Extracting visual information from medical images

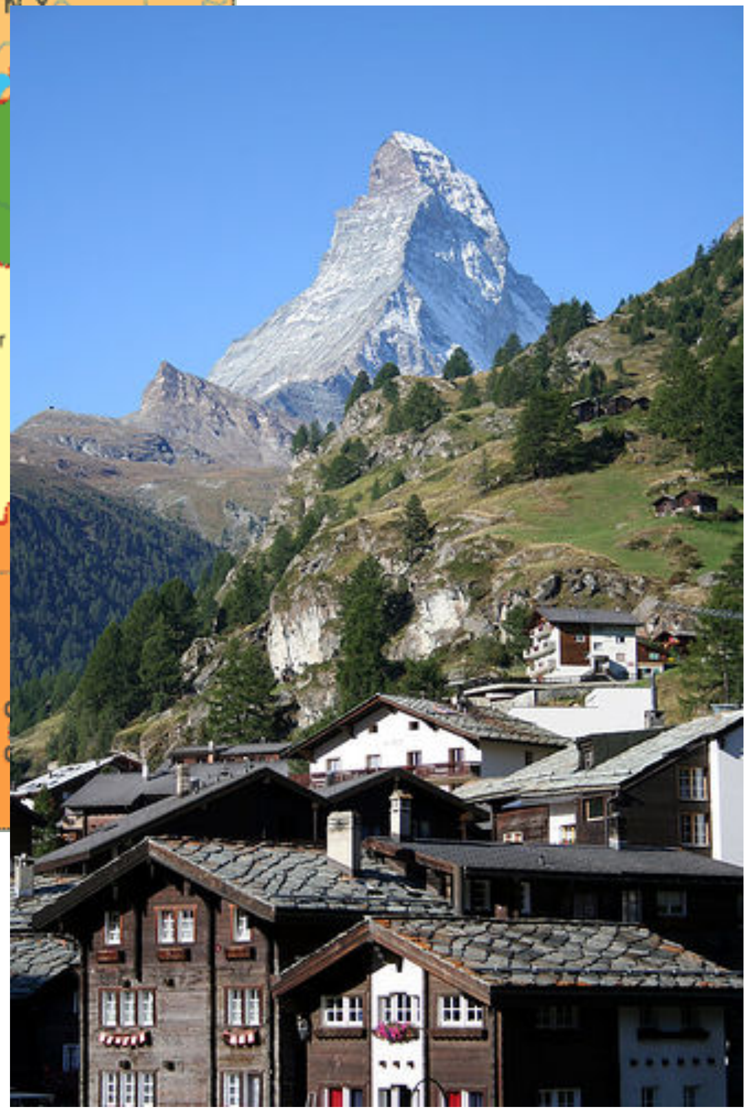
Henning Müller
Jayashree Kalpathy Cramer

Who I am

- Medical informatics studies in Heidelberg, Germany (1992-1997)
  - Exchange with Daimler Benz research, USA
- PhD in image processing, image retrieval, 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)

Where I am

Hes·so VALAIS WALLIS
School of Management & Tourism



Your background

- Who has a computer vision background?
- Who has a good background in machine learning?
- Who is working in a clinical position?

Objectives of this course part

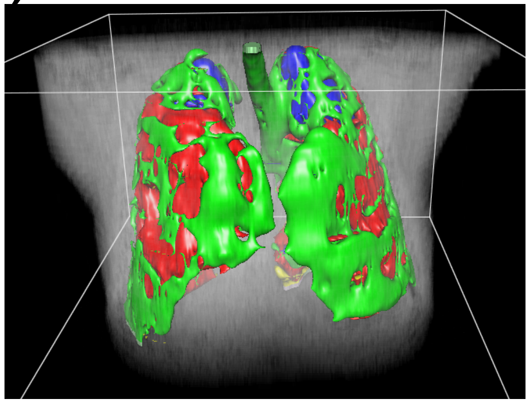
- Learn about how to extract visual information from images automatically
  - For decision support, similar case retrieval, etc.
- Make the link between clinical data and visual data algorithmically (influence of age on images...)
- How to make the different between diversity of anatomy and the small regions of interest in images carrying most of the information
- Separate the model-driven aspects of information extraction from the data-driven parts
  - Machine learning in medical imaging

Links with other courses

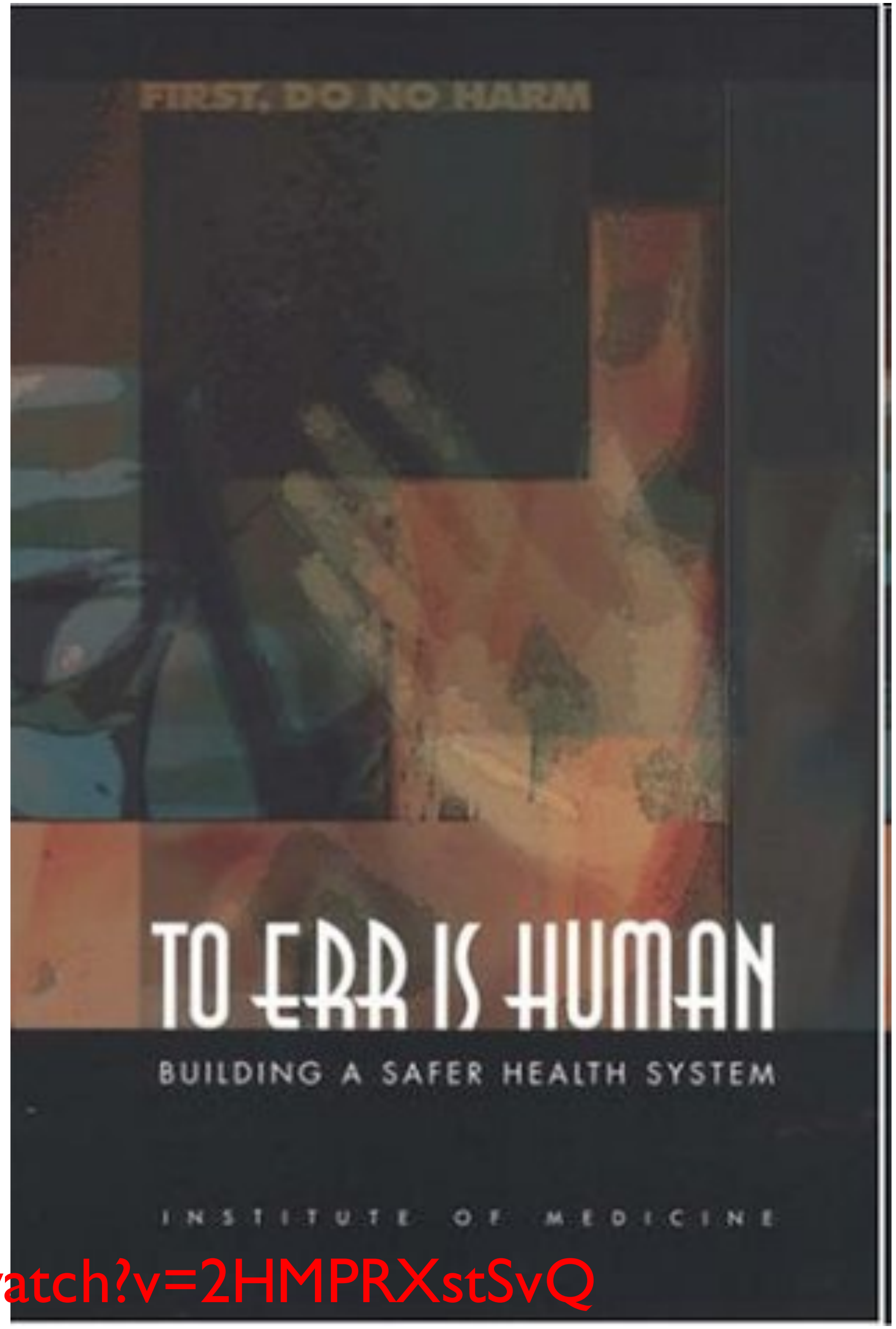
- Characteristics of images produced for research and for clinical work are essential
  - DICOM header information, protocols
- Visual information of images (content) is of limited usefulness if the context is unknown
- Much can be learned from interactive image analysis and visualization for automation
  - 3D Slicer
- Clinical applications are always the final goal, so translational aspects are important

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

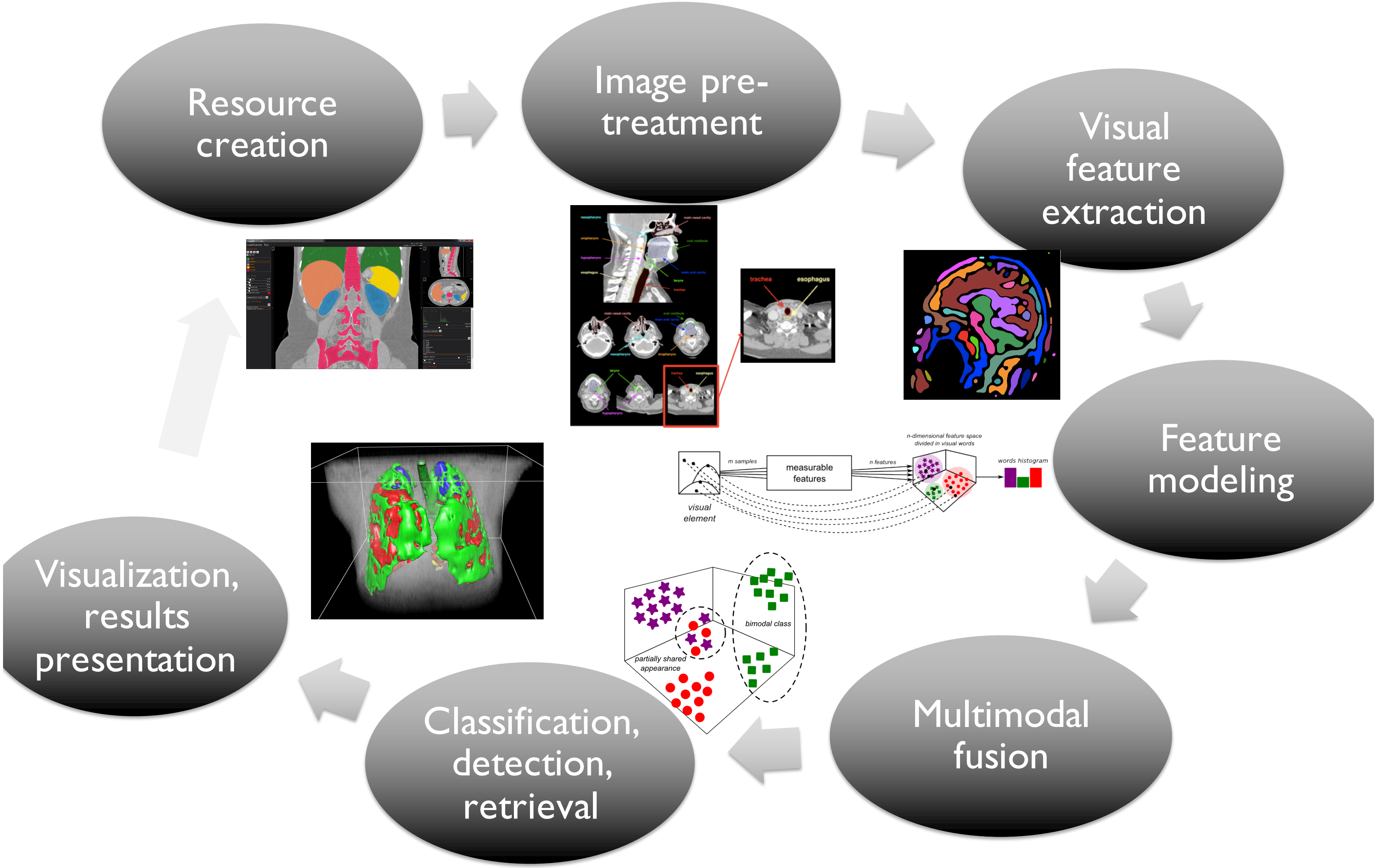


Why decision support?

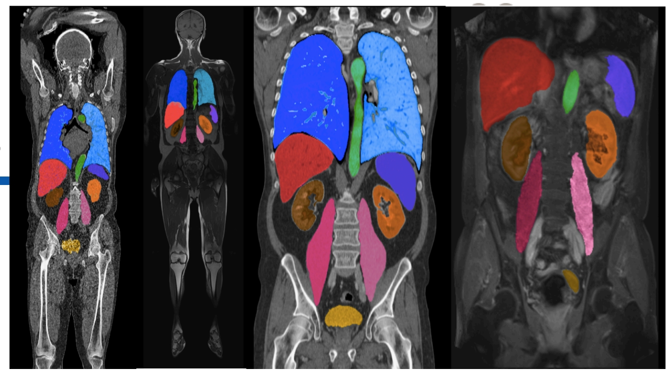


Geoff Hinton on radiology
https://www.youtube.com/watch?v=2HMPRXstSvQ

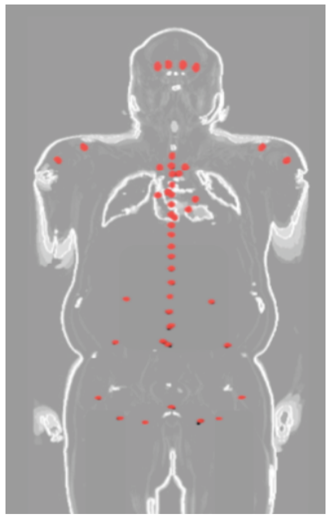
Steps in visual decision support

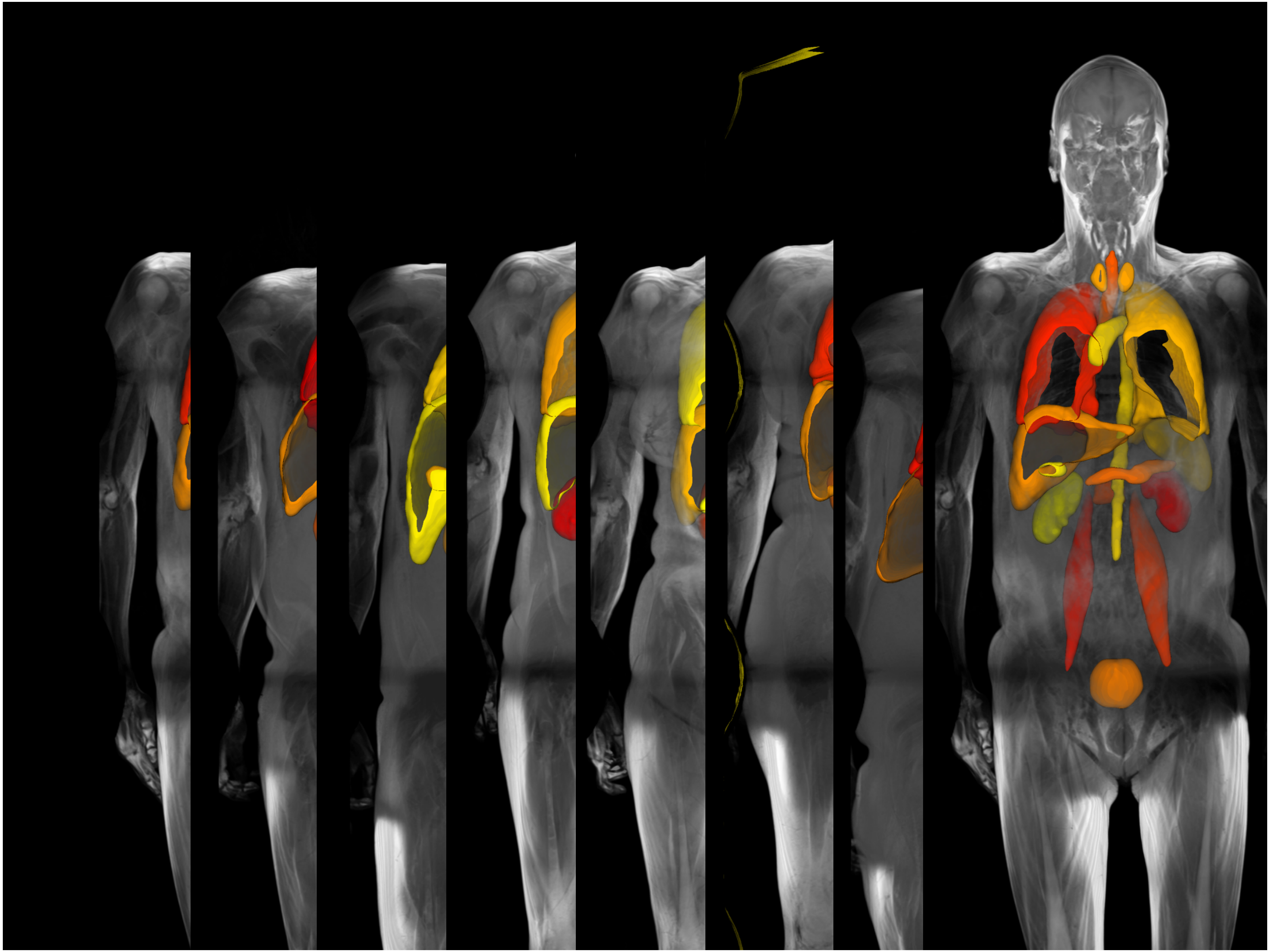


Annotating medical image regions



- 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





Automatic segmentation techniques

- Based purely on grey levels/morphology (lung)
  - Requiring often some cleaning (on borders, holes)
- Simple region growing based on grey levels/texture
  - Based on a seed point in the structure
- Atlas-based annotations
  - Use manually annotated examples and register them to a new case
- Model-based segmentations using shape priors
- Deep learning-based segmentations

Comparing two algorithms

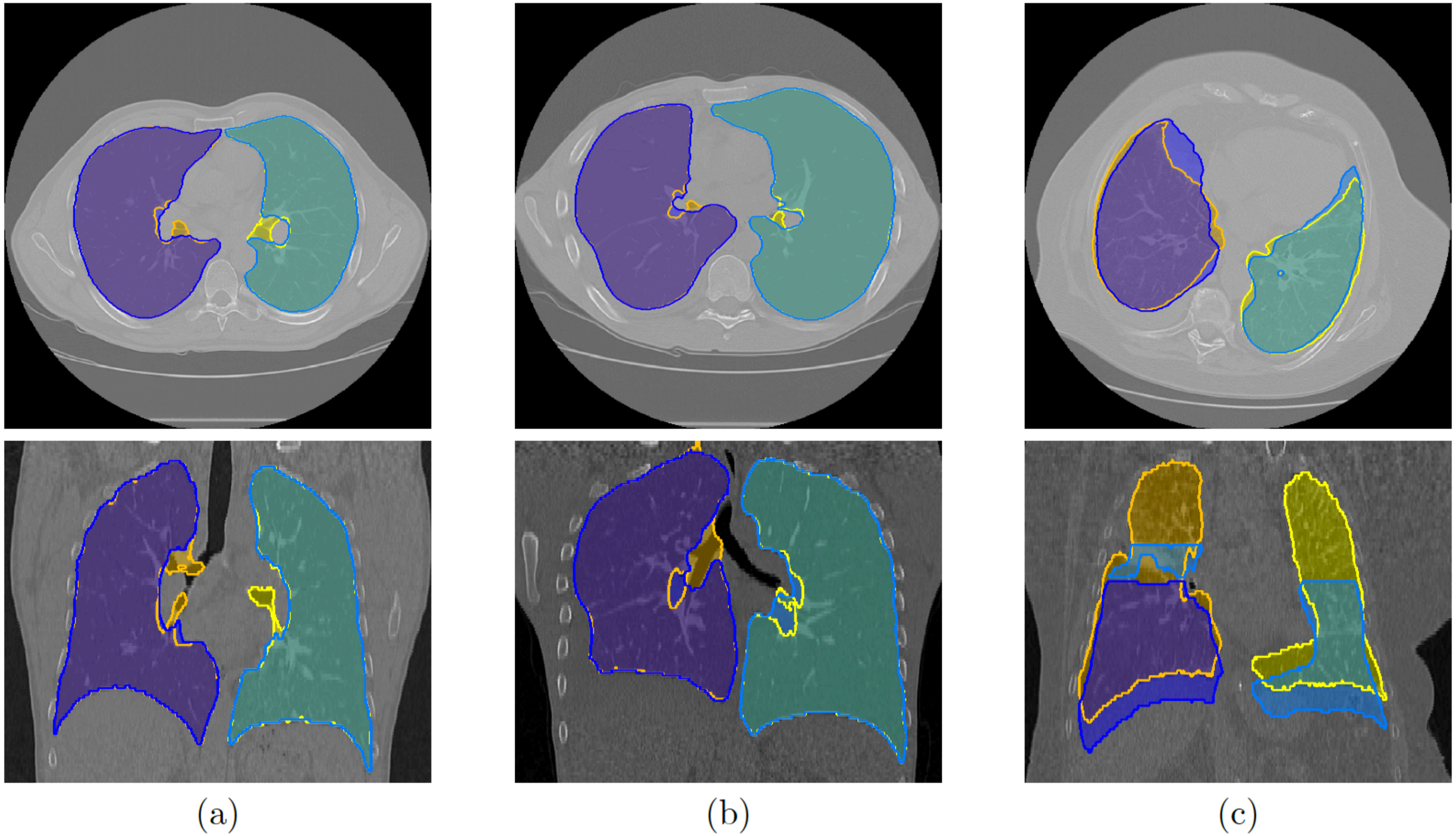
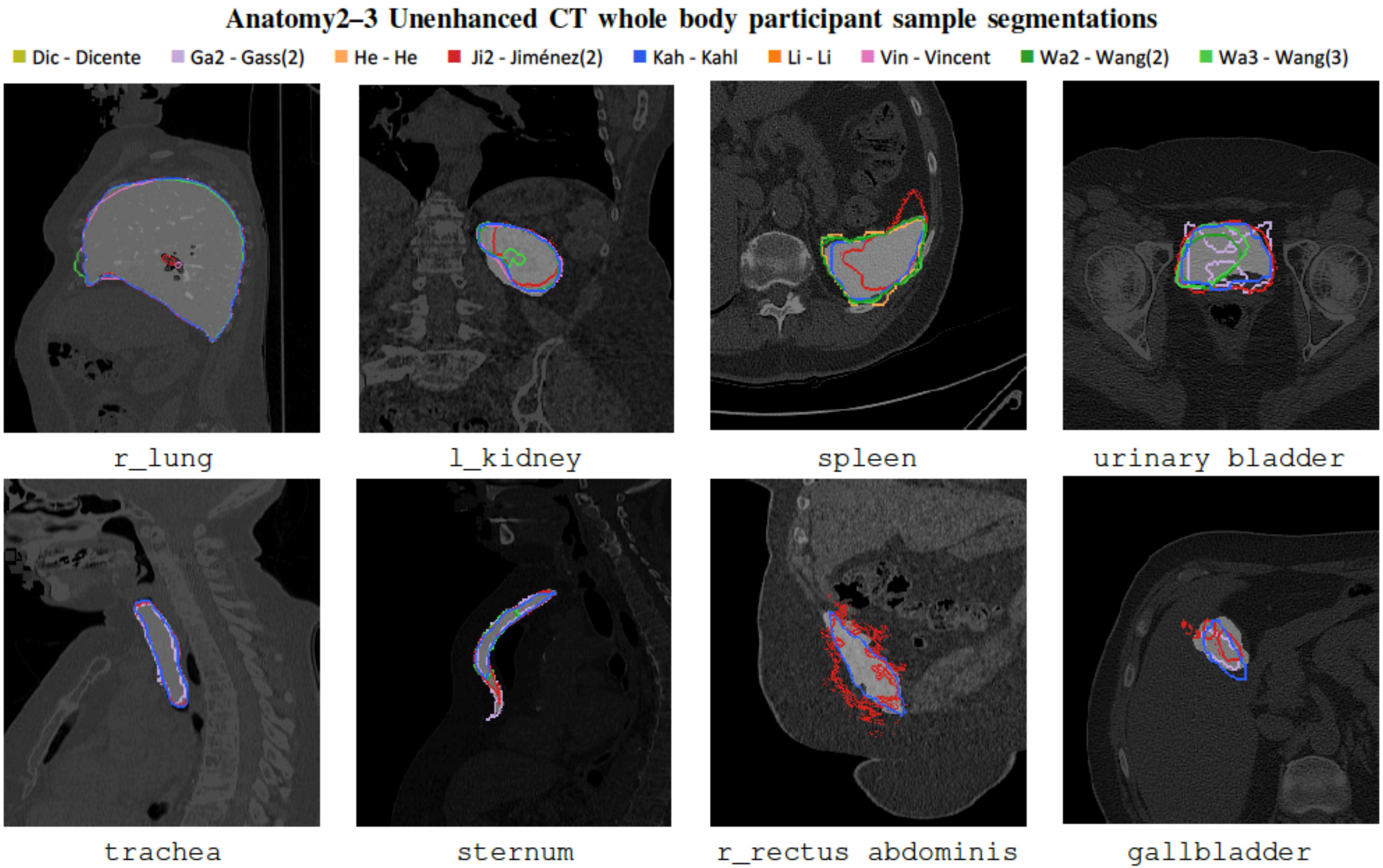


Figure 1: Three examples of the lung segmentation obtained by the two segmentation methods (axial and coronal views). The CIP segmentation is colored in dark blue (right lung) and light blue (left lung) and the DBC segmentation in orange (right lung) and yellow (left lung). When comparing both segmentations, examples (a) and (b) had a Dice coefficient of 0.9920 and 0.9943 respectively. On the other hand, (c) had a Dice coefficient of 0.6788 and was thus discarded from the final database.

