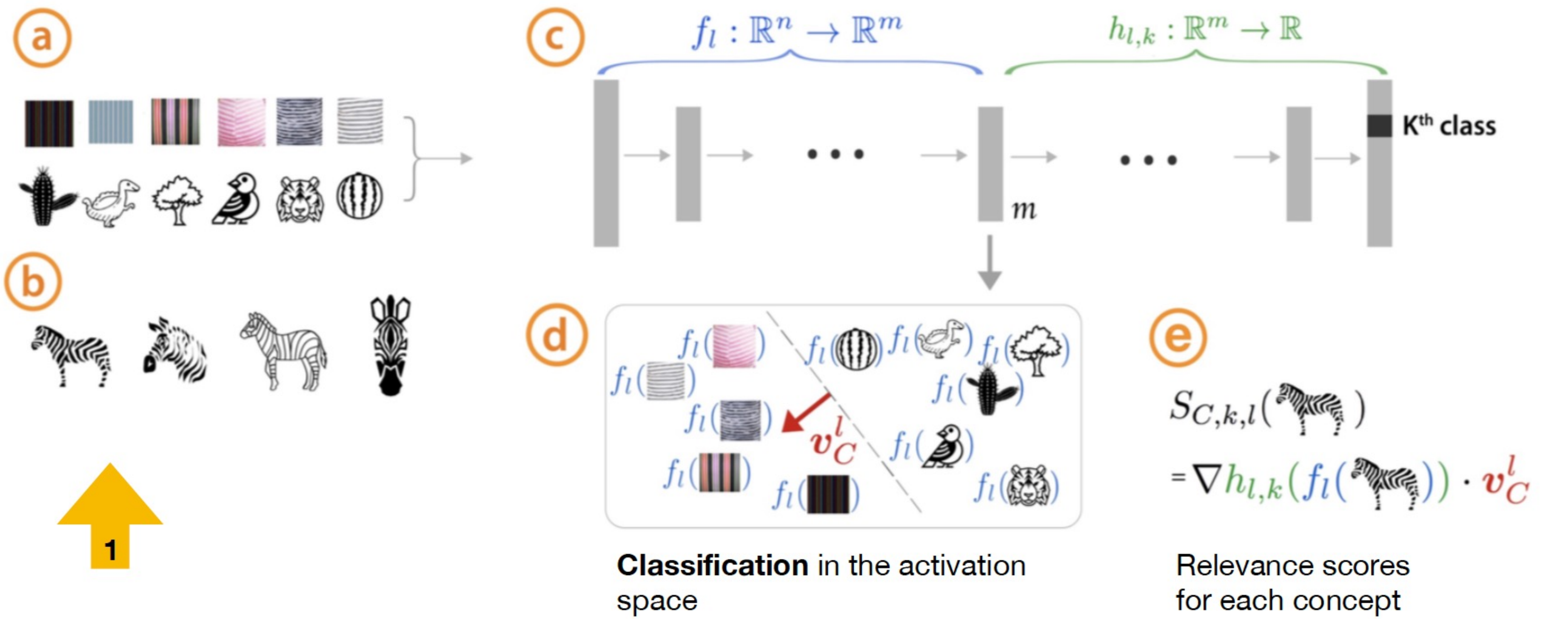


Retrieval performance

- DL features work better at high recall and higher scales
- CEDD is OK at low scale and lower recall
- User tests were done with a small group of pathologists



Interpretability with Concept Activation Vectors



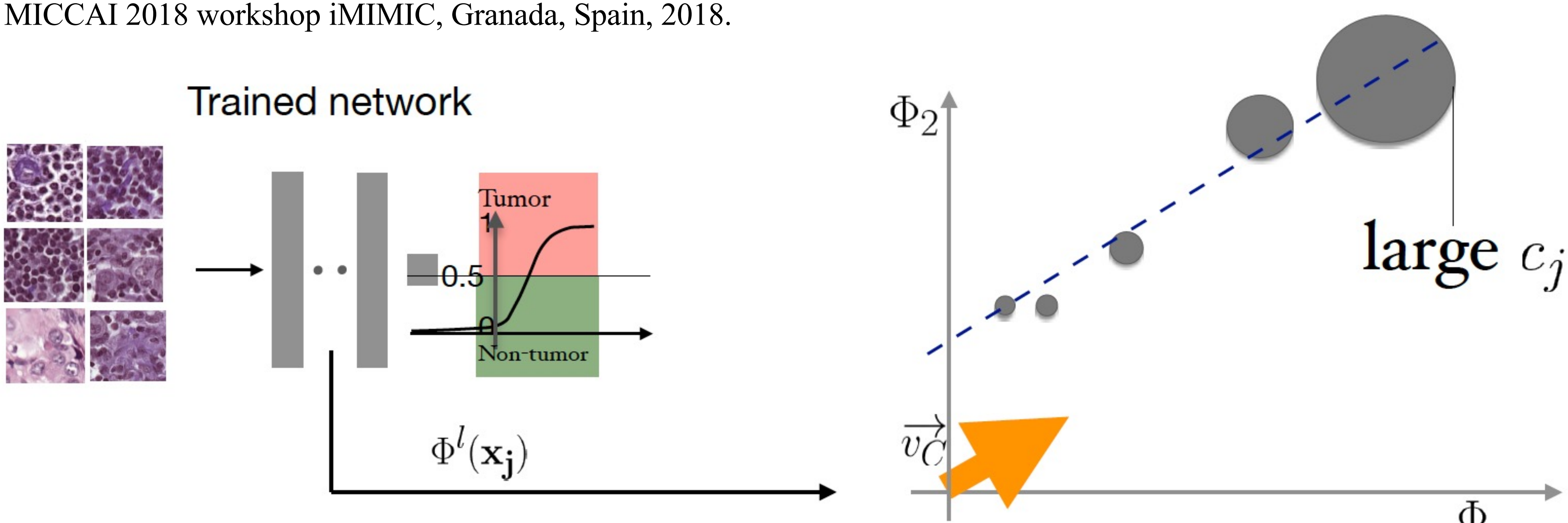
[Kim, 2018]

Kim, B., Wattenberg, M., Gilmer, J., Cai, C., Wexler, J., & Viegas, F. (2018, July). Interpretability beyond feature attribution: Quantitative testing with concept activation vectors (tcav). In *International conference on machine learning* (pp. 2668-2677).

Regression concept vectors

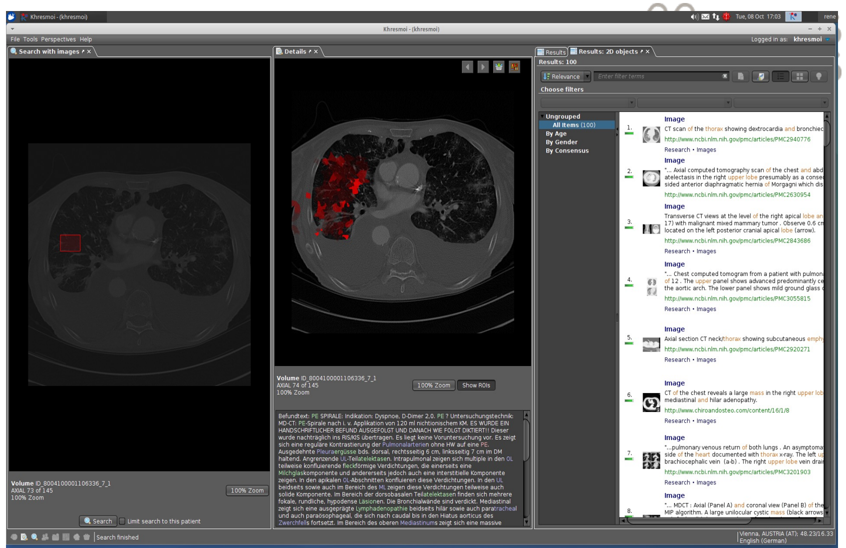
- Identify existing features and check how the decision layers correlate to these features
  - i.e.: nuclei size, internal heterogeneity, borders, …
  - How much can a decision be explained with these?