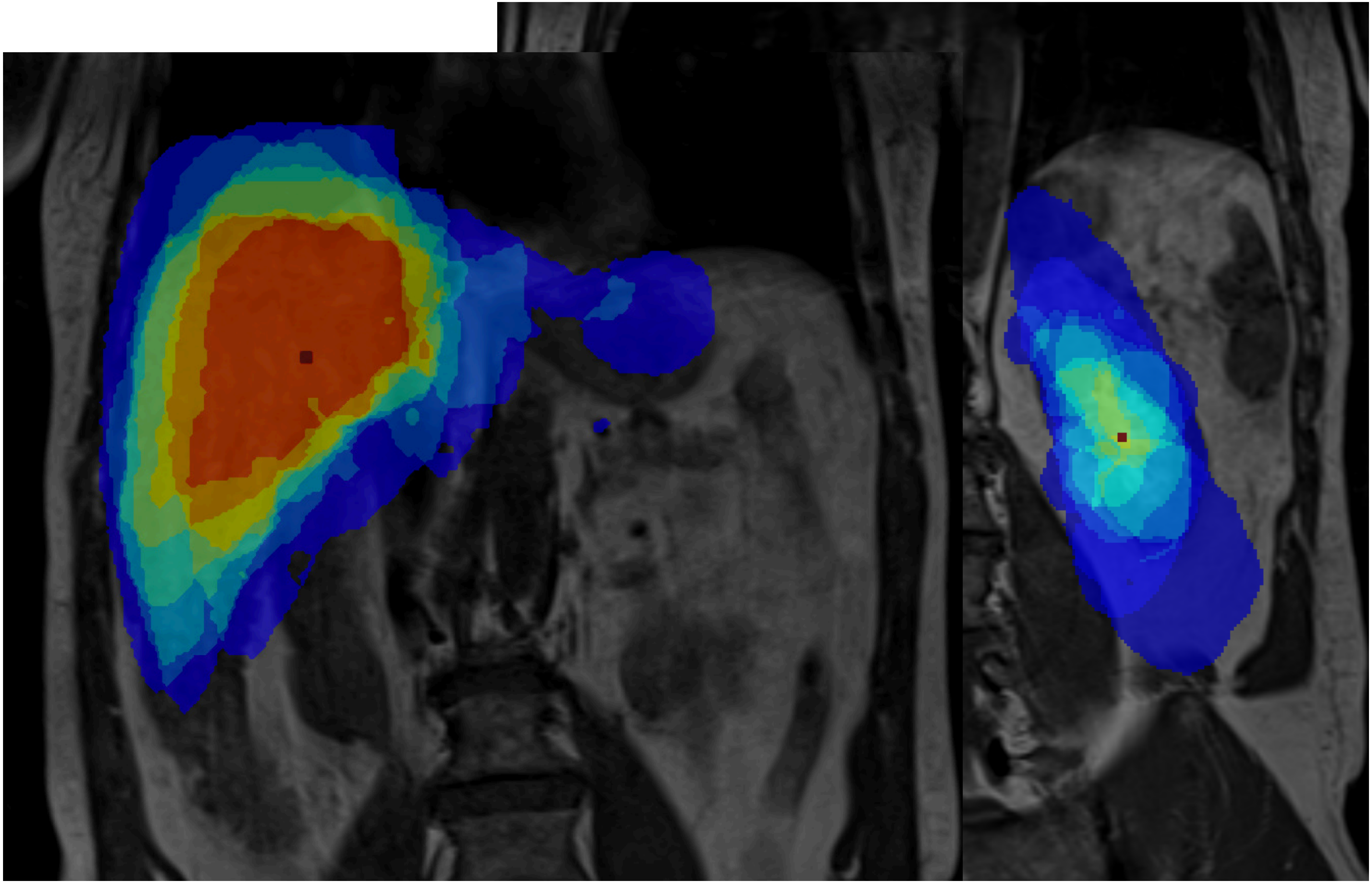
Yashin Dicente Cid, Artem Mamonov, Andrew Beers, Armin Thomas, Vassili Kovalev, Jayashree Kalpathy-Cramer, Henning Müller, Making Sense of Large Datasets without Annotations: Analyzing Age-related Correlations from Lung CT Scans, SPIE MI, 2017.

Comparing several organs/tools



Oscar Alfonso Jiménez del Toro et al., Cloud–based Evaluation of Organ Segmentation and Landmark Detection Algorithms: VISCERAL Anatomy Benchmarks, *IEEE Transactions on Medical Imaging*, 2016.

Example atlas-based segmentation



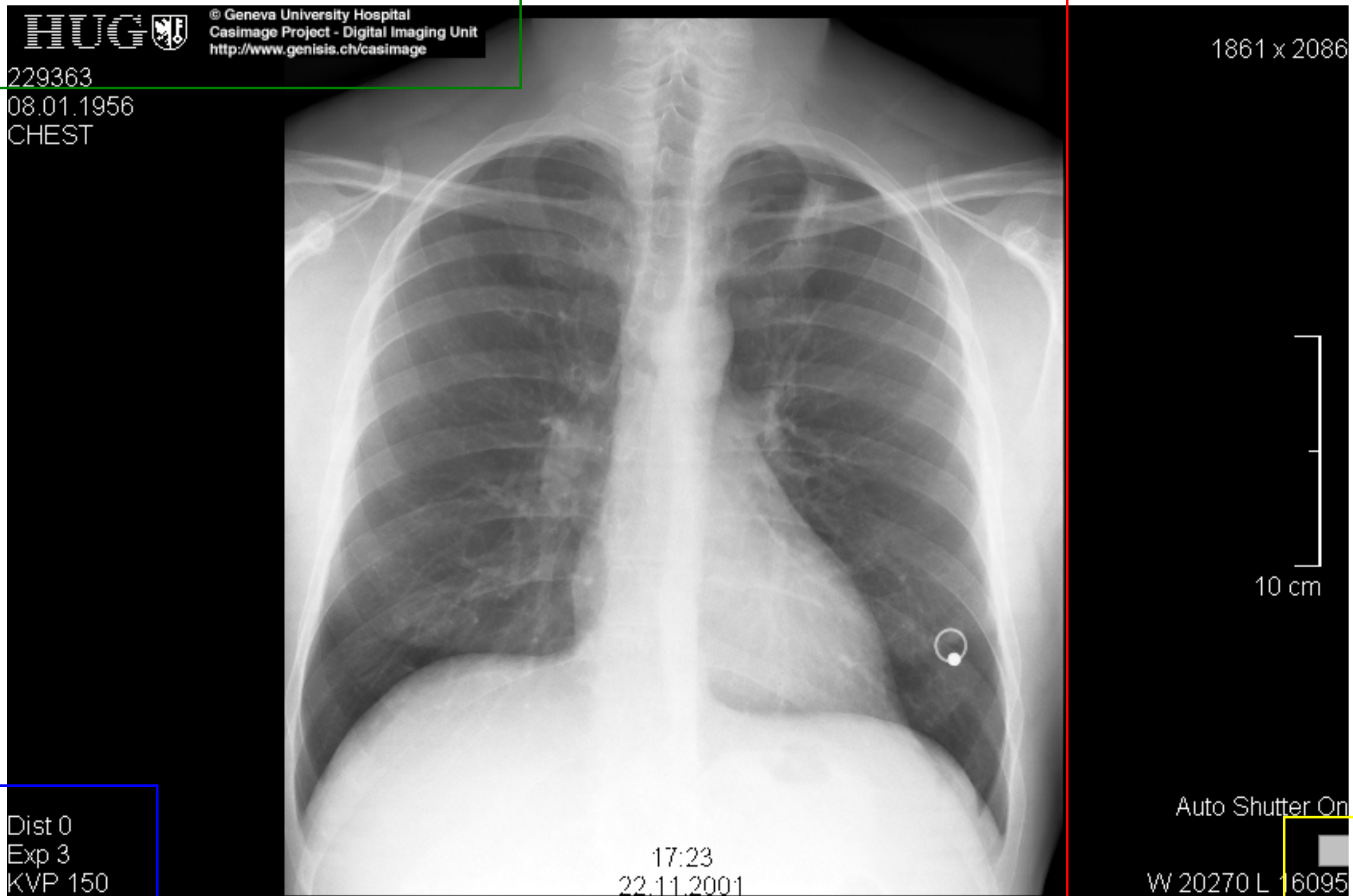
Detection of regions of interest

- Avoids having to read the entire image data (efficiency) or missing important findings
  - Finding a needle in a haystack
  - Incidental findings, how to deal with them
- Make results less observer-dependent/subjective
  - Extract quantitative criteria locally
- Input for a more detailed analysis of regions of interest, removing noise
  - Comparing apples with apples …
  - Automating the analysis workflow

Removing noise

logo

Text

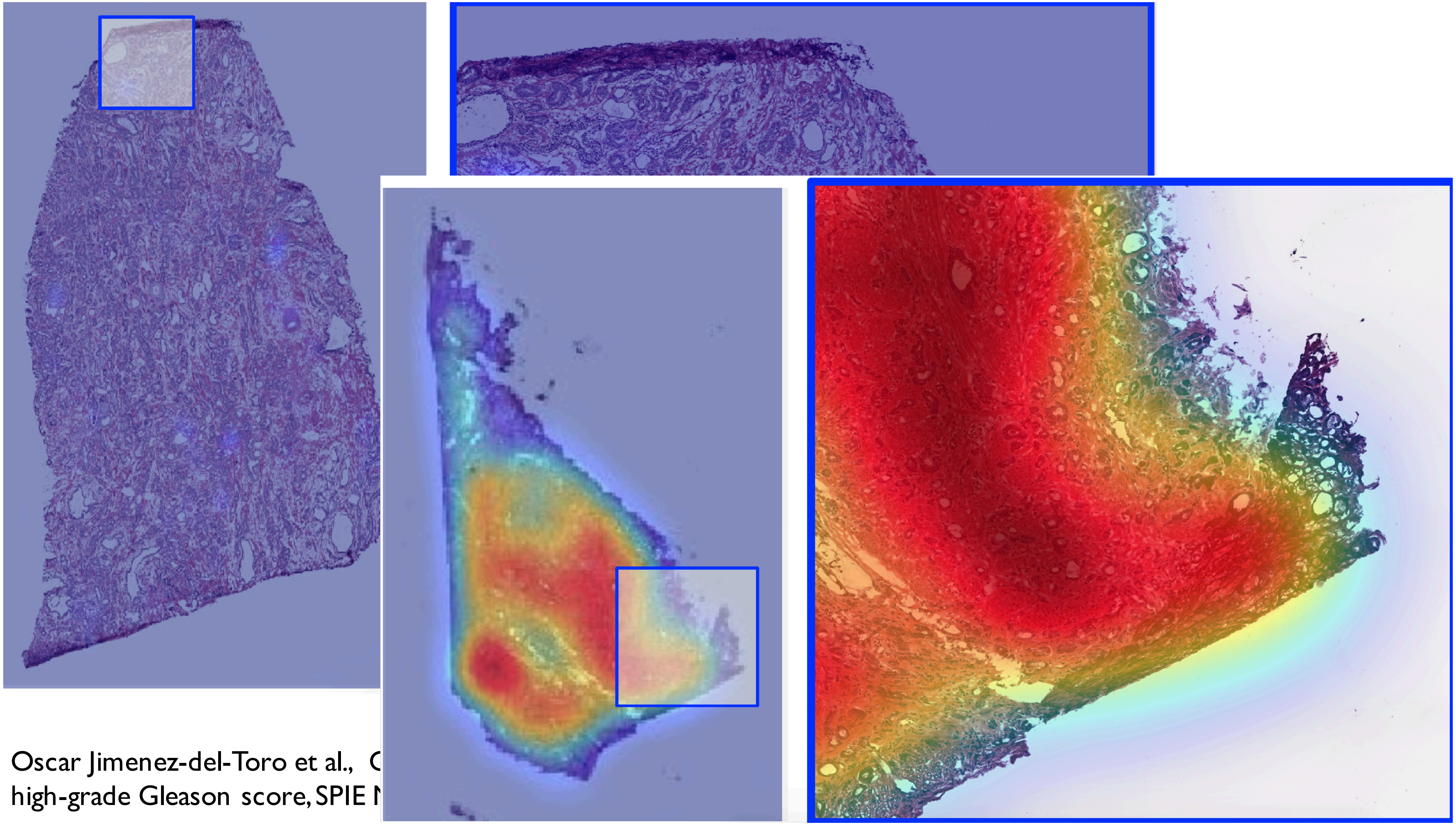


Specific other problems

Large parts without information

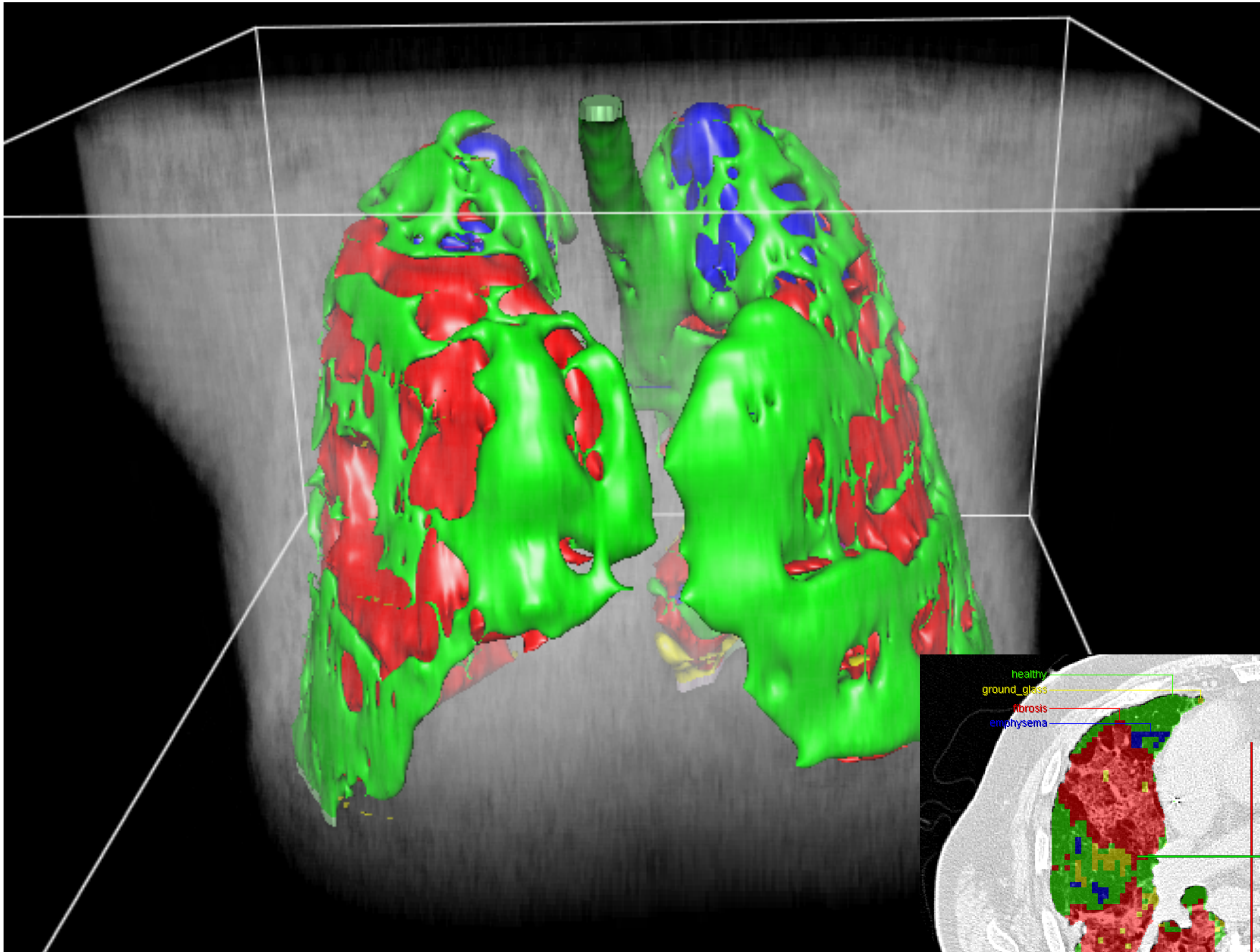
Examples: ROI detection (DL)

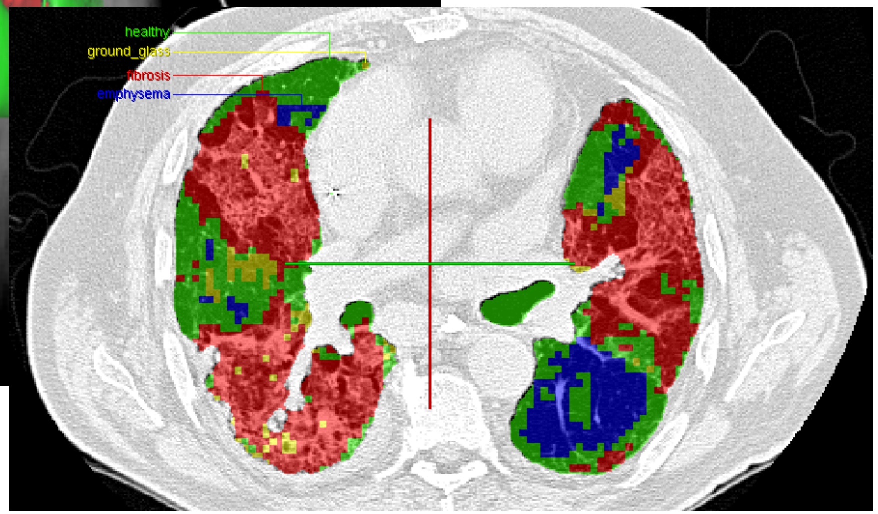
- Find areas with high vs. low Gleason grades



Oscar Jimenez-del-Toro et al., C
high-grade Gleason score, SPIE N

Example: 3D ROIs in lung CT





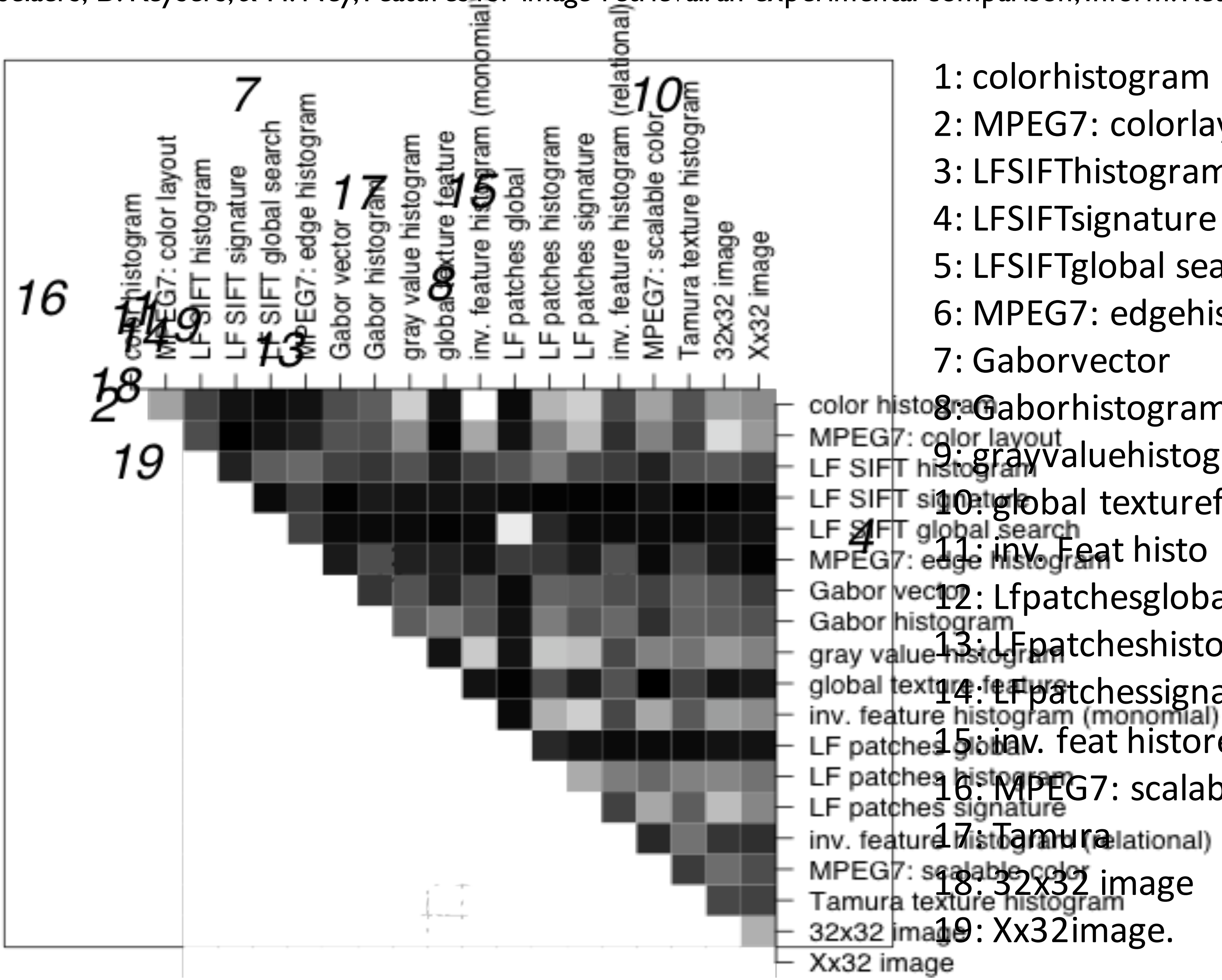
Types of visual features

- 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

The semantic gap …

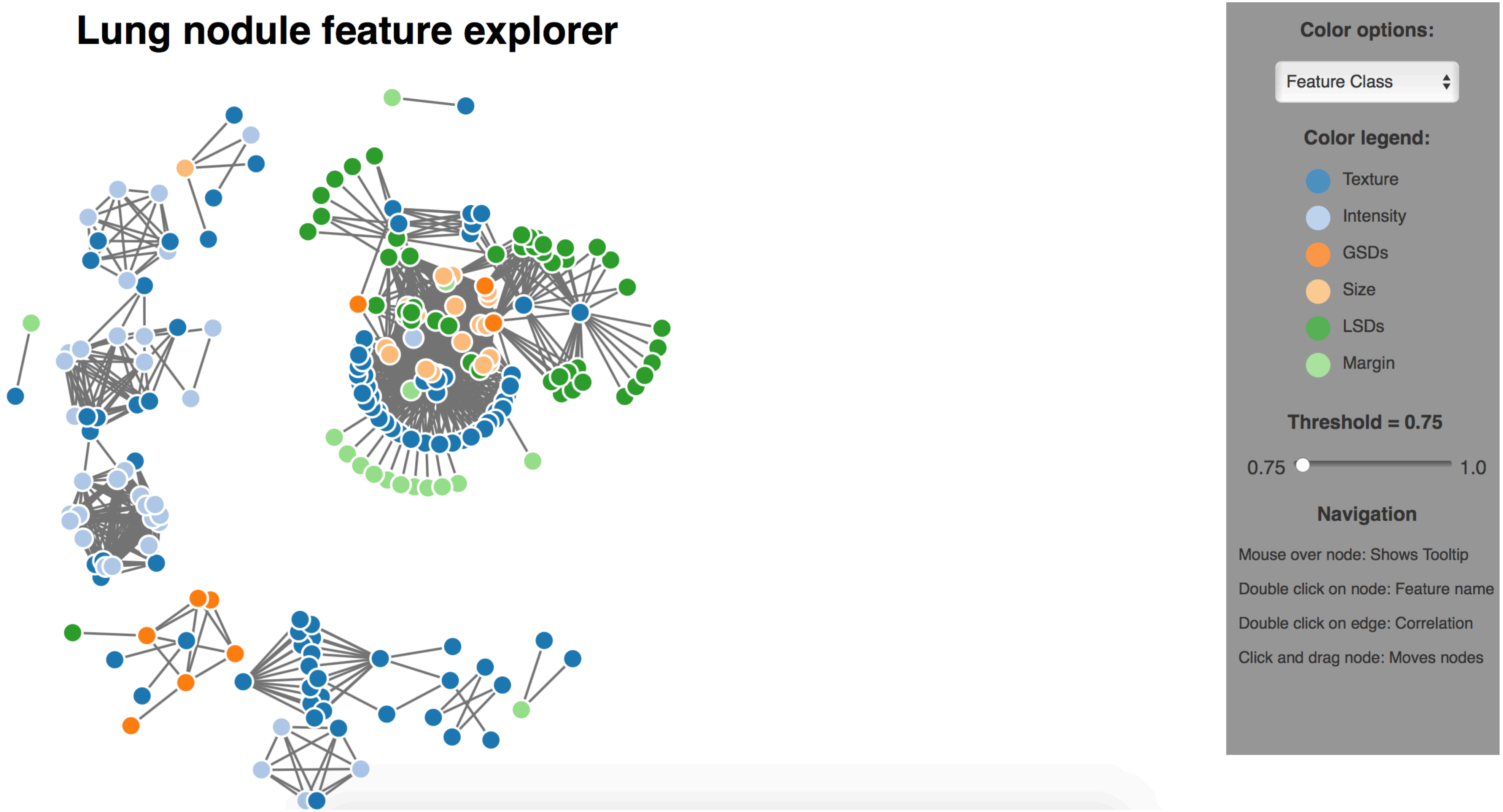
Correlation between Features

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

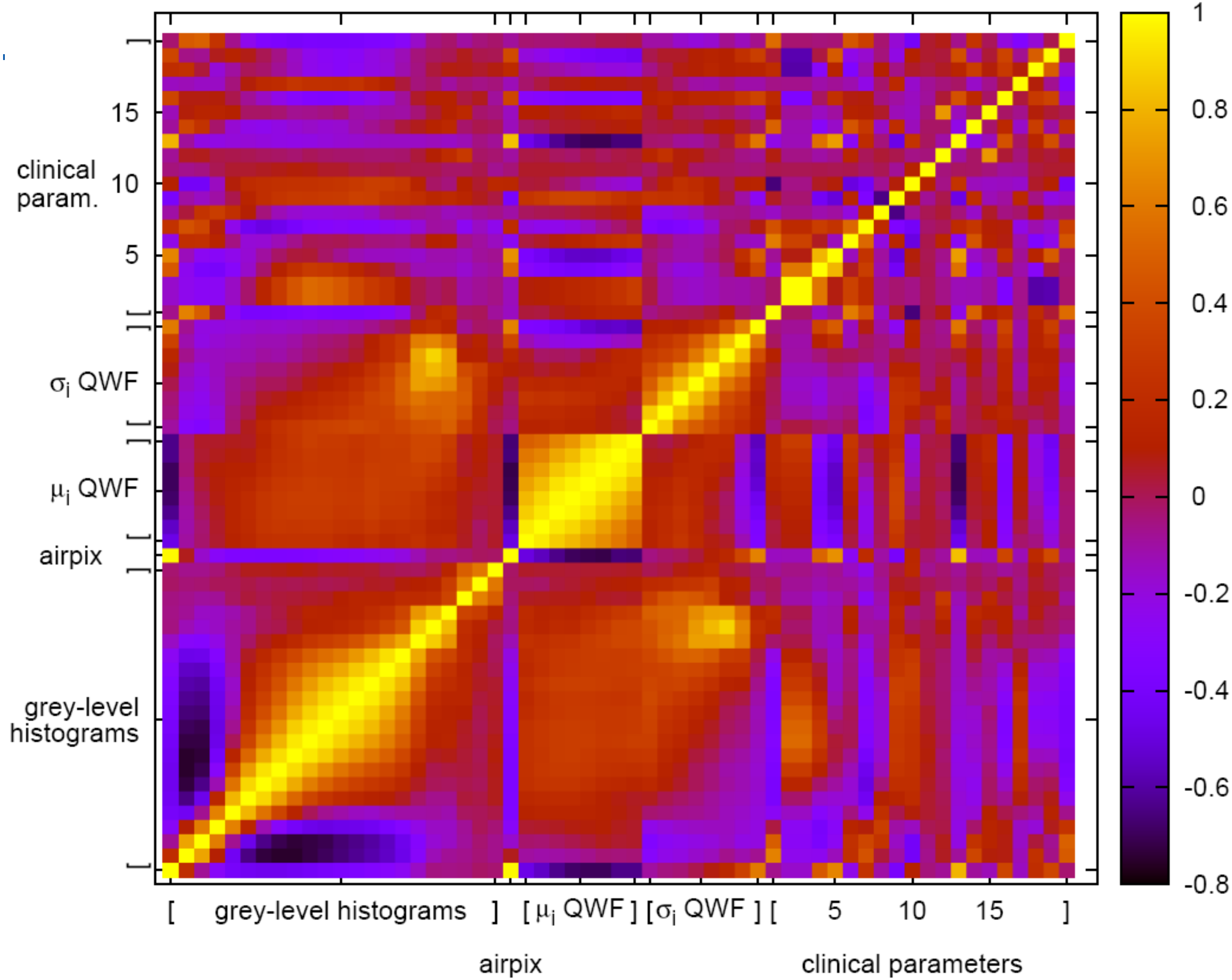


1: colorhistogram
2: MPEG7: colorlayout
3: LFSIFThistogram
4: LFSIFTsignature
5: LFSIFTglobal search
6: MPEG7: edgehistogram,
7: Gaborvector
8: Gaborhistograms
9: grayvaluehistogram
10: global texturefeature,
11: inv. Feat histo color
12: Lfpatchesglobal
13: LFpatcheshistogram
14: LFpatchessignature,
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



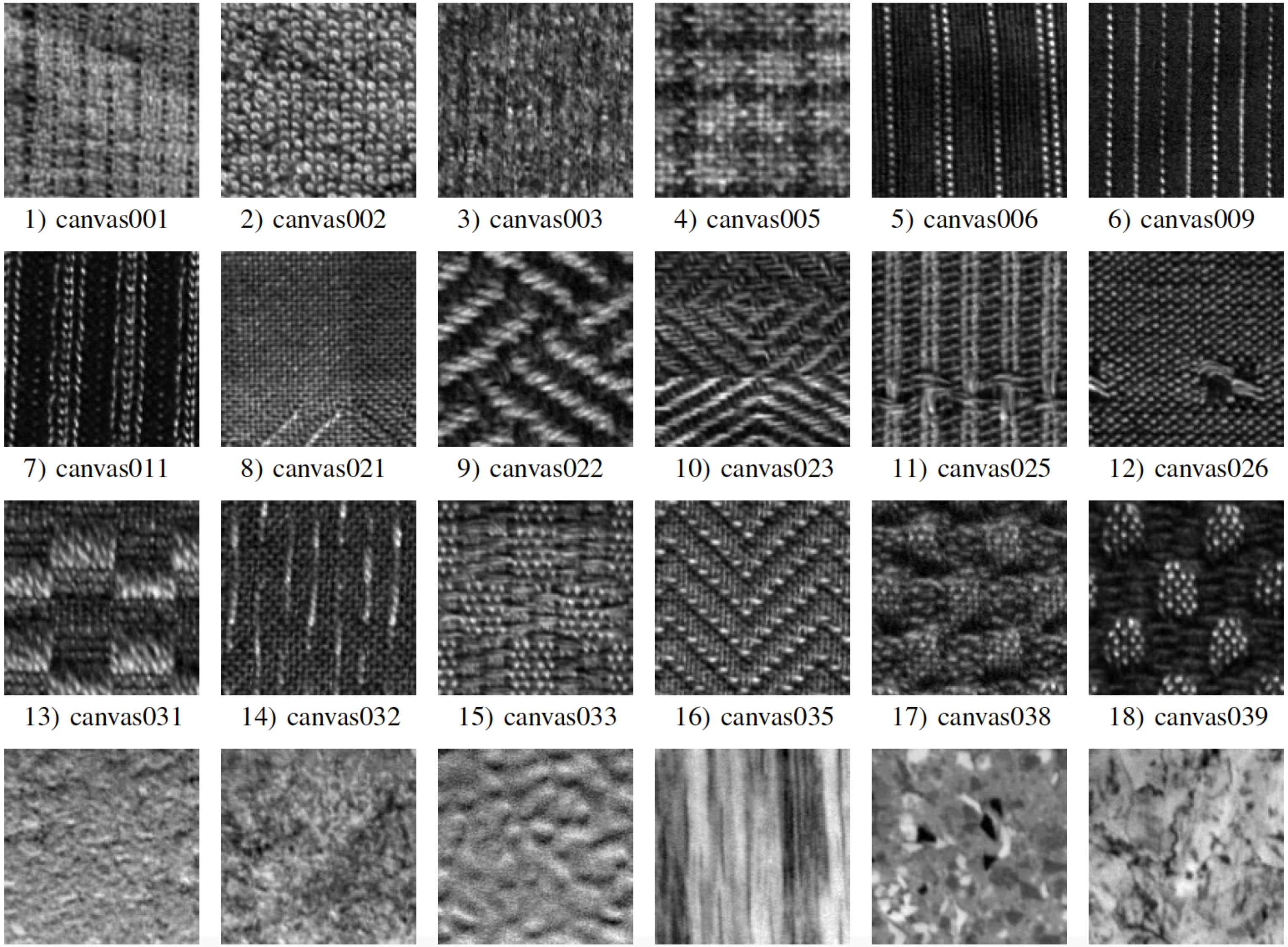
clinical param.
15
10
5
σi QWF
μi QWF
airpix
grey-level histograms
[grey-level histograms] [μi QWF] [σi QWF] [5 10 15]
airpix
clinical parameters
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8

Color/grey level features

- Resolution in the feature space is important for comparisons (particularly using histograms)
  - Level/window settings to concentrate on the interesting aspects (for example in the case of CT)
- Different color spaces (RGB, HSV, CIE Lab/Luv) that correspond more or less to human perception
- Global counting or relationships of local colors
  - Often using a simple histogram intersection

Texture

- "the feel, appearance, or consistency of a surface or a substance"



1) canvas001 2) canvas002 3) canvas003 4) canvas005 5) canvas006 6) canvas009
7) canvas011 8) canvas021 9) canvas022 10) canvas023 11) canvas025 12) canvas026
13) canvas031 14) canvas032 15) canvas033 16) canvas035 17) canvas038 18) canvas039
19) tile005 20) tile006 21) carpet002 22) carpet004 23) carpet005 24) carpet009

Texture features

- Tamura texture features (1978)
    - Coarseness – coarse vs. fine
    - Contrast – high vs. low
    - Directionality – directional vs. non-directional
    - Linelikeness – line-like vs. non-line-like
    - Regularity – regular vs. irregular
    - Roughness – rough vs. smooth
- Grey level co-occurrence matrices (Haralick, 1979)
    - Choice of scales, directions
- Gabor filters, Wavelets
    - Steerable wavelets allow for texture learning (Riesz)
- Local Binary Patterns (LBP)

Gray-Level Co-Occurrence Matrices

- Statistical descriptor for texture properties of an image by comparing neighboring pixels
  - Direction and distance
    - Features extracted from several matrices in general
  - Extract features from matrix
    - Entropy
    - Contrast
    - Correlation
    - …

I

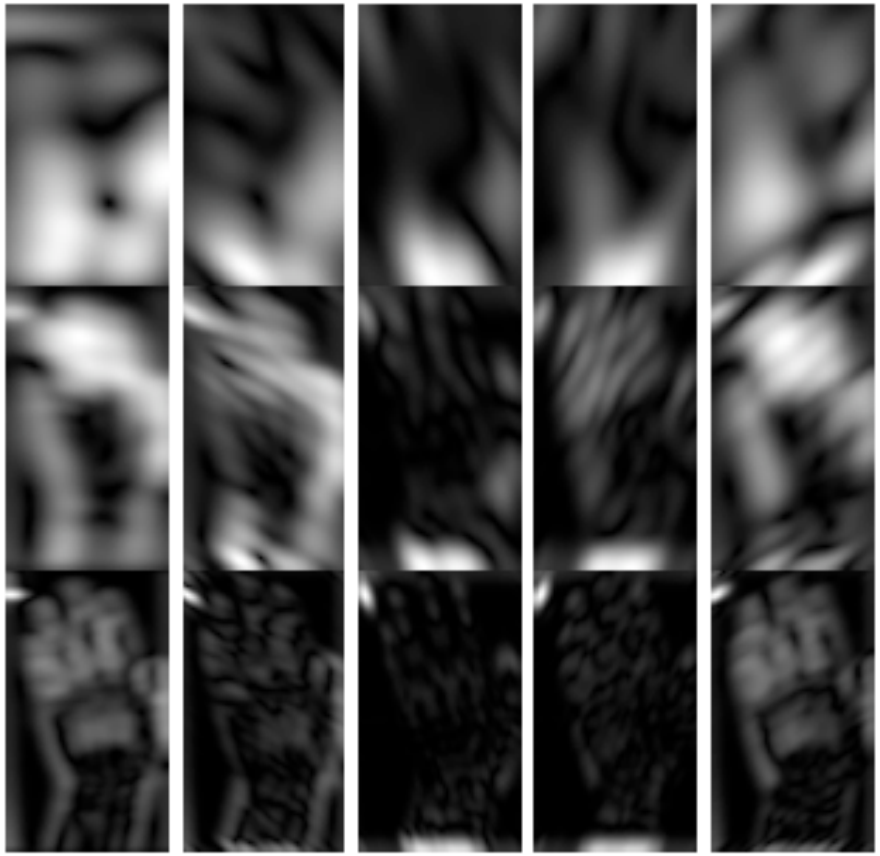
1	1	5	6	8
2	3	5	7	1
4	5	7	1	2
8	5	1	2	5

GLCM

	1	2	3	4	5	6	7	8
1	1	2	0	0	1	0	0	0
2	0	0	1	0	1	0	0	0
3	0	0	0	0	1	0	0	0
4	0	0	0	0	1	0	0	0
5	1	0	0	0	0	1	2	0
6	0	0	0	0	0	0	0	1
7	2	0	0	0	0	0	0	0
8	0	0	0	0	1	0	0	0

Gabor Features

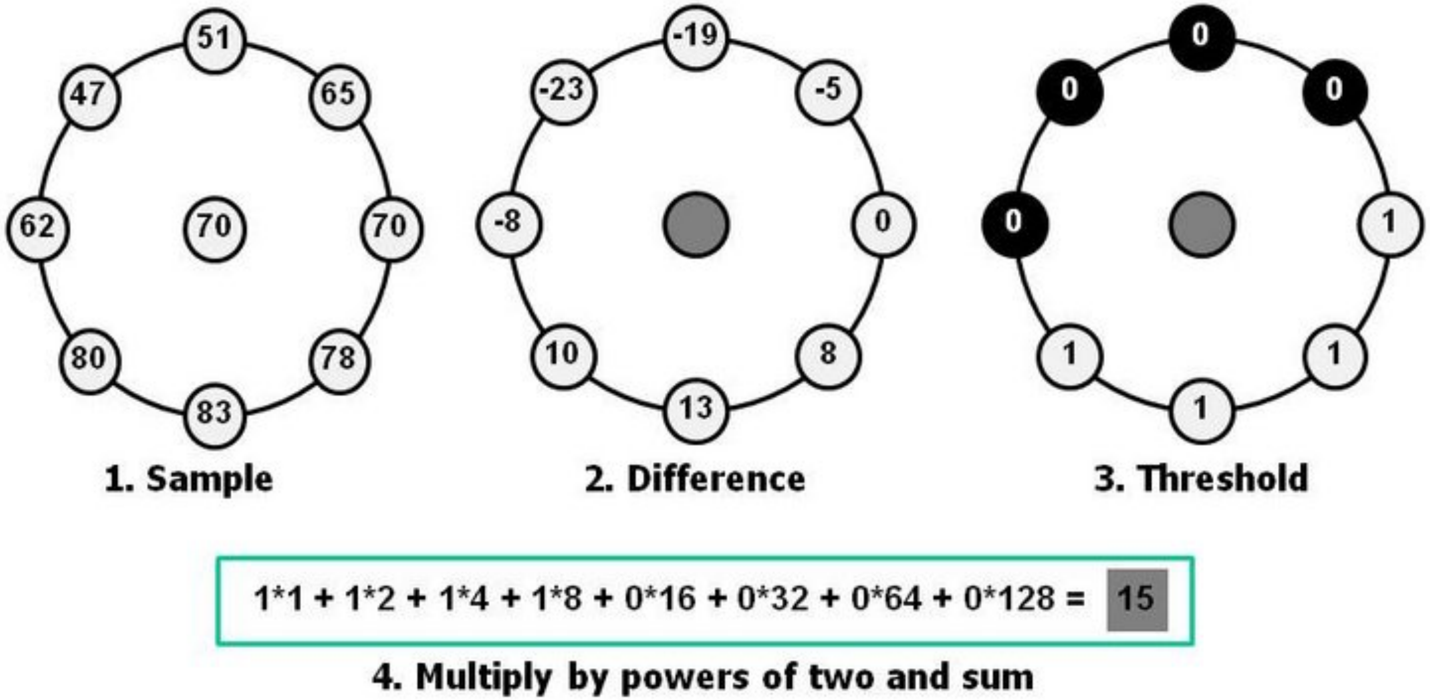
- Obtain several values per pixel denoting spatial frequencies and directions
  - Often four directions, three scales, similar to GLCM



Local binary patterns

The value of the LBP code of a pixel (x_c, y_c) is given by:

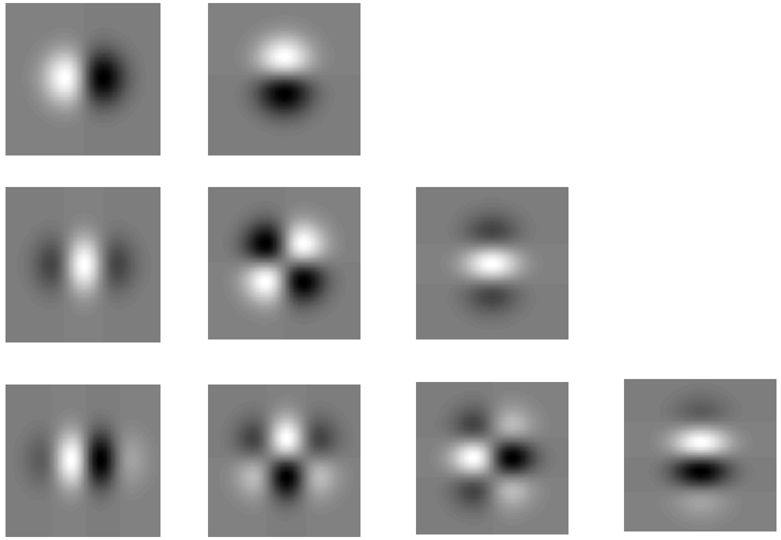
$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c)2^p \qquad s(x) = \begin{cases} 1, if\ x \geq 0; \\ 0, otherwise. \end{cases}$$



Ojala, T., Pietikäinen, M. and Harwood, D. (1996), A Comparative Study of Texture Measures with Classification Based on Feature Distributions. Pattern Recognition 19(3):51-59.