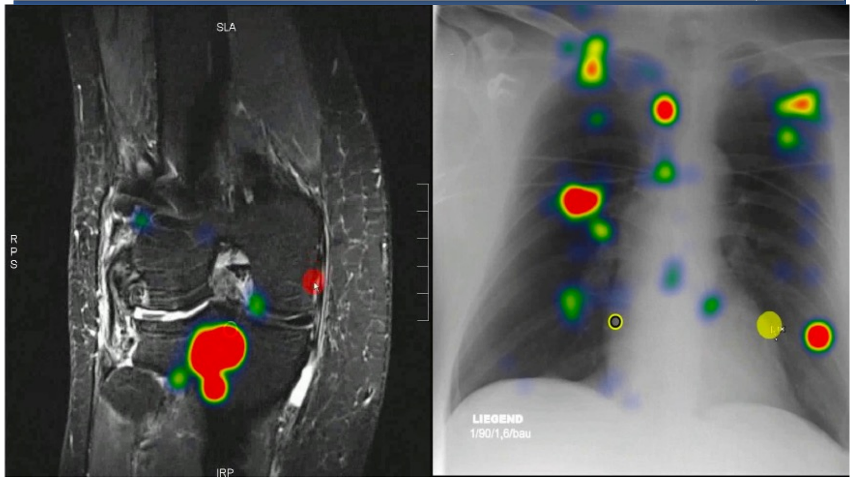
M. Graziani, V. Andreaczyk, H. Müller, Regression Concept Vectors for Bidirectional Explanations in Histopathology, MICCAI 2018 workshop iMIMIC, Granada, Spain, 2018.



Surveys and user tests

- Ask physicians and then build systems on this
- Record the screens
- Outcomes
  - Small things change much!
    - What people are used to
  - Task changes behavior
  - Integration into the workflow
  - Certified labels of found cases, …