The Riesz transform

- The Riesz transform implements Nth-order directional derivatives:



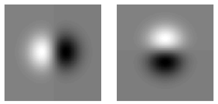
$N = 1$

$N = 2$

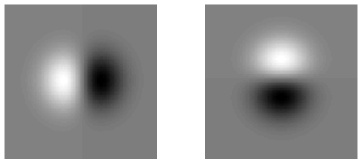
$N = 3$

- At multiple scales:

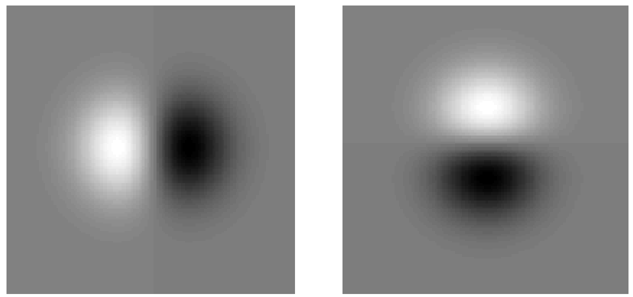
$N = 1$, 1st scale:



$N = 1$, 2nd scale:



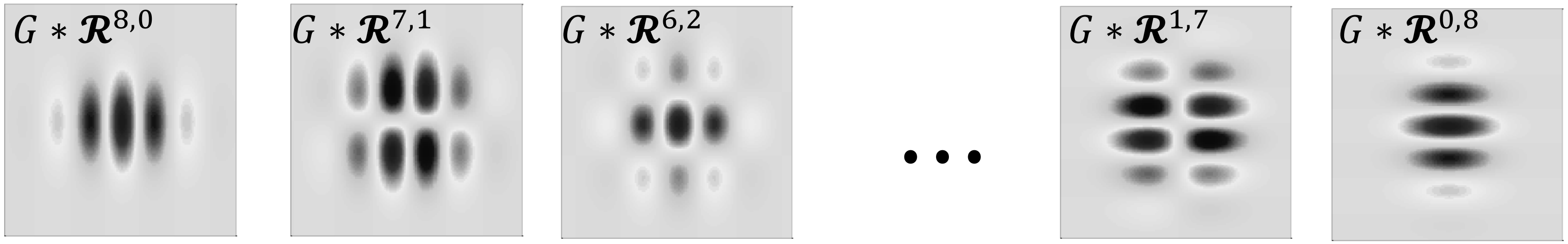
$N = 1$, 3rd scale:



The Riesz transform:

- A Riesz filterbank constitutes a dictionary of basic textures:

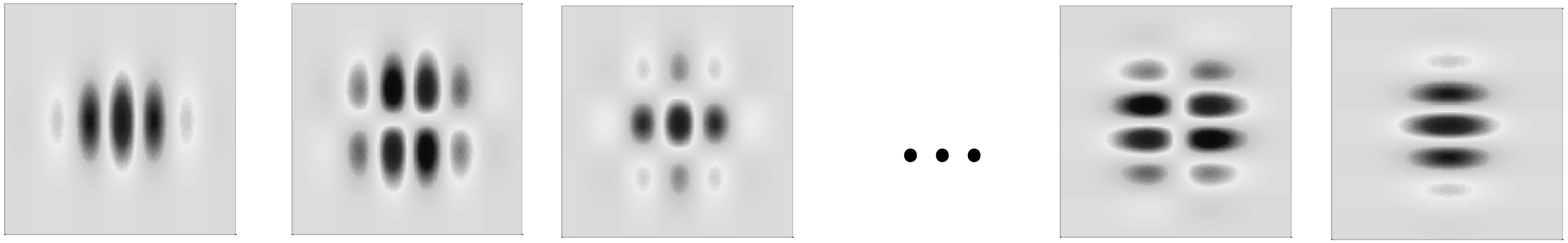
Riesz filterbank ($N = 8$)



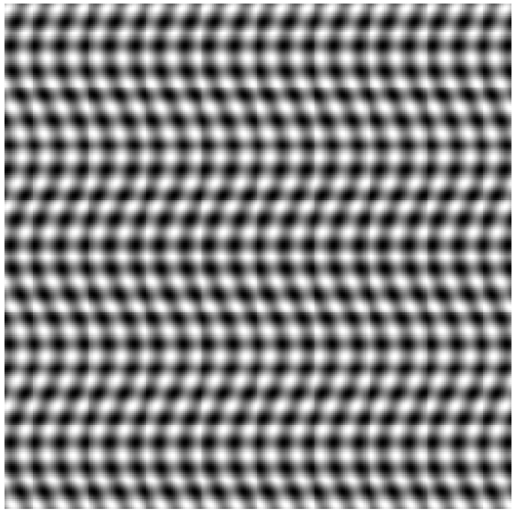
- Texture signatures are built using a linear combination of the Riesz templates

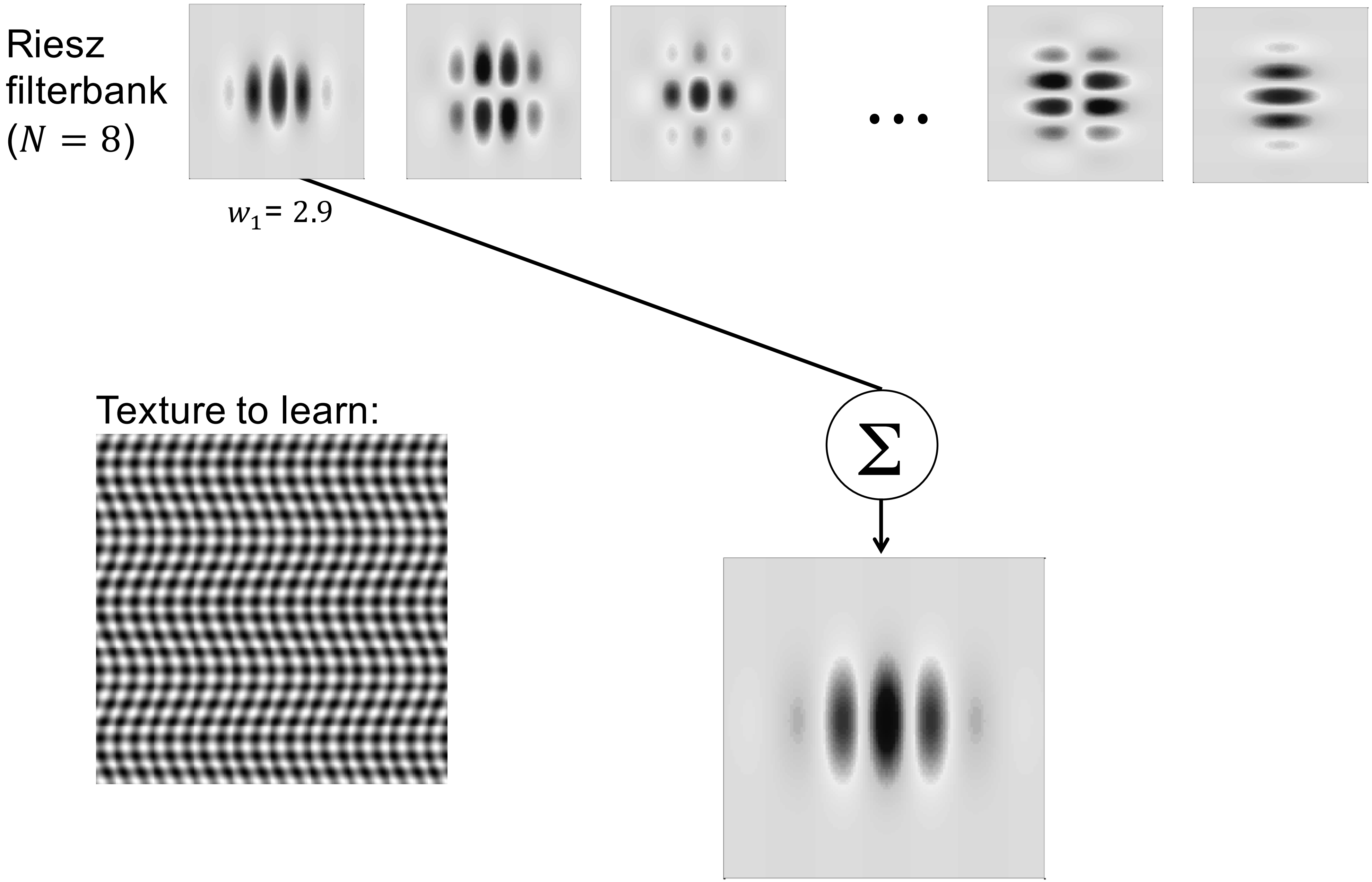
$$\Gamma_c{}^N = w_1(G * \boldsymbol{\mathcal{R}}^{N,0})_{s_1} + w_2(G * \boldsymbol{\mathcal{R}}^{N-1,1})_{s_1} + \cdots + w_{4N+4}(G * \boldsymbol{\mathcal{R}}^{0,N})_{s_4}$$

Riesz filterbank $(N = 8)$

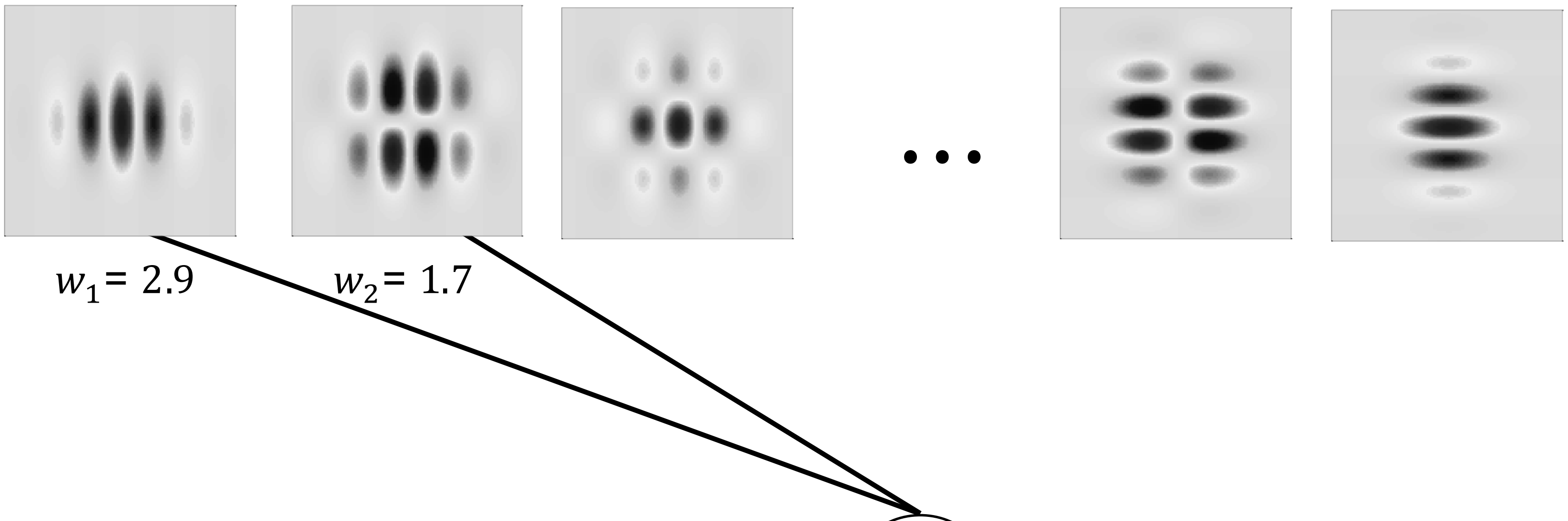


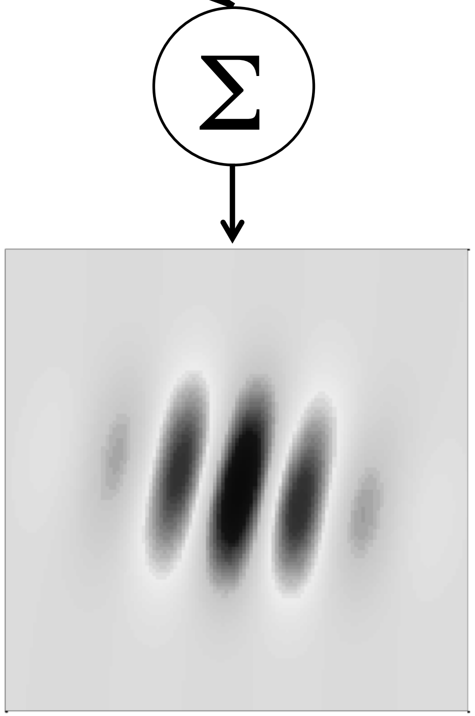
Texture to learn:



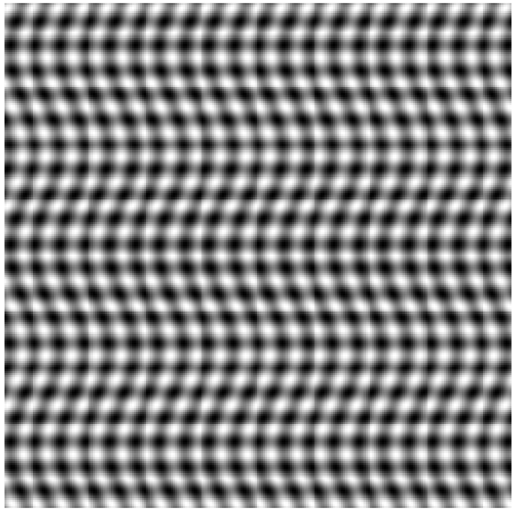

Riesz
filterbank
$(N = 8)$
$w_1 = 2.9$
Texture to learn:
Σ

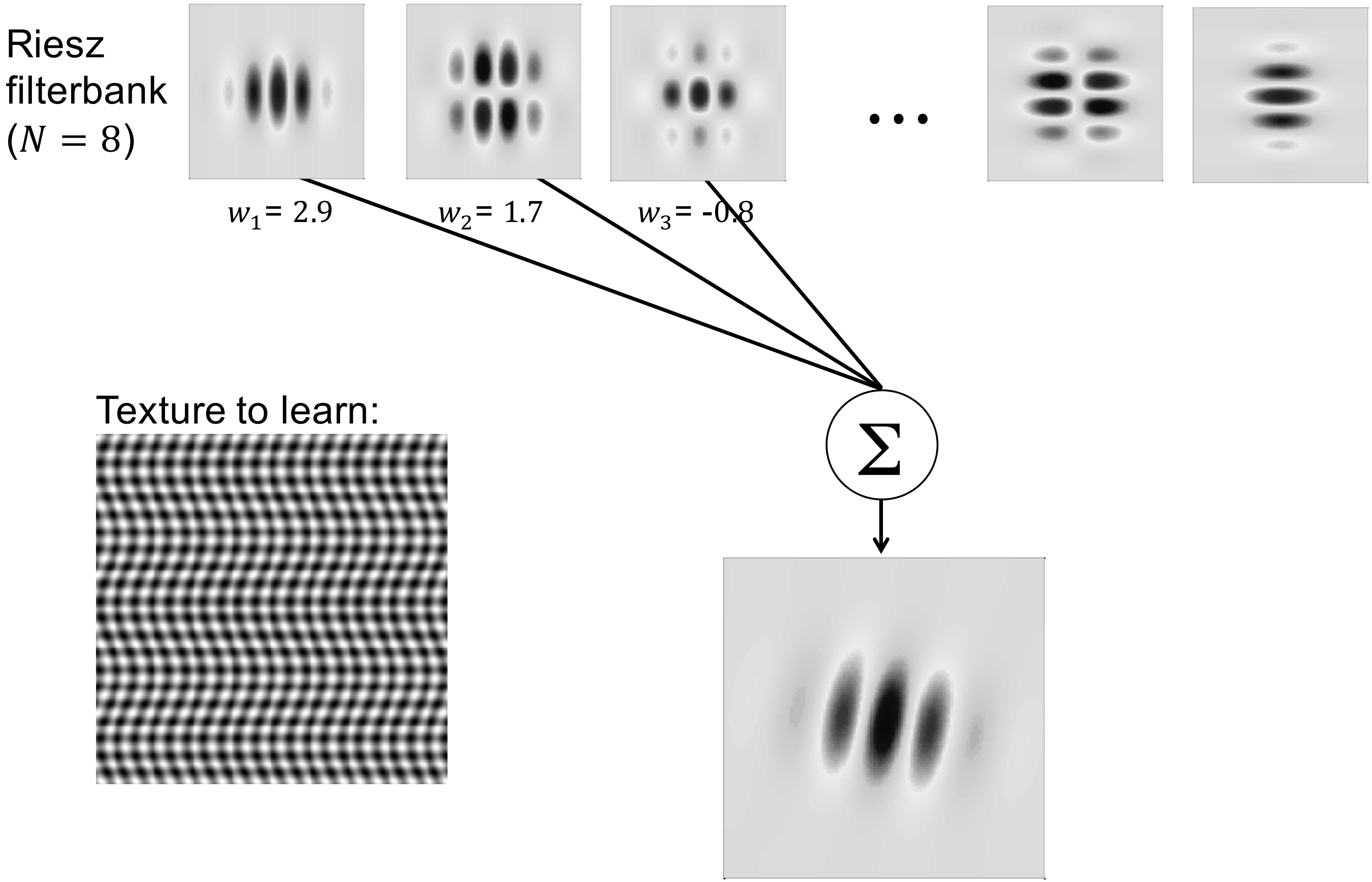
Riesz filterbank ($N = 8$)



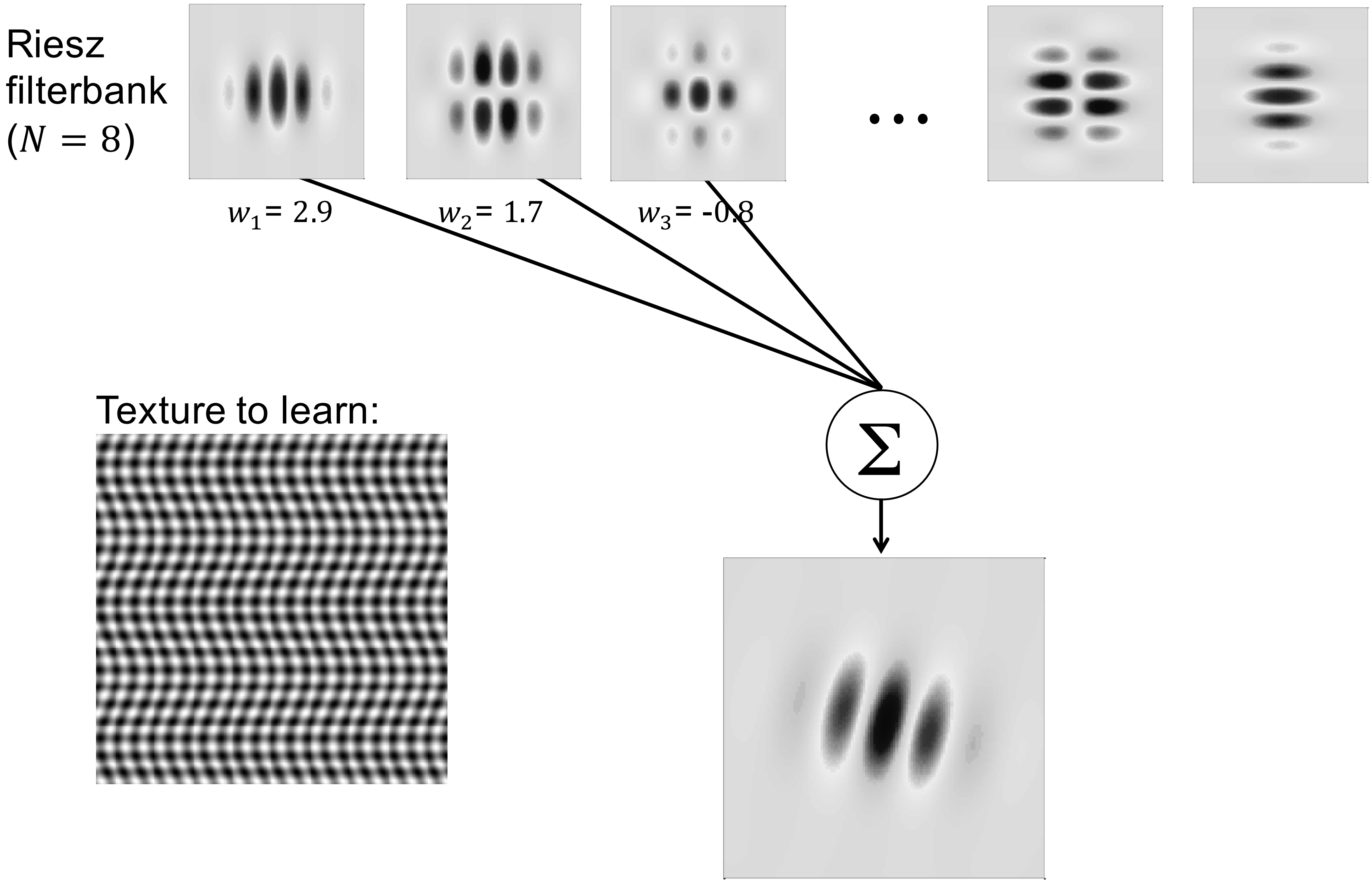


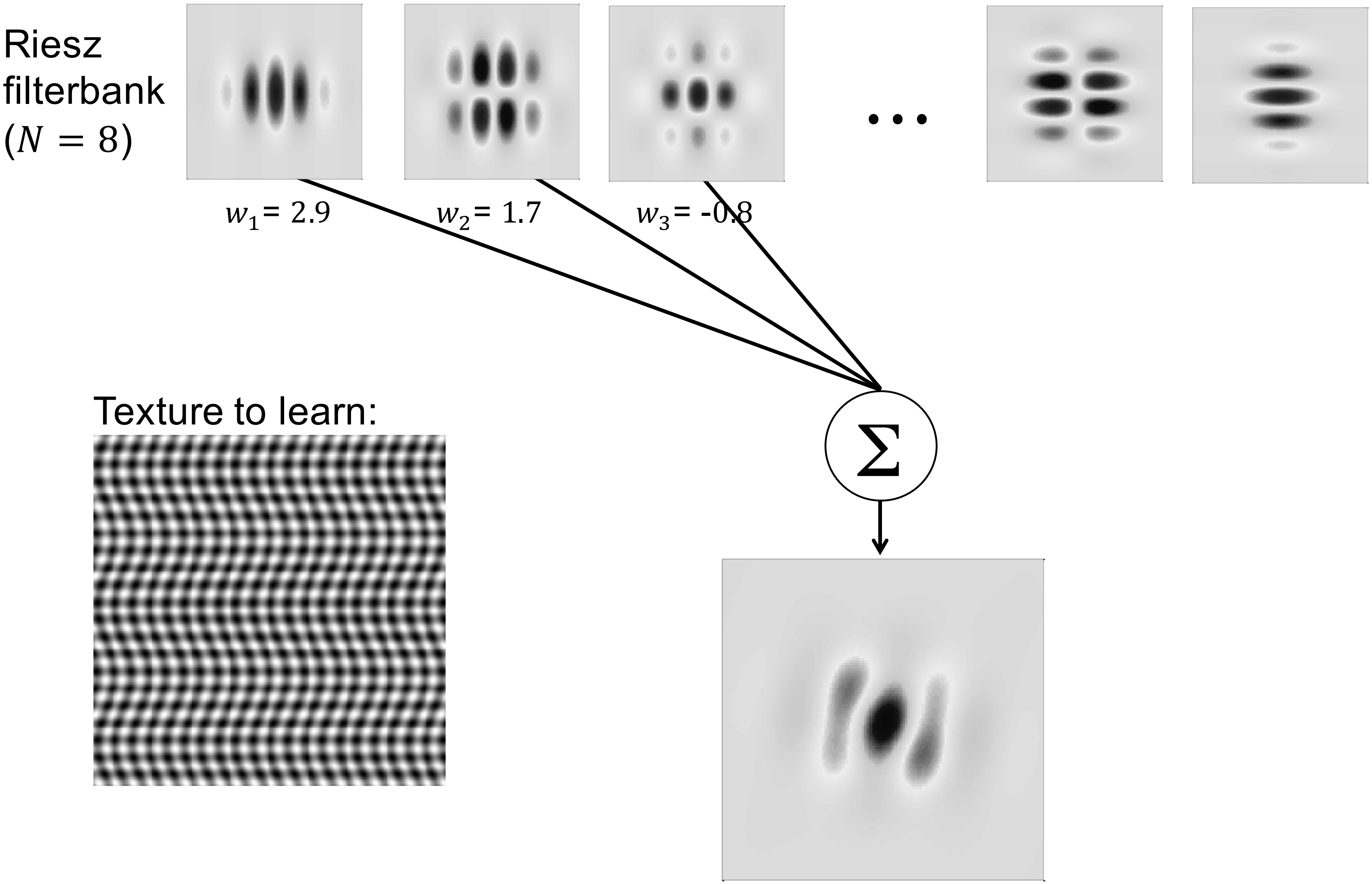
Texture to learn:

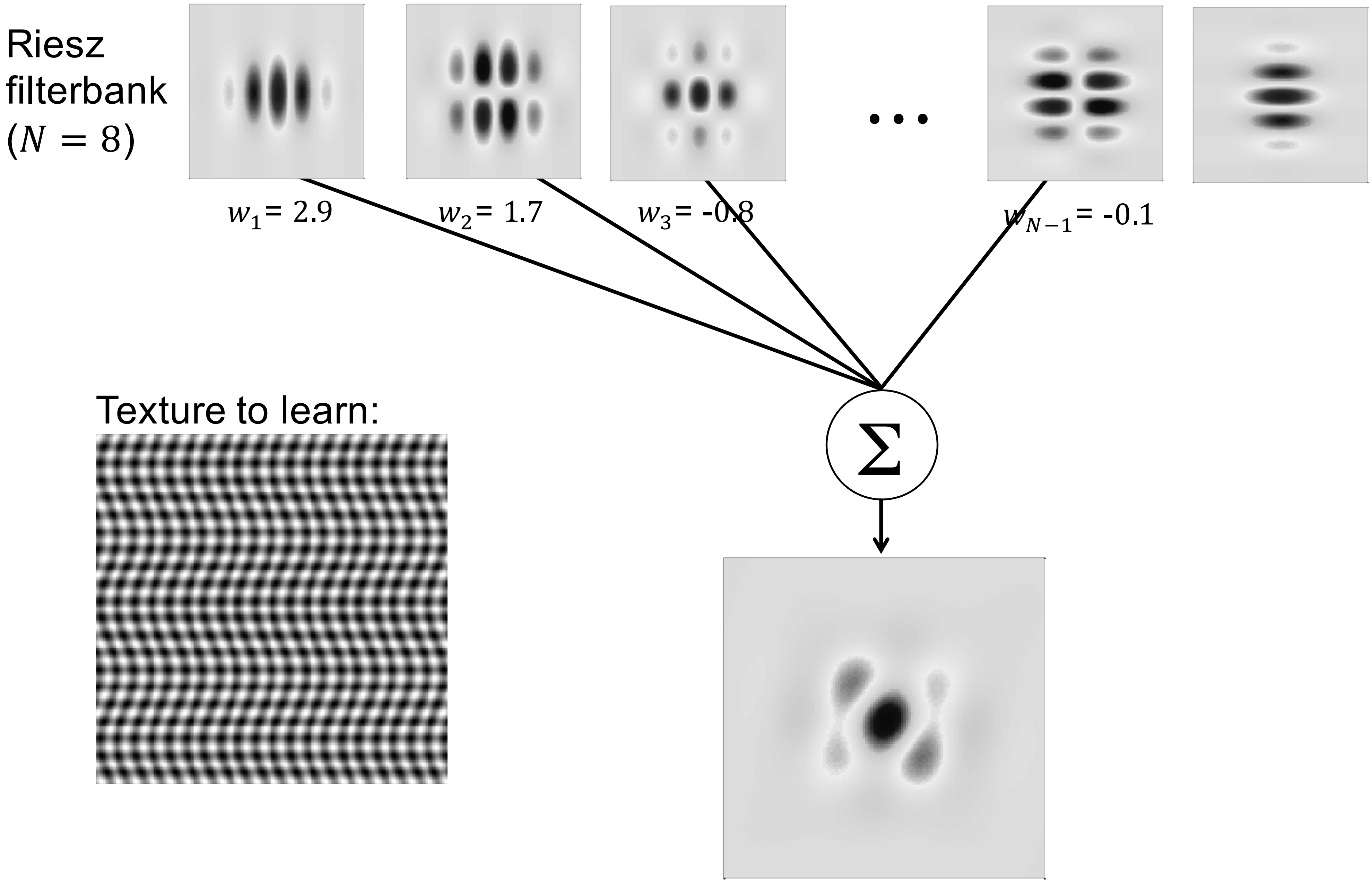




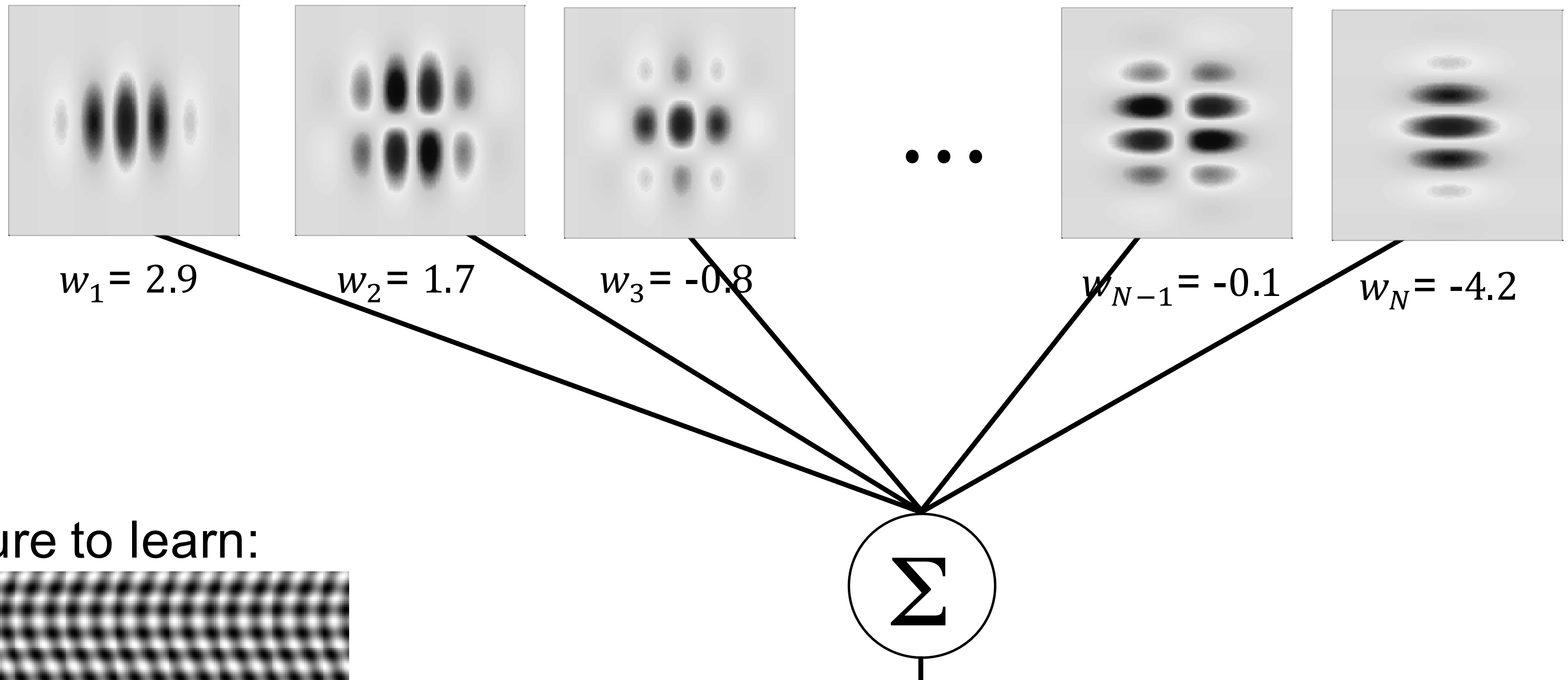
Riesz
filterbank
$(N = 8)$
$w_1 = 2.9$
$w_2 = 1.7$
$w_3 = -0.8$
. . .
Σ
Texture to learn:

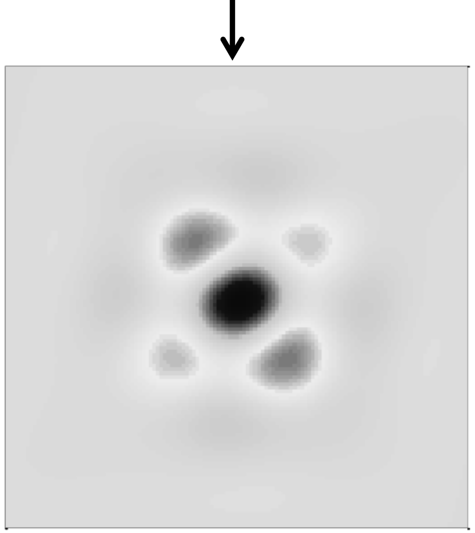

Riesz filterbank ($N = 8$)
w_1= 2.9
w_2= 1.7
w_3= -0.8
...
Σ
Texture to learn:


Riesz filterbank $(N = 8)$
$w_1 = 2.9$
$w_2 = 1.7$
$w_3 = -0.8$
…
Σ
Texture to learn:

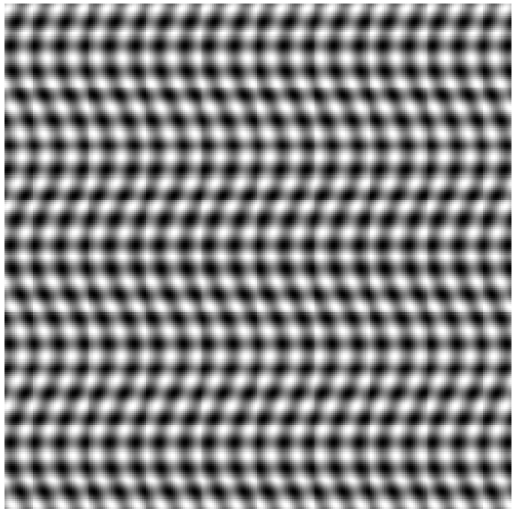

Riesz filterbank ($N = 8$)
w_1= 2.9
w_2= 1.7
w_3= -0.8
w_{N-1}= -0.1
Texture to learn:
Σ

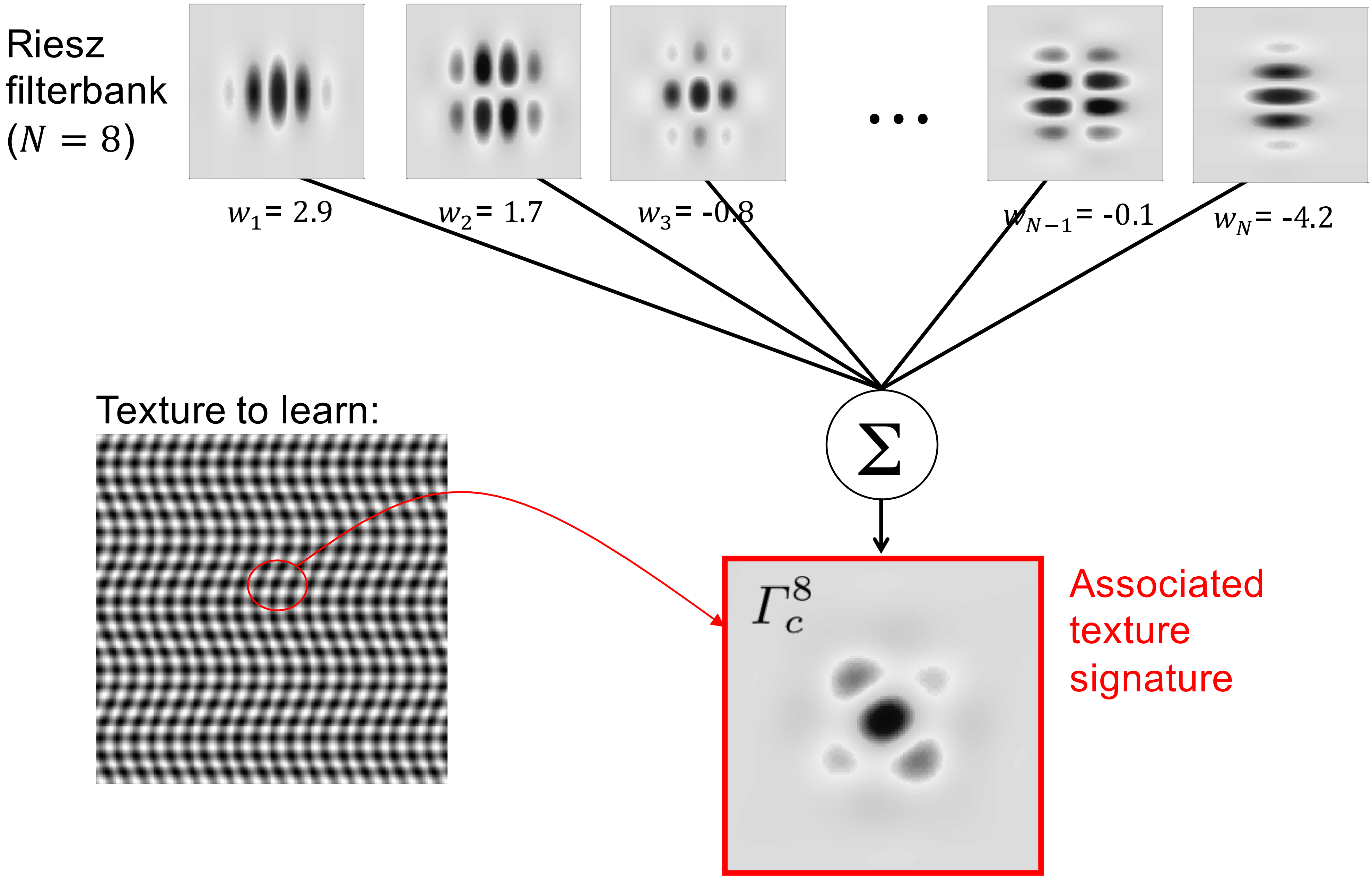
Riesz filterbank ($N = 8$)





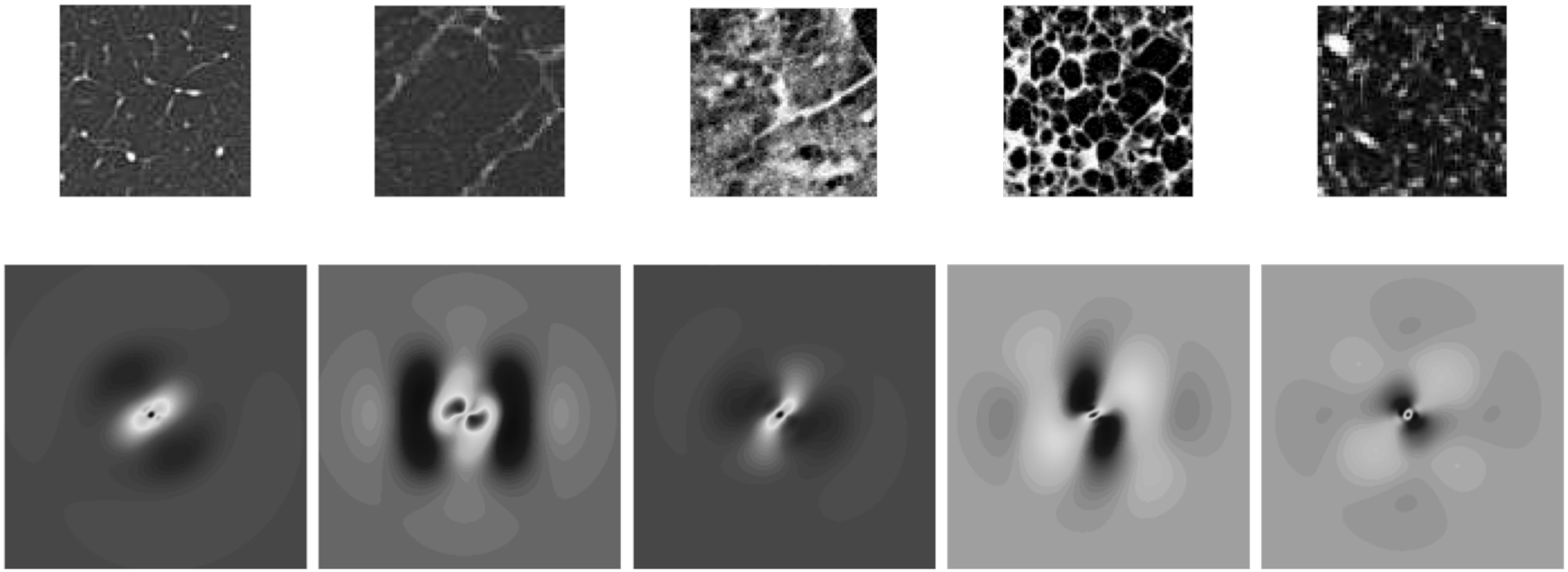
Texture to learn:




Riesz filterbank ($N = 8$)
w_1= 2.9
w_2= 1.7
w_3= -0.8
w_{N-1}= -0.1
w_N= -4.2
Σ
Texture to learn:
Γ_c^8
Associated texture signature

Learned 3D signatures

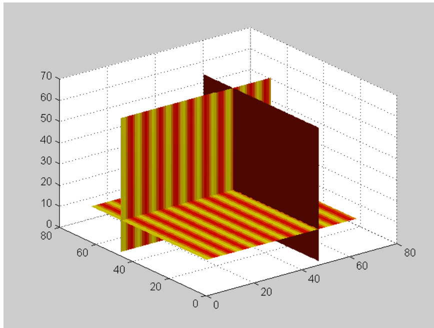
- Learn combinations of Riesz wavelets as digital signatures using SVMs
  - Create signatures to detect small local lesions and visualize them

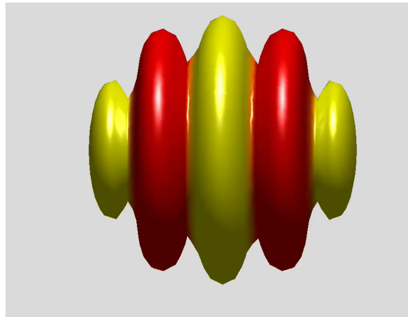


Adrien Depeursinge, Antonio Foncubierta–Rodriguez, Dimitri Van de Ville, and Henning Müller, Rotation–covariant feature learning using steerable Riesz wavelets, IEEE Transactions on Image Processing, volume 23, number 2, page 898-908, 2014.

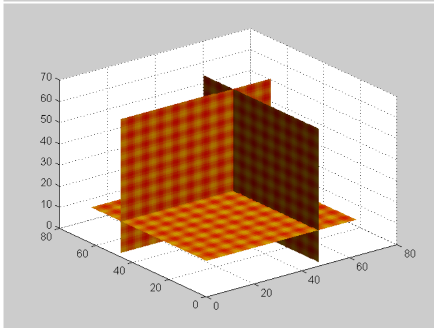
Learning Riesz in 3D

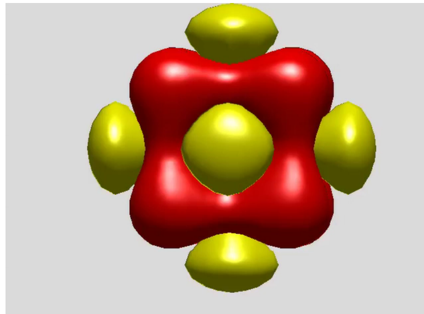
- Most medical tissues are naturally 3D
- But modeling gets more complex
  - Vertical planes