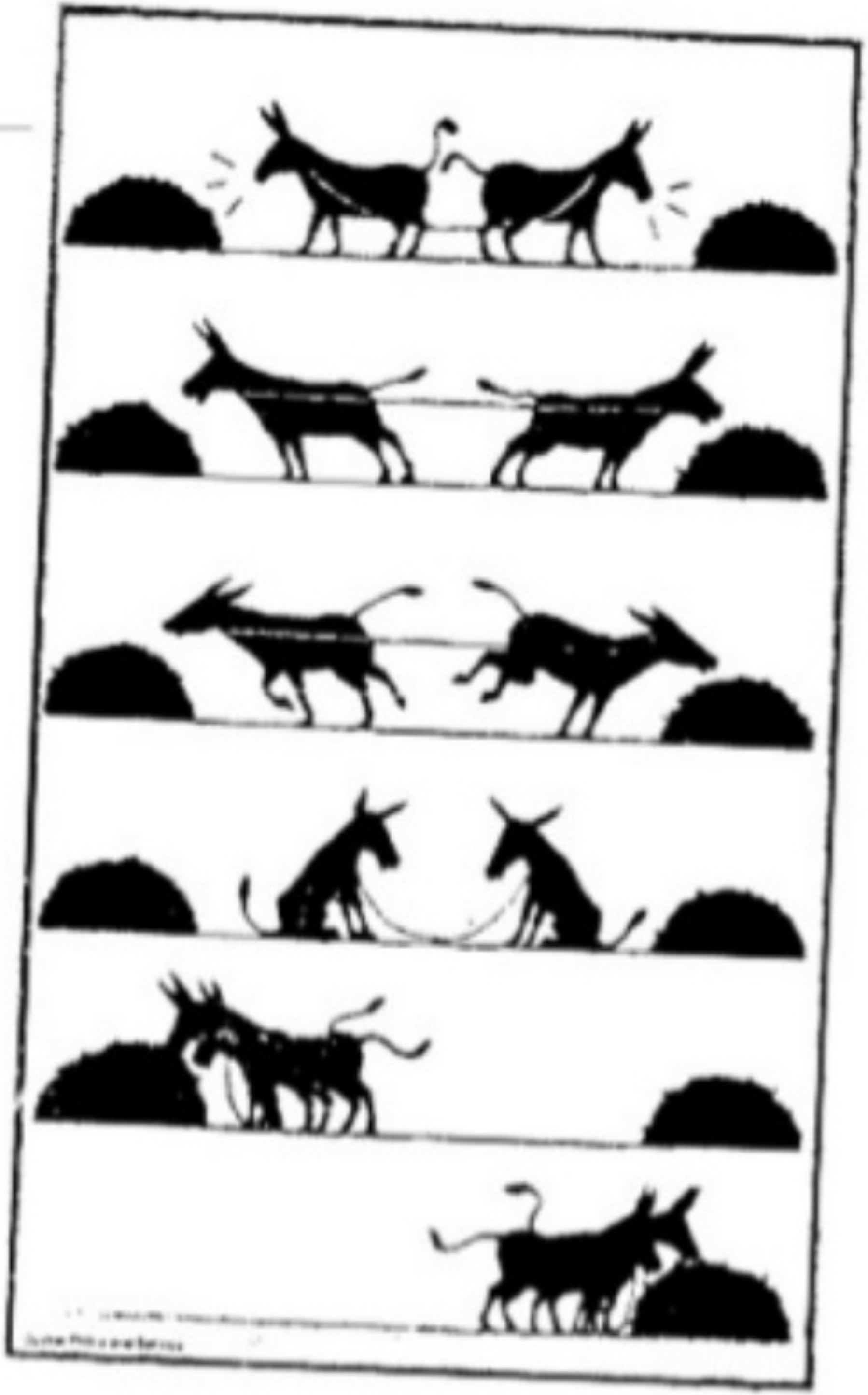


Markonis, D., Holzer, M., Dungs, S., Vargas, A., Langs, G., Kriewel, S., & Müller, H. (2012). A survey on visual information search behavior and requirements of radiologists. *Methods of information in Medicine*, *51*(06), 539-548.

Markonis, D., Holzer, M., Baroz, F., De Castaneda, R. L. R., Boyer, C., Langs, G., & Müller, H. (2015). User-oriented evaluation of a medical image retrieval system for radiologists. *International journal of medical informatics*, *84*(10), 774-783.

Scientific environment

- Competition
- Coopetition
- Cooperation



Early history of ImageCLEF

- 2003: first image retrieval task, 4 participants
- 2004: 17 participants for three tasks (~200 runs)
  - Medical task for visual image retrieval added
- 2005: 24 participants for fours tasks (~300 runs)
  - Two medical tasks
- 2006: 30 participants for four tasks (~300 runs)
  - LTU database of objects for object classification
- 2007: 35 participants (>1000 runs)
  - Hierarchical classification
- 2008: 45 participants submitted results (>2000 runs)
  - 63 registrations, wiki task
- … each year until 2021, 2022 already foreseen

Müller, H., Clough, P., Deselaers, T., & Caputo, B. (Eds.). (2010). *ImageCLEF: experimental evaluation in visual information retrieval* (Vol. 32). Springer.

Medical tasks (2004-2018)