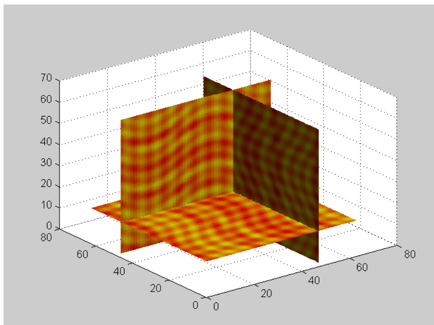


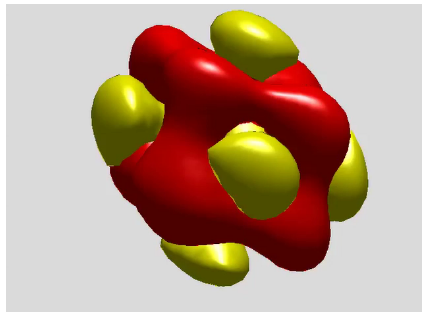
  - 3-D checkerboard





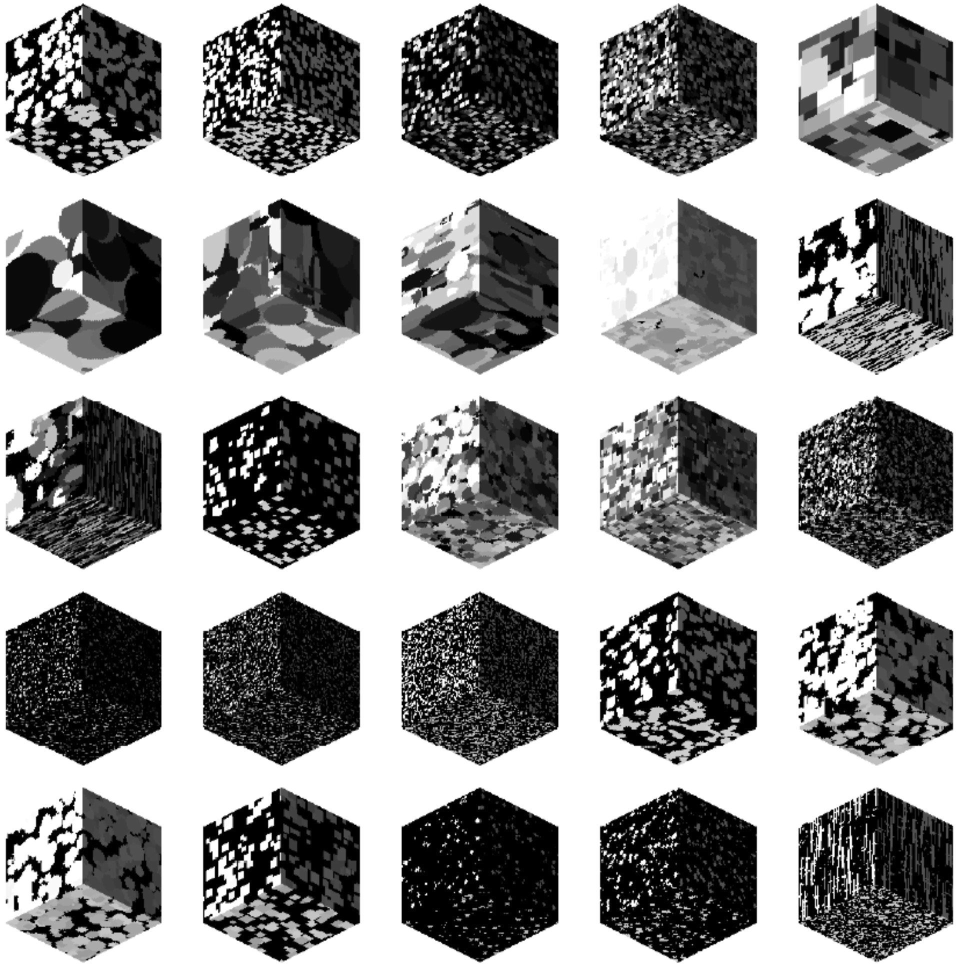
  - 3-D wiggled checkerboard

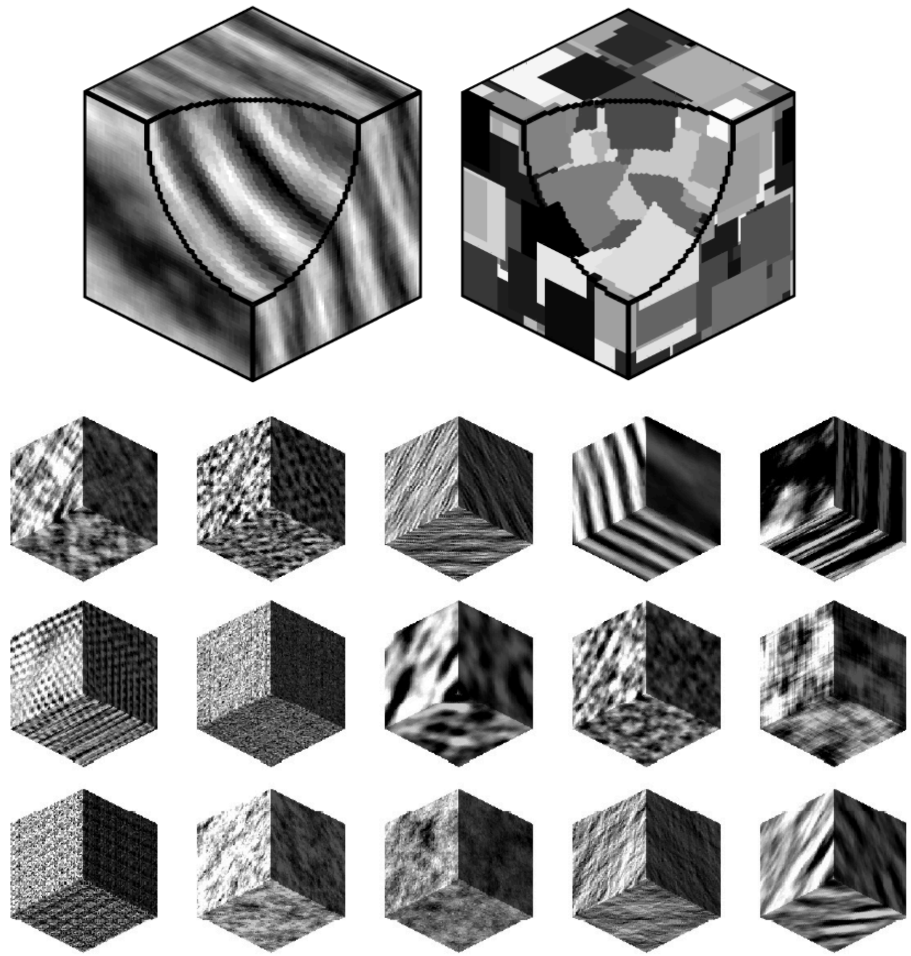




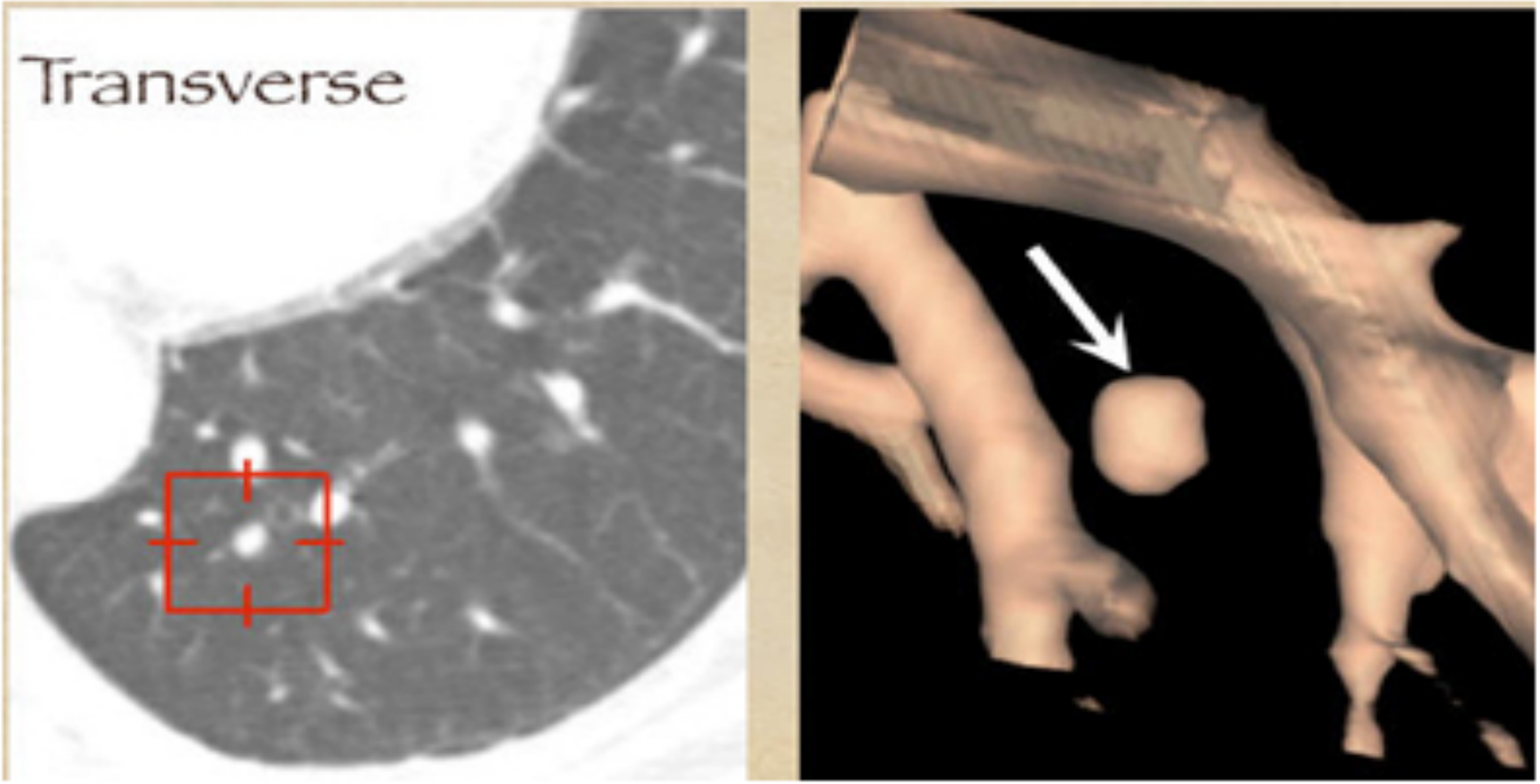
Solid 3D texture

- Hard to visualize
- Most texture measures are translated to 3D





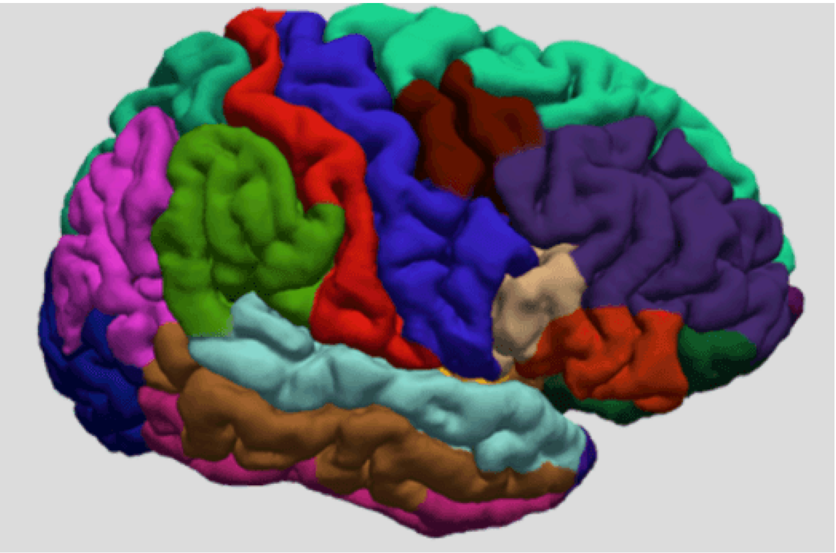
3D can be really important



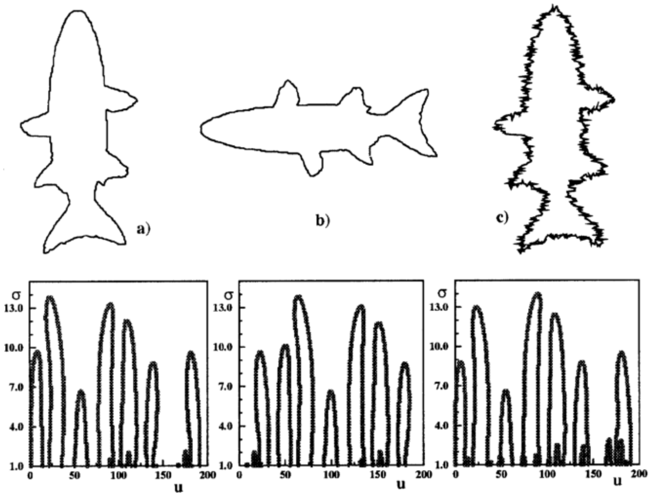
CT finding (left) has the appearance of an adjacent vessel in transverse-section reconstruction and was not called by any of the four LIDC readers. After viewing transverse, coronal, sagittal, and volume-rendered reconstructions (right), all four university readers called the finding a lung nodule.

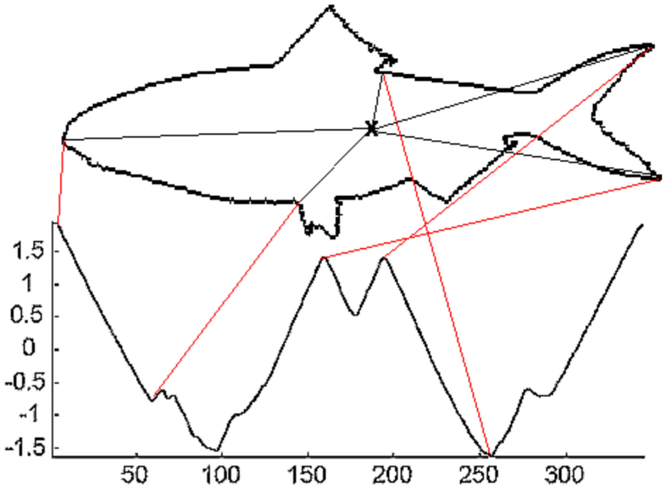
Interest Points – local information

Shape features



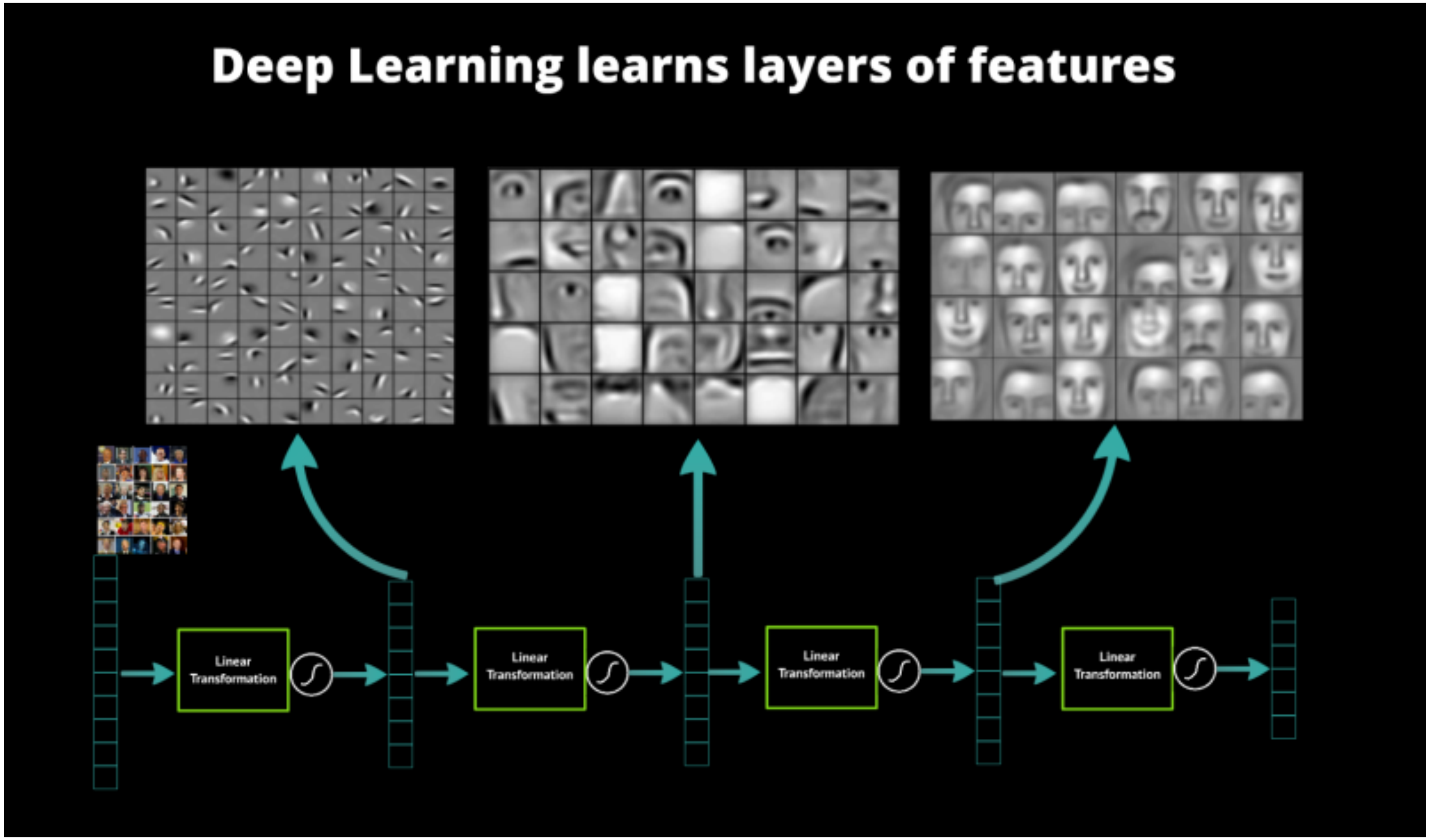
- 2D vs. 3D
- Invariances (rotations, size)
- Bounding box, aspect ration, circularity, centroid, chain code, …
- Invariant statistical moments (Zernicke moments)
- Curvature scale space





Deep learning-based features

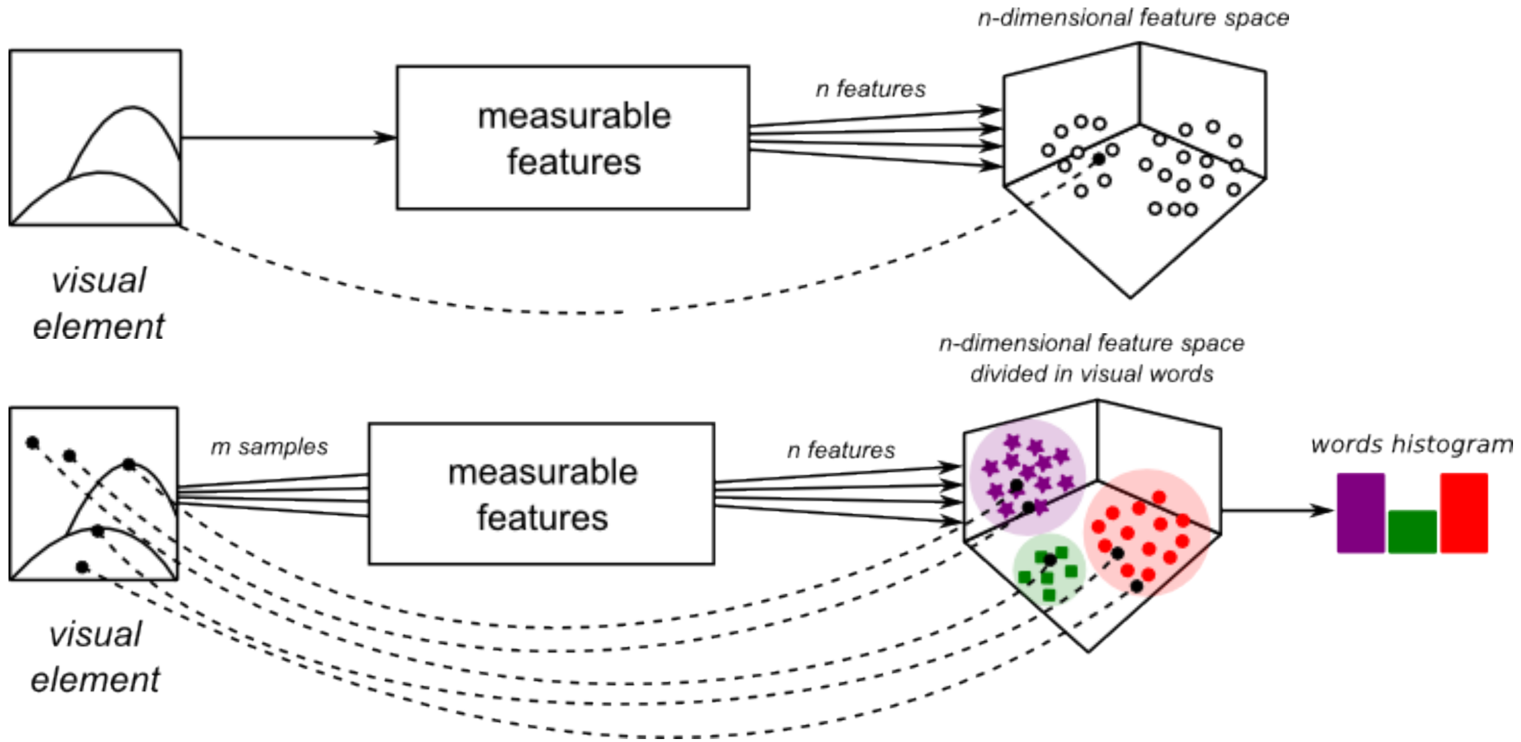
- Deep learning has several layers of neurons
- Usually “features” are not used and all is learned
  - Specific output layers can be used as features



Datarobot.com, a primer on deep learning

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

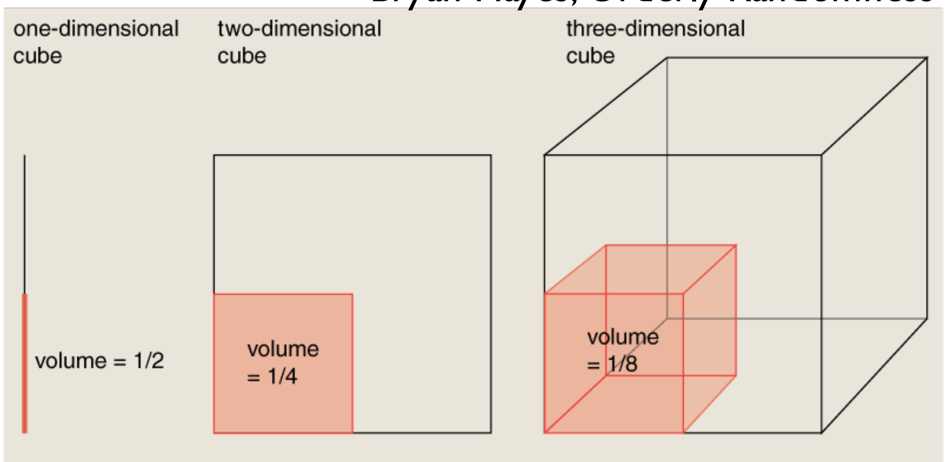


A Foncubierta, AG Seco de Herrera, H Müller, Medical Image Retrieval using a Bag of Meaningful Visual Words, ACM MM workshop on medical multimedia retrieval, Barcelona, Spain, 2013.

Curse of dimensionality

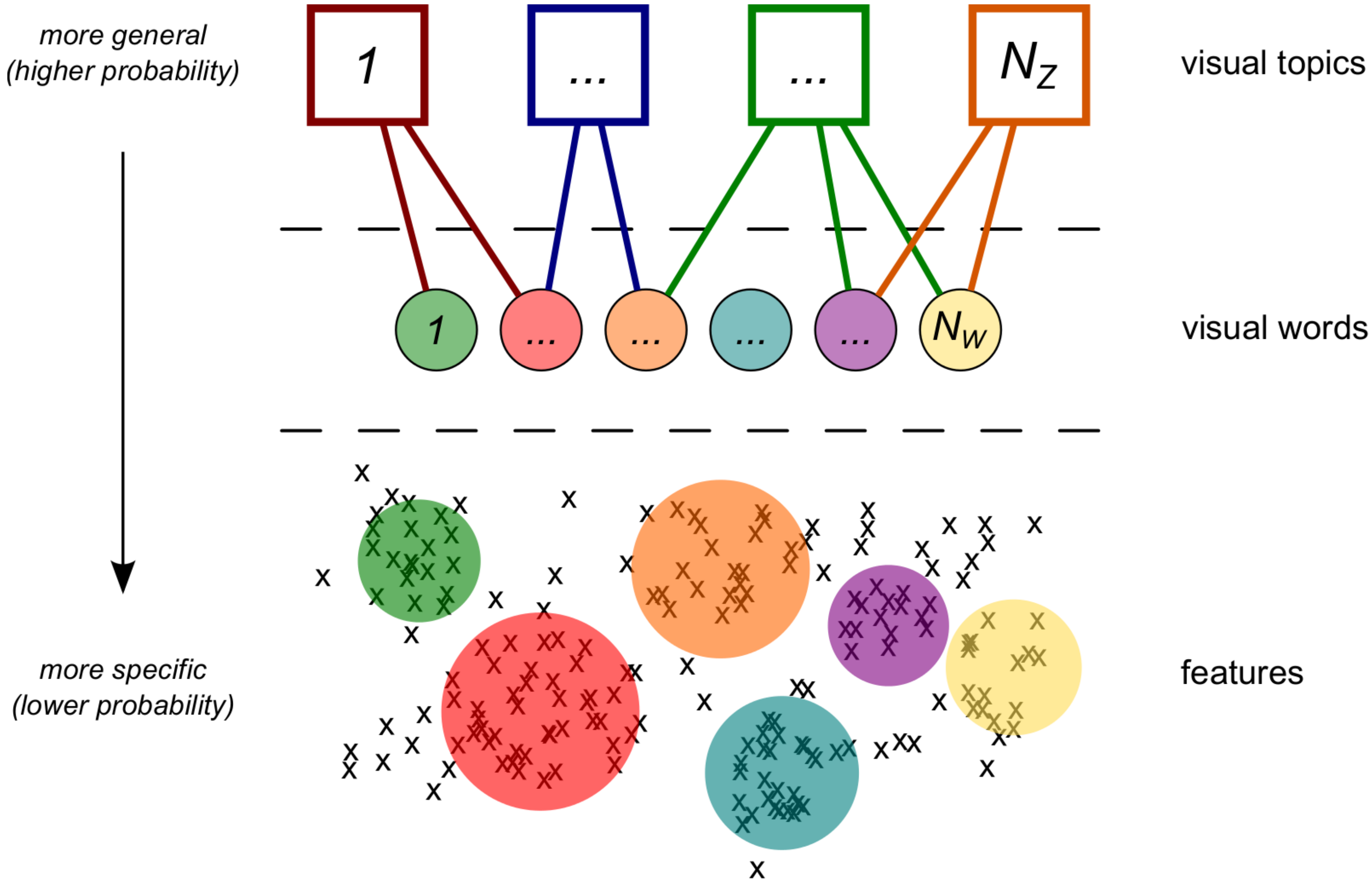
- "The curse of dimensionality refers to various phenomena that arise when analyzing and organizing data in high-dimensional spaces (often with hundreds or thousands of dimensions) that do not occur in low-dimensional settings such as the three-dimensional physical space of everyday experience." Wikipedia
- Increasing numbers of features mean that generalization requires exponential data amounts
- Volume of a space increase with more dimensions, so data gets increasingly sparse
  - Distance between all items becomes similar
  - Automatic classification is then hard (using kNN)

Bryan Hayes, Orderly Randomness



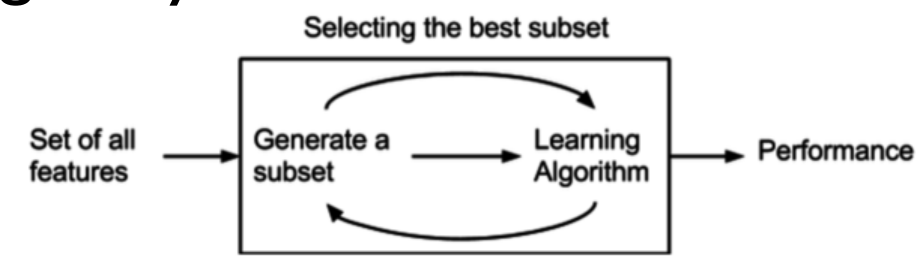
Latent semantic analysis

- Find underlying links between (visual) words
  - Find synonyms, homonyms, so ambiguous words



Feature selection strategies

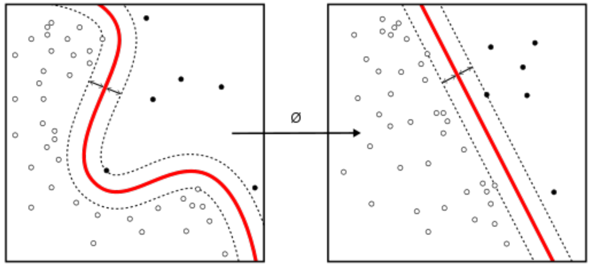
- Variable selection, attribute selection, …
- Goal: simpler models, lower storage, faster training, avoid overfitting, remove redundancy
- Many approaches
  - Exhaustive search, best first, greedy forward and backward, …
  - Filter, wrapper, embedded



- Criteria to select features
  - Correlation between features, mutual information, accuracy of the created models

Dimensionality reduction

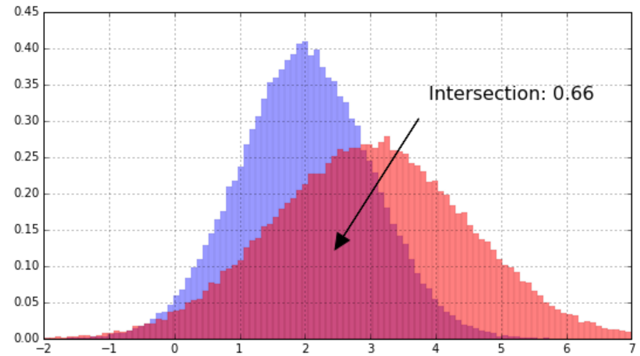
- Feature selection
- Principal component analysis (PCA)
  - Linear mapping of data onto fewer dimensions
  - Mapping to 2D, 3D allows to visualize data
- Kernel PCA
  - Nonlinear space, maximizing variance

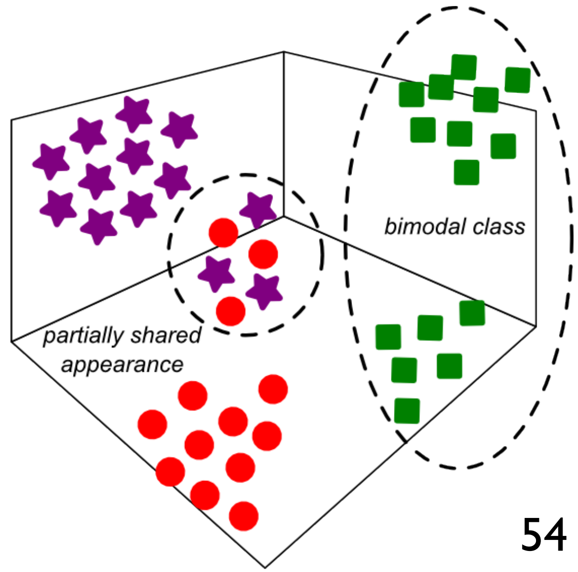


- Linear Discriminant Analysis (LDA)
  - Finding a linear combination to best separate classes
- …

Distance measures

- Visual features represent structures in an n-dimensional space
  - Hopefully our visual features separate the items well
- Many distance metrics exist
  - Histogram intersection
  - City block distance
  - Euclidean distance
  - Earth Movers distance
  - Mahalanobis, Bhattacharyya, …

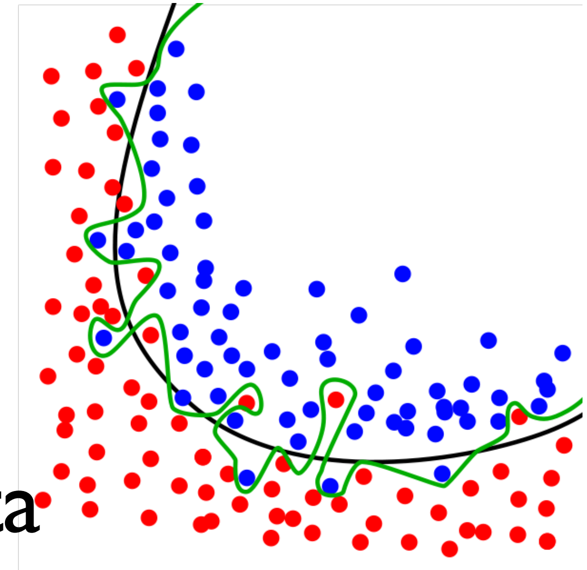




Overfitting

Hes·so VALAIS WALLIS
School of Management & Tourism

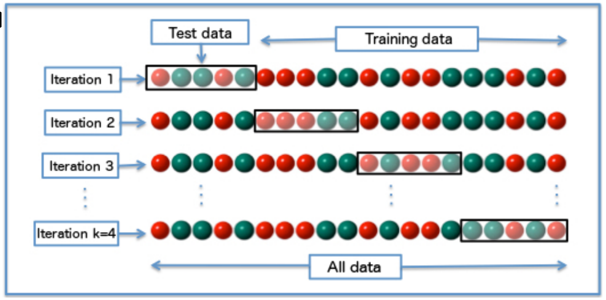
- "In overfitting, a statistical model describes random error or noise instead of the underlying relationship. Overfitting occurs when a model is excessively complex, such as having too many parameters relative to the number of observations." Wikipedia



- Over fit models do not generalize
  - Not good on new or unseen data
- Real risk in learning with many parameters or training on test data
  - Manual tuning to get good results
- A model should perform well on unseen data!
  - Methods such as testing on unseen data can help

Evaluation methodologies for ML

- Clear split of training and test data
  - Sometimes validation data as part of training data
  - All parameters set only on training data
- Leave-one-out cross validation
  - When few items are in the data, maximize training data
  - Leave-one-patient-out, limits overfitting
- N-fold cross validation
  - Separate data in N folds, then use each time one for testing and all others for training
  - Random splits can make this non-reproducible



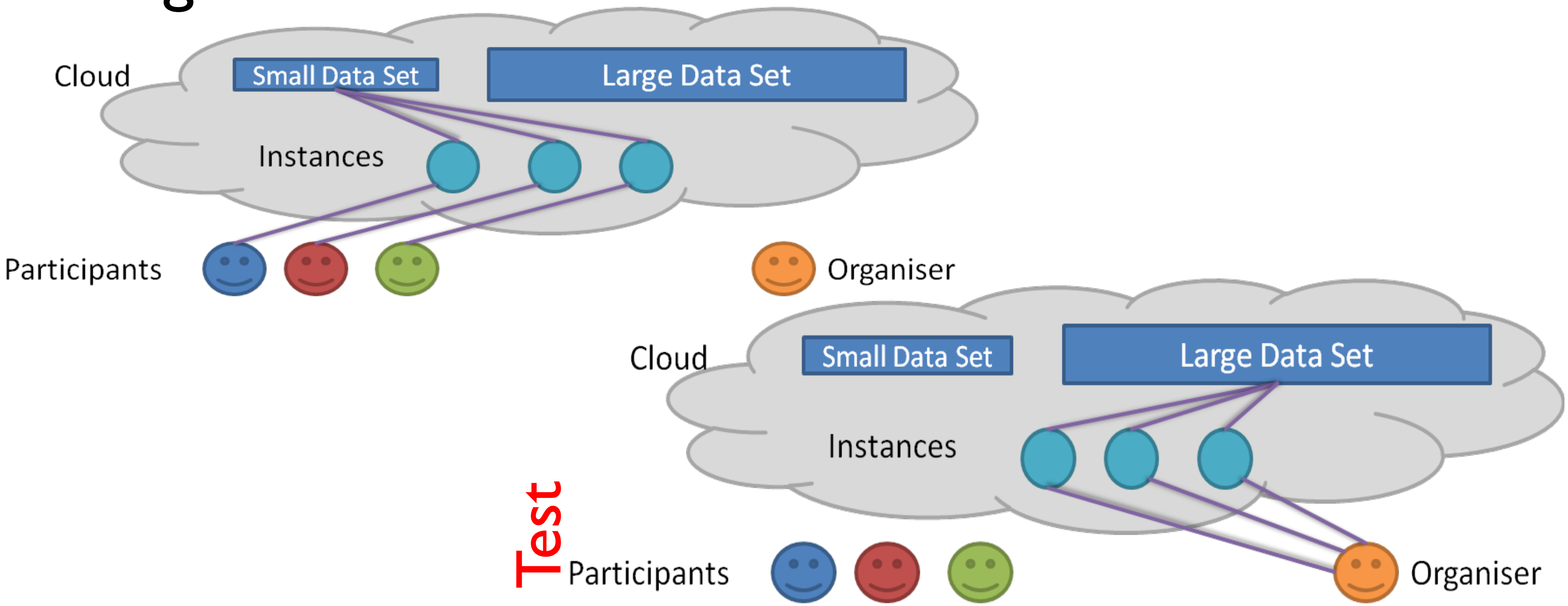
Scientific challenges

- Proposed at most conferences and workshops in machine learning and medical imaging
  - RSNA maybe to follow soon
- Same data, same evaluation methodology, workshop to discuss results among persons
  - Really make results comparable
- Commercial platforms
  - Kaggle
  - TopCoder
  - …

Evaluation-as-a-Service

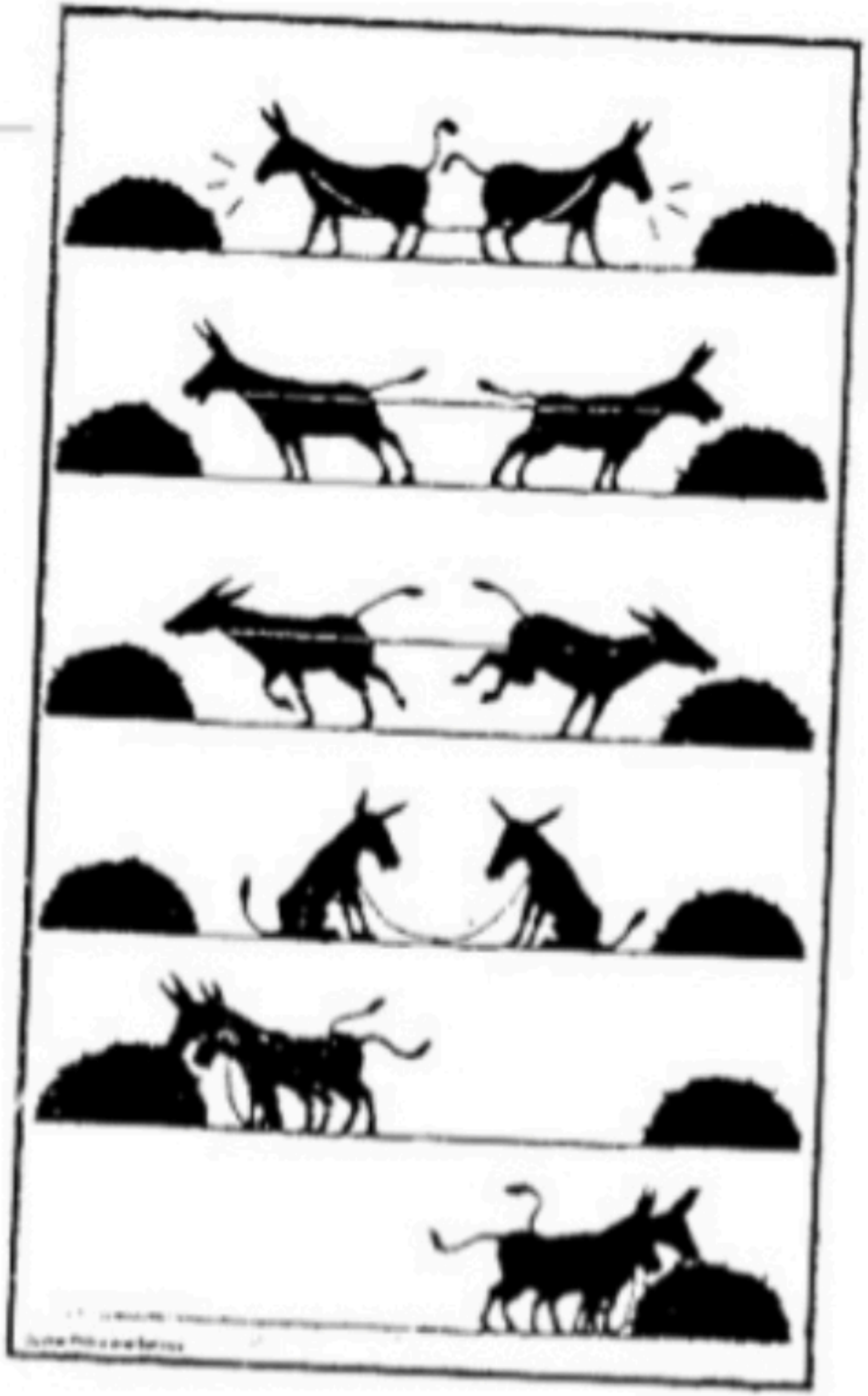
Hes·so // VALAIS WALLIS
School of Management & Tourism

- Cloud-based evaluation, bringing the algorithms to the data



Scientific environment

- Competition
- Coopetition
- Cooperation

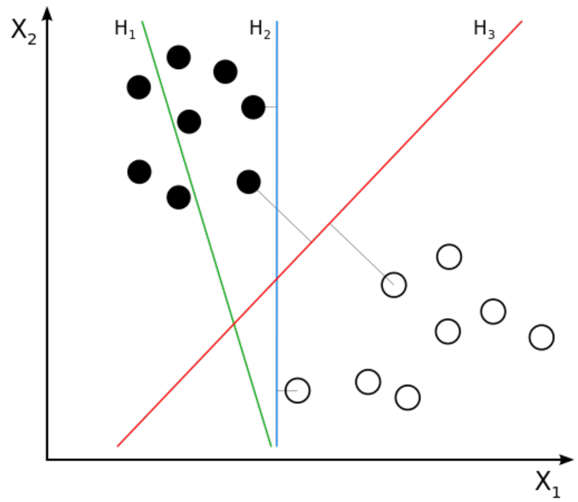


Supervised vs. unsupervised ML

- Unsupervised
  - Clustering, find structure in a data space (group items)
  - Number of clusters can be given (k-means) or not
- Weakly supervised
  - Use some existing, potentially erroneous labels
  - Images from the web with keywords close to images
    - Keywords from the radiology report to match with images
- Supervised (classification)
  - Class labels exist and are considered of good quality
  - On different levels: ROI (local), organ (regional), full image
- Reinforcement learning

Machine learning approaches

- Key nearest neighbors (kNN), simplest approach
  - Parameter free, local approach
- Decision trees
  - Random forests
- Support Vector machines (SVMs)
- Neural networks
  - More on deep learning later
- Boosting
- Linear classifiers such as naïve Bayes

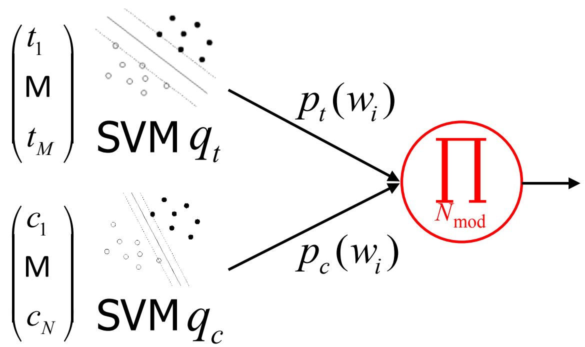


Information fusion

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, …

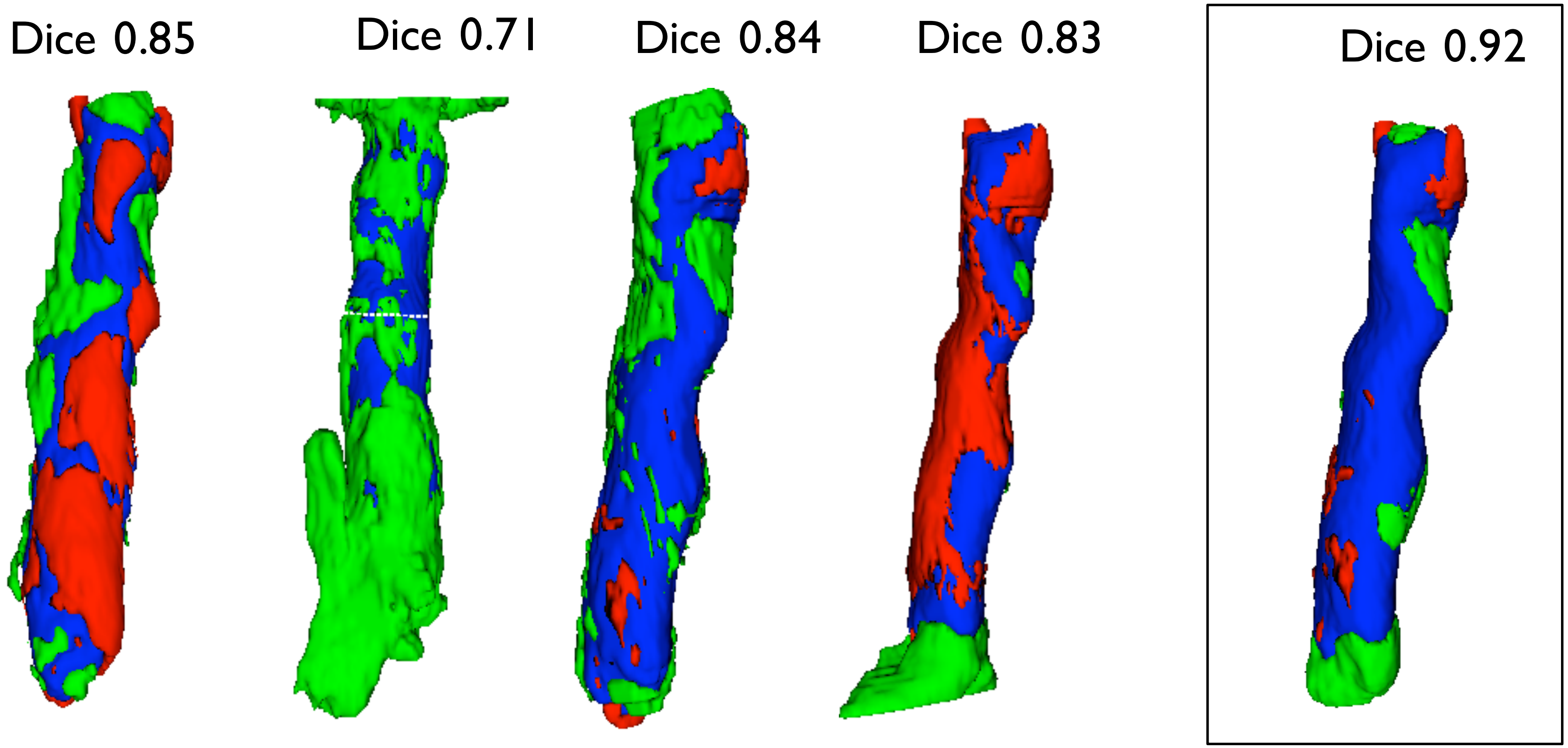
- Applications

Silver corpus (example trachea)

- Executable code of the VISCERAL scientific challenge for segmentation
  - Run it on new data, do label fusion of the results
    - Silver corpus that can be used for training

Participant segmentations

Silver Corpus

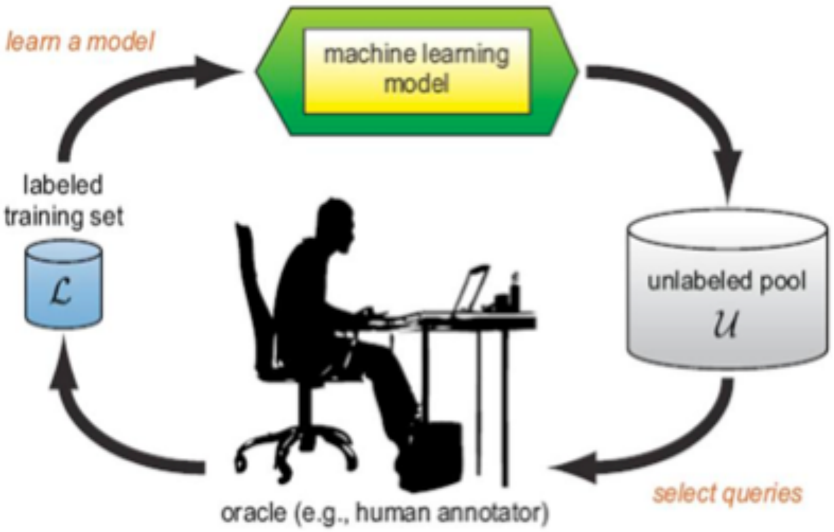


How to create labels for learning?

- Manual expert annotators
  - Expensive, time-consuming, does not scale well, quality
- Use associated documents
  - Radiology report, anamnesis, ICD codes
  - Quality not guaranteed
- Crowdsourcing
  - Non-expert annotations
  - Strict quality control is needed and focused task definition
- Leverage outcomes of automatic tools

Active learning

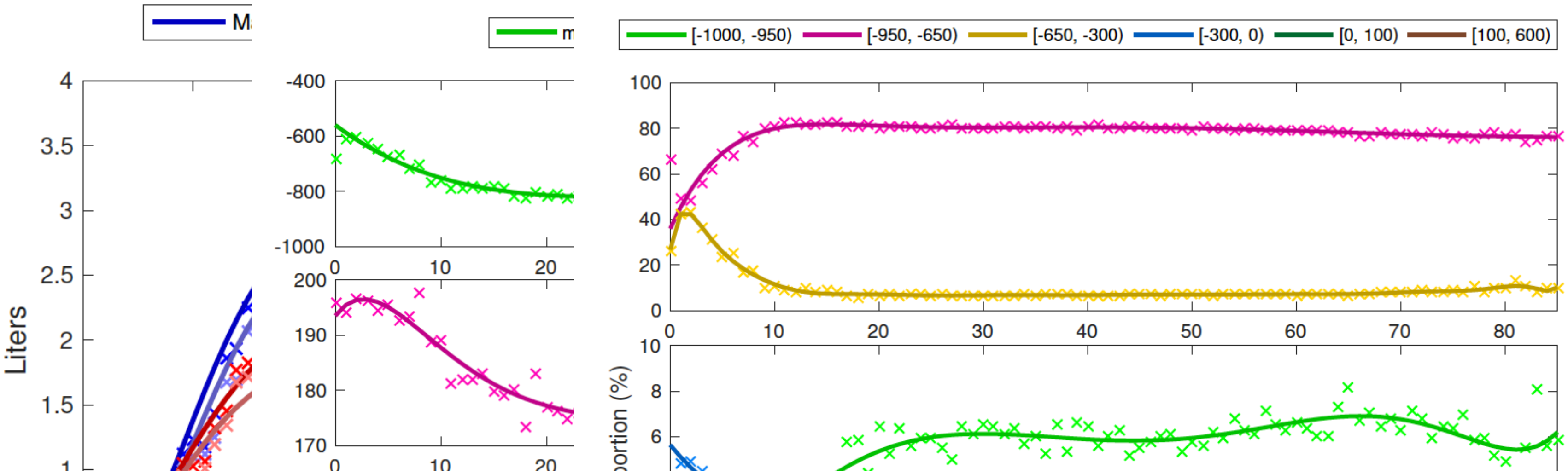
- Let the algorithm decide which non-labelled instances are best to be annotated
  - Usually iteratively to maximize information gain
- Interactive way to limit the amount of annotated data needed
- Visualization can make things easier