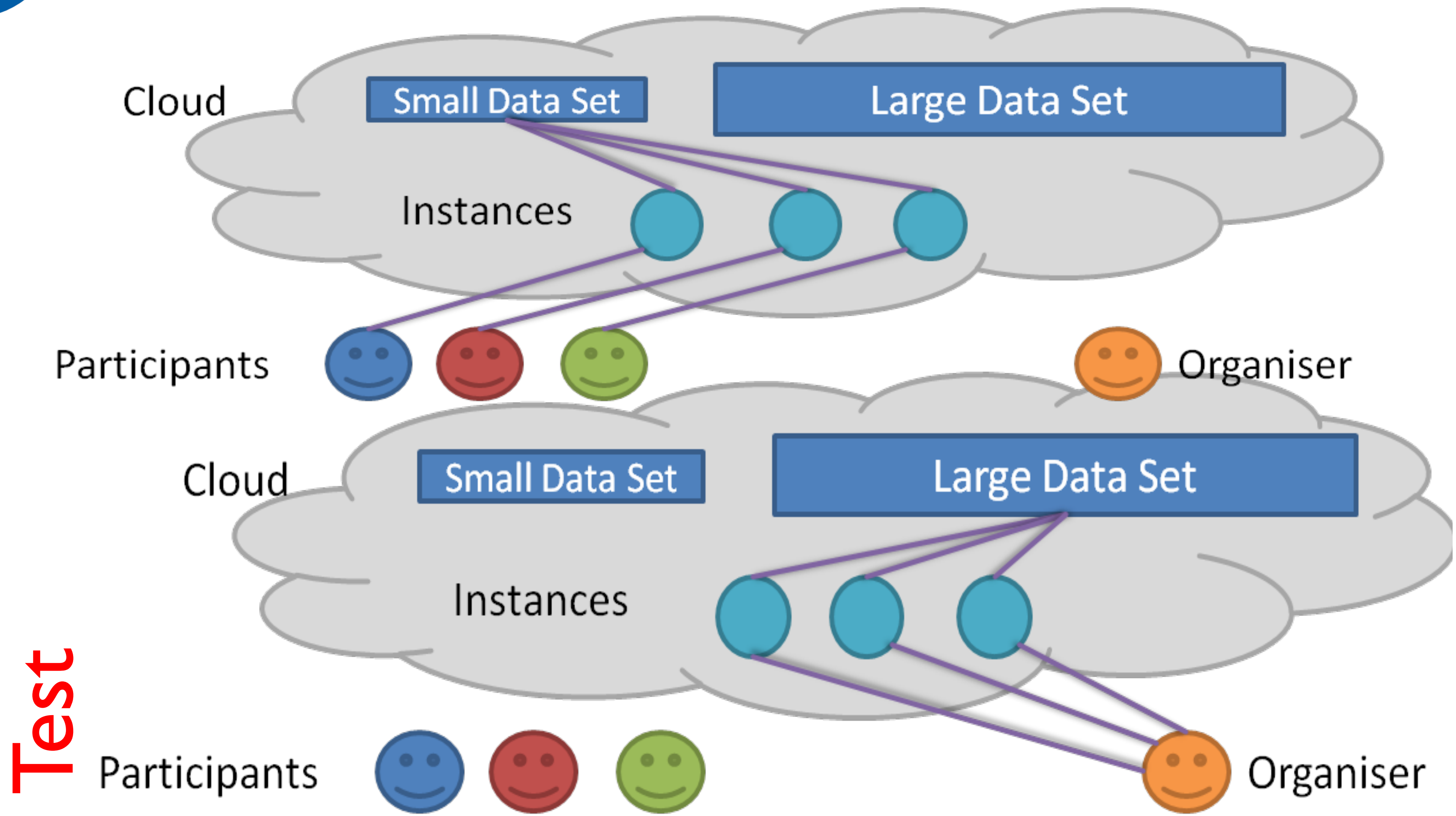
Task type	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
image–based retrieval	x	x	x	x	x	x	x	x	x	x					
image type classification		x	x	x	x	x									
case–based retrieval							x	x	x	x					
image modality classification							x	x	x	x					
compound figure separation										x		x	x		
liver CT annotation											x				
subfigure classification												x	x		
compound figure detection												x	x		
caption prediction												x	x	x	x
tuberculosis														x	x
visual question answering															x

Müller, H., Kalpathy-Cramer, J., & de Herrera, A. G. S. (2019). Experiences from the ImageCLEF medical retrieval and annotation tasks. In *Information Retrieval Evaluation in a Changing World* (pp. 231-250). Springer.

Evaluation-as-a-Service (EaaS)

- Moving the algorithms to the data, not vice versa
  - Required when data are: very large, changing quickly, confidential (medical, commercial, …)
- Different approaches
  - Source code submission, APIs, VMs local or in the cloud, Docker containers, specific frameworks
- Allows for continuous evaluation, component-based evaluation, total reproducibility, updates, …
  - Workshop March 2015 in Sierre on EaaS
  - Workshop November 2015 in Boston on cloud-based evaluation (http://www.martinos.org/cloudWorkshop/)

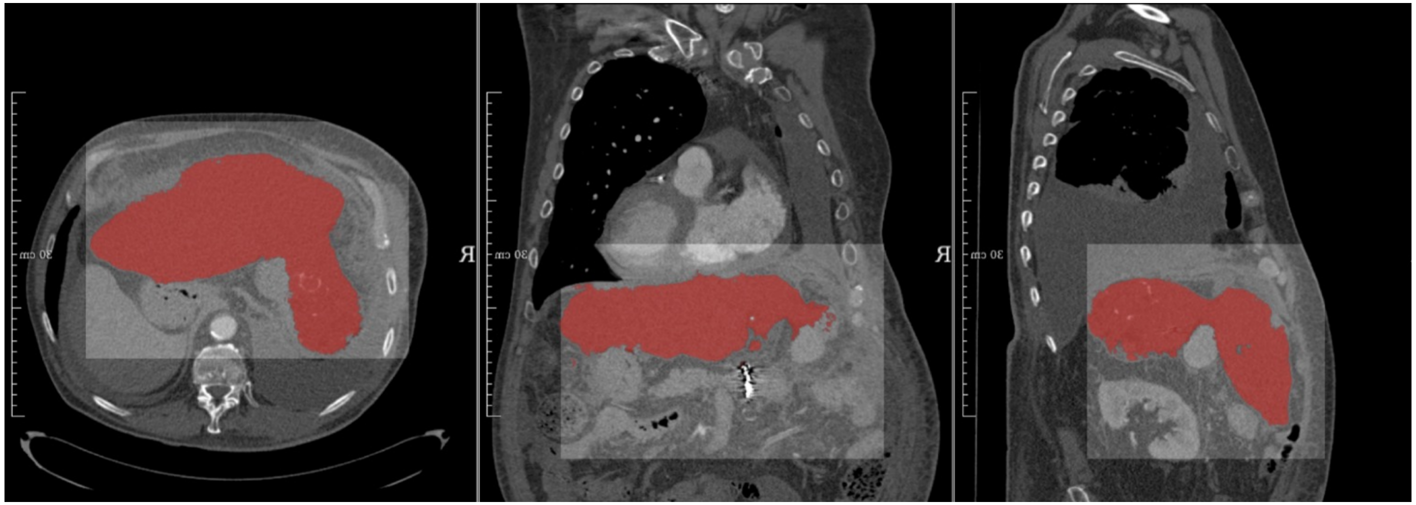
Hopfgartner, F., Hanbury, A., Müller, H., Eggel, I., Balog, K., Brodt, T., ... & Kato, M. P. (2018). Evaluation-as-a-service for the computational sciences: overview and outlook. Journal of Data and Information Quality (JDIQ), 10(4), 1-32.



A. Hanbury, H. Müller, G. Langs, M. A. Weber, B. H. Menze, T. Salas Fernandez, Bringing the algorithms to the data: cloud–based benchmarking for medical image analysis, CLEF conference, Springer Lecture Notes in Computer Science, 2012.

Case-based similarity retrieval

ROI



Organ Mask

RadLex terms

RID480, Aorta	RID5227,Sklerose,0
RID58,Leber	RID3822,Zirrhose,0
RID1384,Mediastinum	RID3798,Lymphadenopathie,0
RID1327,Oberlappen der linken Lunge	RID3953,Granulom,0
RID1362,Pleura	RID4872,Erguss,0
RID1315,Unterlappen der rechten Lunge	RID28493,Atelektase,0

3D Volume

O. A. Jimenez del Toro, A. Hanbury, G. Langs, A. Foncubierta Rodriguez, H. Müller,
Overview of the VISCERAL Retrieval Benchmark 2015, Proceedings of MRMD 2015,
Springer LNCS 9059, 2015.

Conclusions

- Medical image retrieval has evolved strongly in the past 25 years
  - Many new techniques, types of interaction, …
- Many resources for medical images are now available openly (TCIA, TCGA, Zenodo, …)
  - But specific annotations are hard to get
  - Learn from weak labels
- Clinical use is still relatively rare
  - Requires certification, good integration, user tests
- All available data need to be used (text, images, …)
  - Multimodal retrieval/integration is the norm!
    - Towards including temporal data

Contact

European Commission | Horizon 2020 European Union funding for Research & Innovation

- More information can be found at
  - http://medgift.hevs.ch/
  - http://publications.hevs.ch/
  - http://www.examode.eu/
- Contact:
  - Henning.mueller@hevs.ch