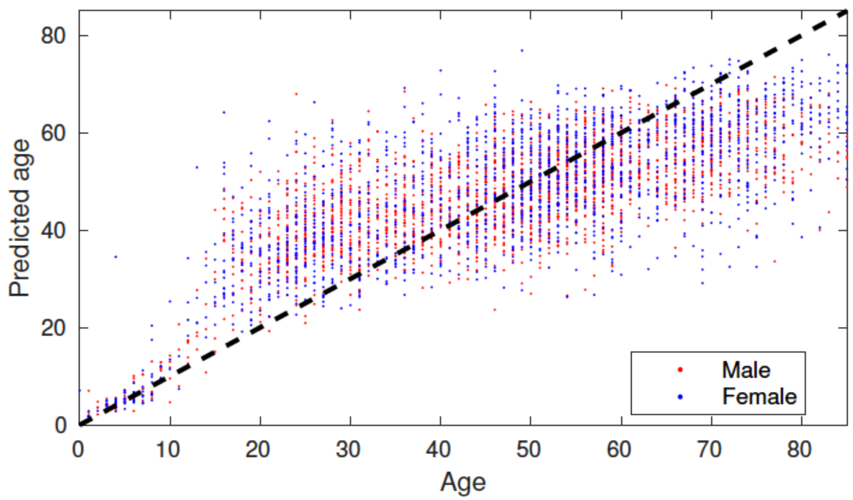


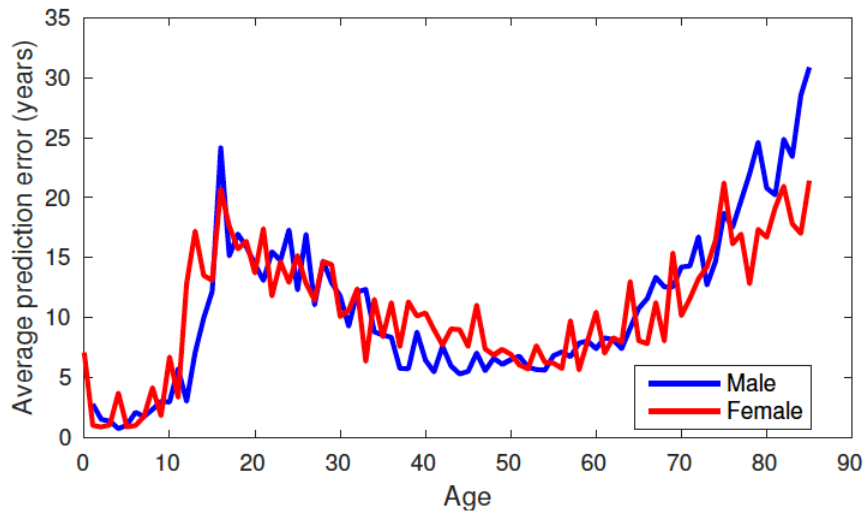
Pressforward.org

Analysis with (almost) no labels

- Age prediction from lung CTs



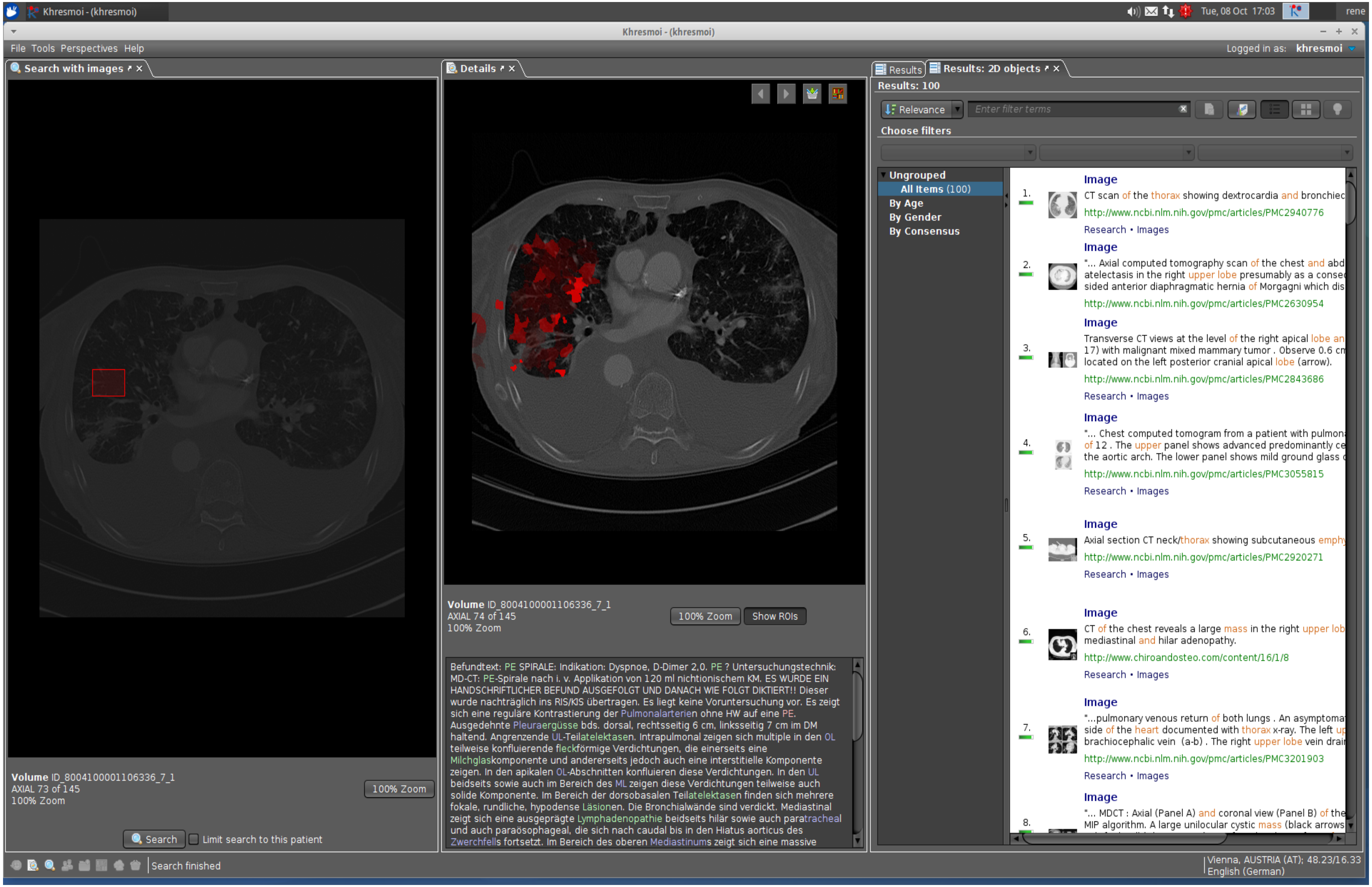




Other applications

- Detection (i.e. of lung nodules)
- Classification of diseases or disease types/states
- Retrieval of similar cases
  - Based on visual or text/structured data
  - What is similarity in this context?
    - Evaluate it based on human judgements or outcomes
- Retrieval of similar visual patterns
  - Maybe mixed with other parameters to be chosen
- How to give flexibility and keep interfaces simple?

Khresmoi4radiology interface

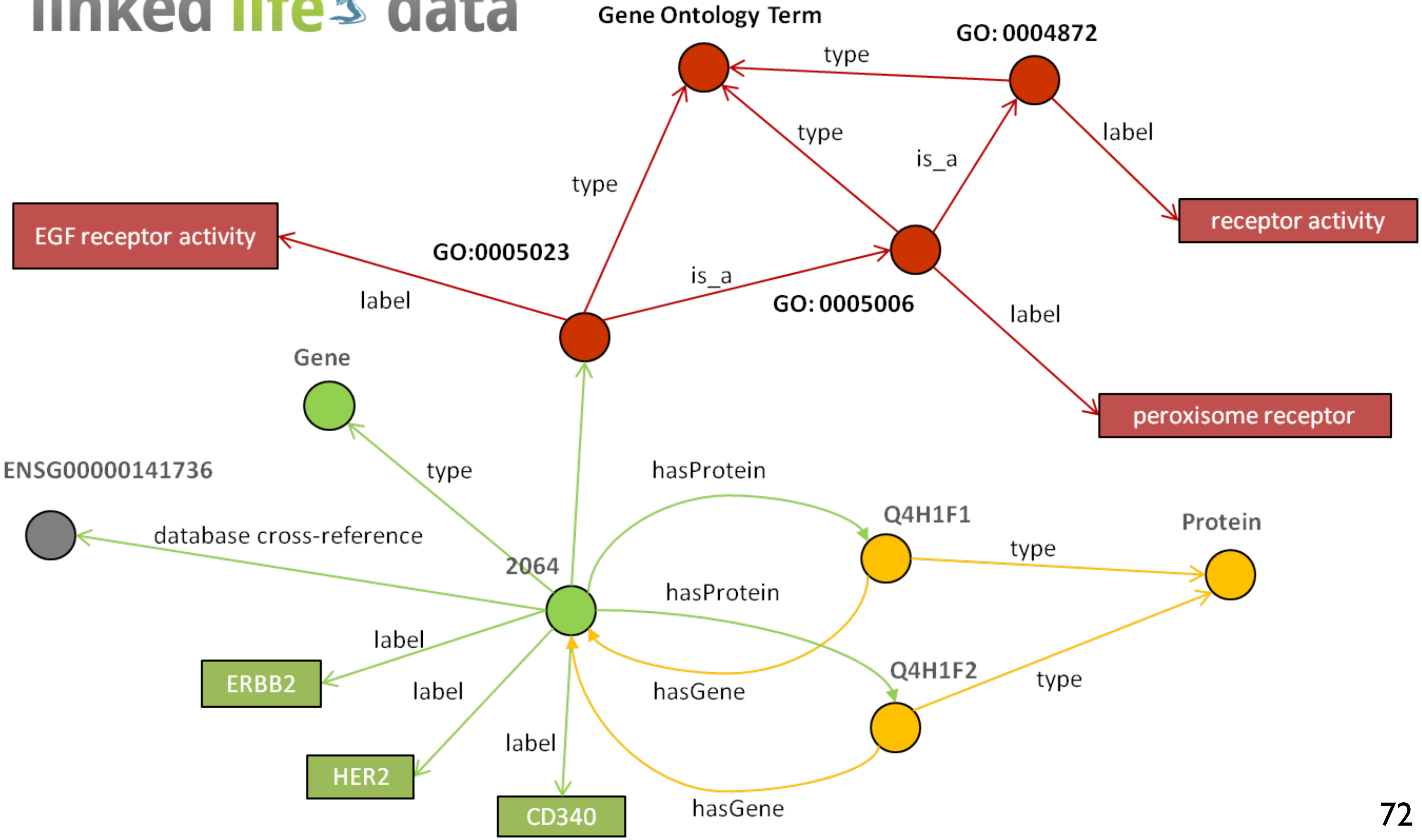


Example for semantic text analysis

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

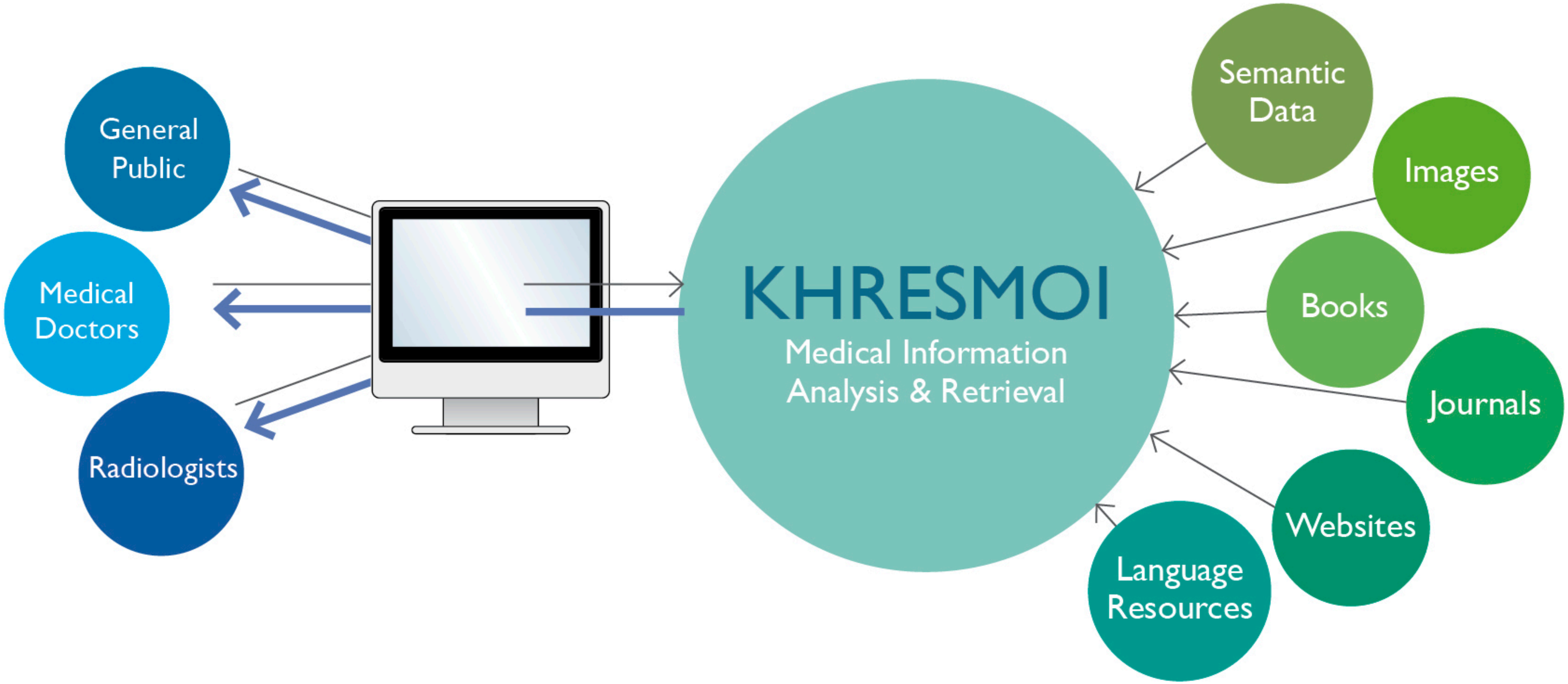
LinkedLifeData



khresmoi
MEDICAL INFORMATION ANALYSIS & RETRIEVAL

Hes·so// VALAIS WALLIS
School of Management & Tourism

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

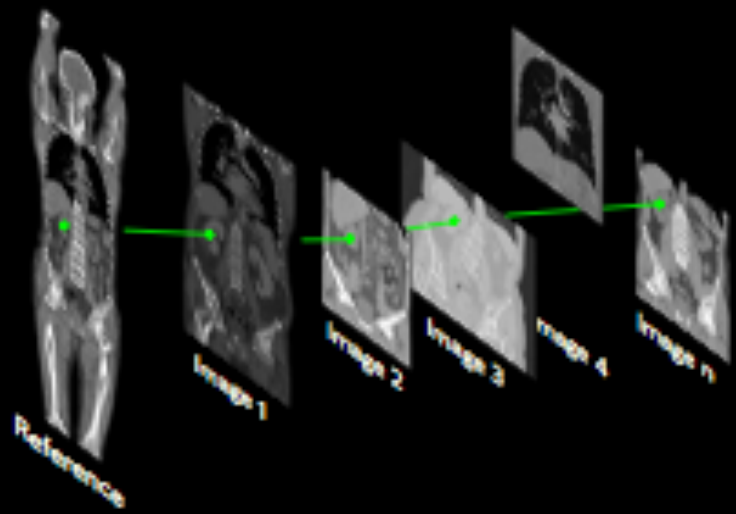


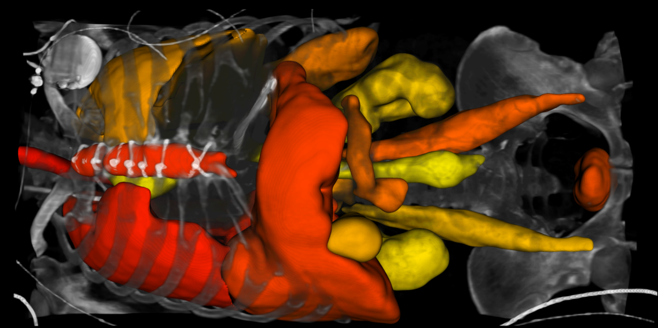
Allan Hanbury, Célia Boyer, Manfred Gschwandtner, Henning Müller, KHRESMOI: Towards a Multi-Lingual Search and Access System for Biomedical Information, Med-e-Tel, pages 412-416, Luxembourg, 2011.

KHRESMOI: towards retrieval in clinical data

Hofmanninger CVPR 2015

- Mapping imaging data to a reference space
- Compare observations across individuals
- Learn from clinical routine data

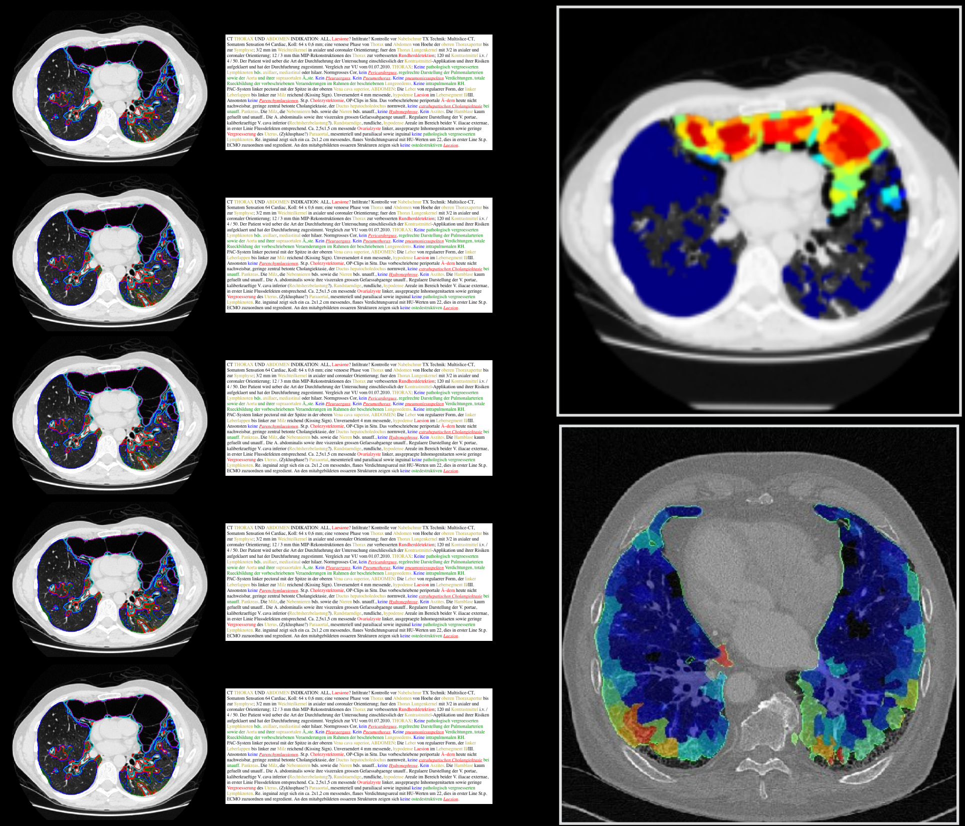




KHRESMOI: towards retrieval in clinical data

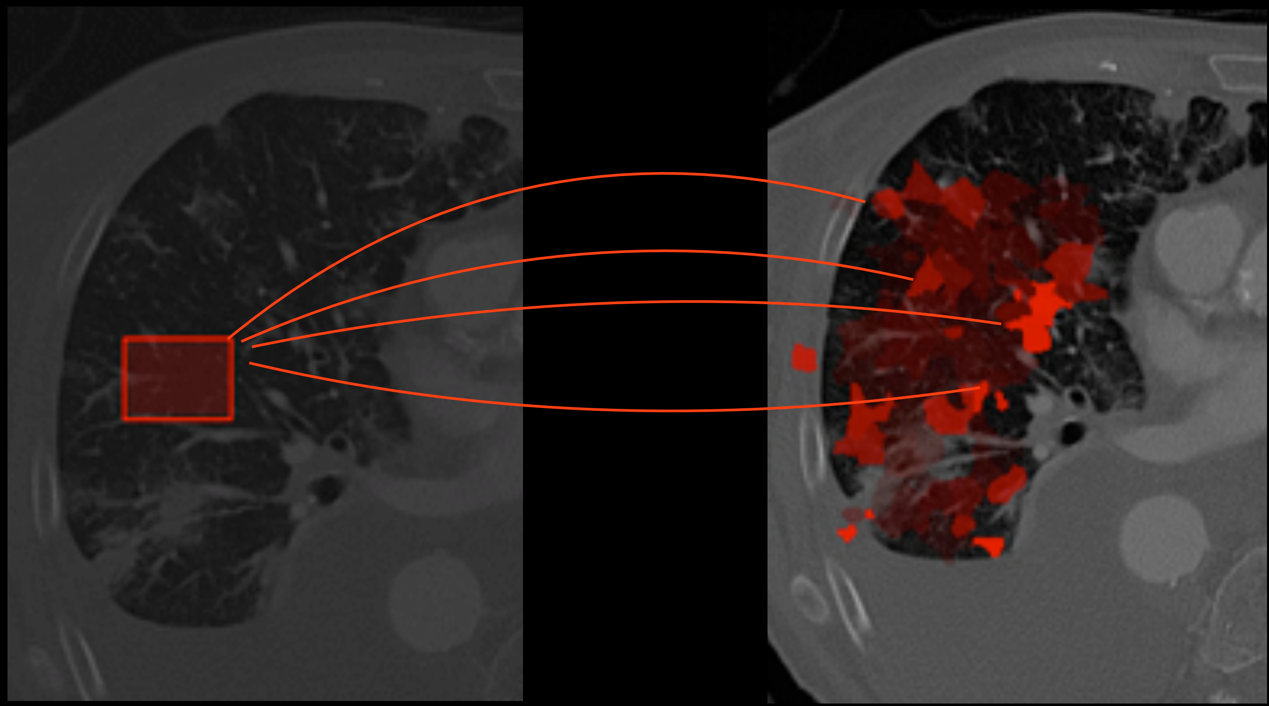
Hofmanninger CVPR 2015

- 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

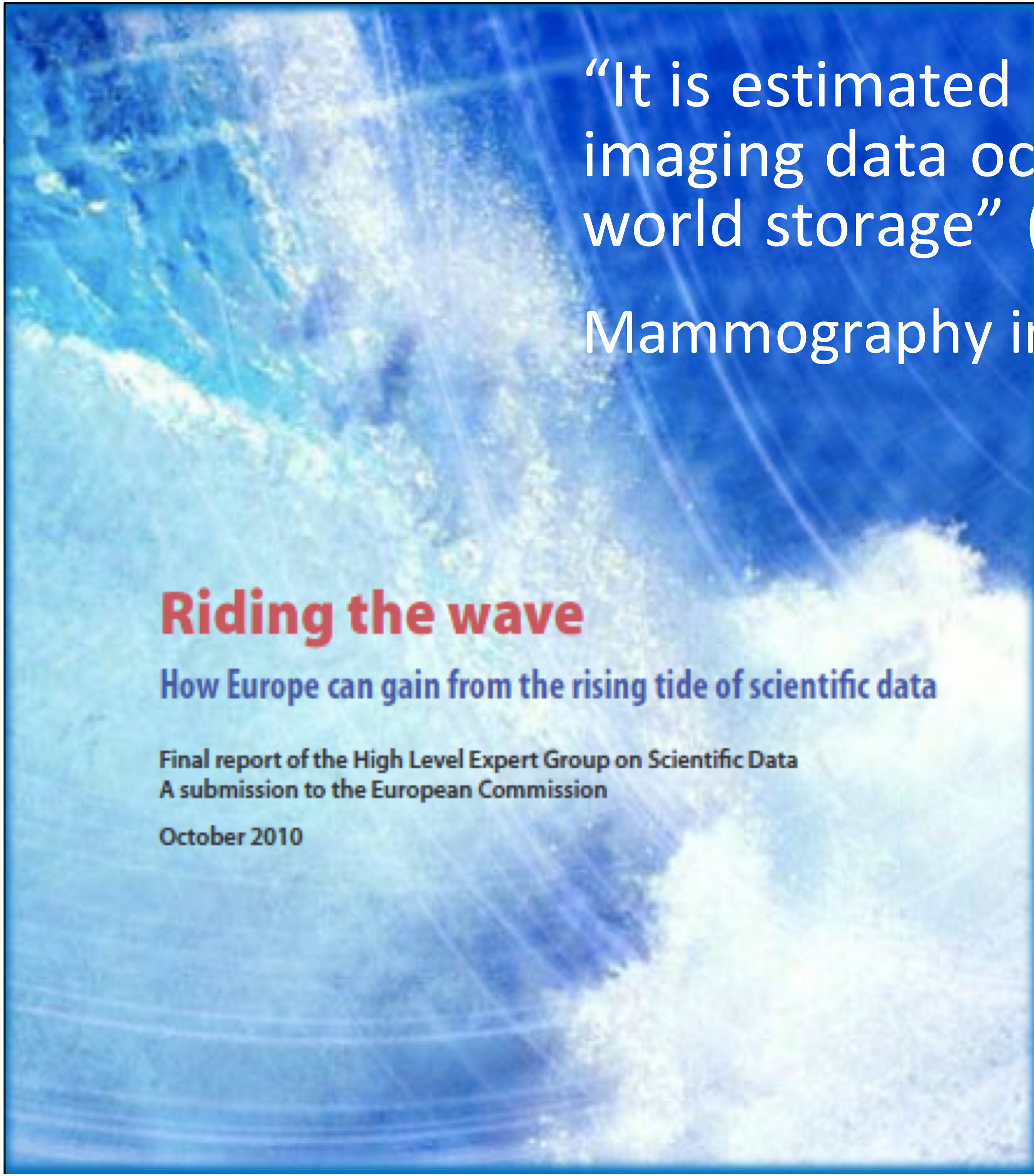
We can use these features for retrieval



- Learn visual features that reflect findings / clinically relevant information
- Establish similarities between cases, and use it for search during clinicial routine

"It is estimated that radiological imaging data occupies 30% of world storage" (2010)

Mammography in the US: 3PB/year



Big data, …

- Three, four, five Vs
  - Volume, velocity, variety, veracity, value
- Big data is when data can not easily be processed by "normal" means
- Thin data (only factual)
  - A mile wide but an inch deep, facts on client behavior
- Thick data (data with context)
  - Understanding the why of the behavior, confirm hypotheses

Conclusions

Contact and more information

- More information can be found at
  - http://khresmoi.eu/
  - http://visceral.eu/
  - http://medgift.hevs.ch/
  - http://publications.hevs.ch/

- Contact:
  - Henning.mueller@hevs.ch