

Multi-Modal Deep Learning

Lecture Series | Summer 2025

Institute for AI and Informatics in Medicine

29.04.2025

Objects, Measurements, Data

RECAP FROM EARLIER

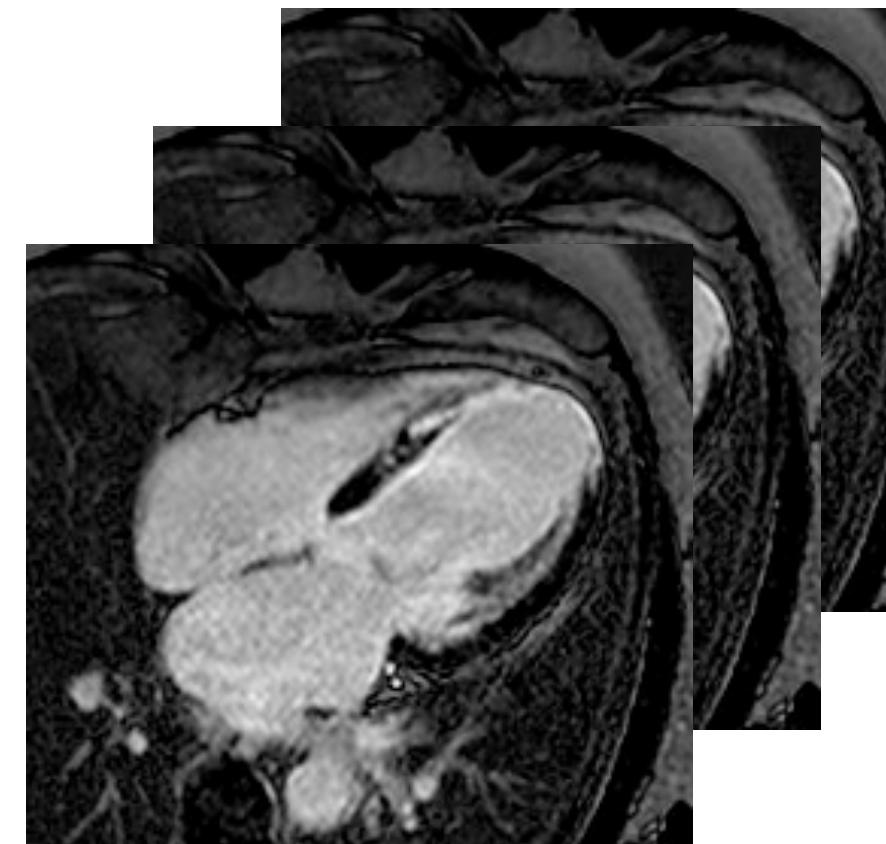
What is a modality?

"Type of data used as input (conditioning) or output (prediction) of a model"

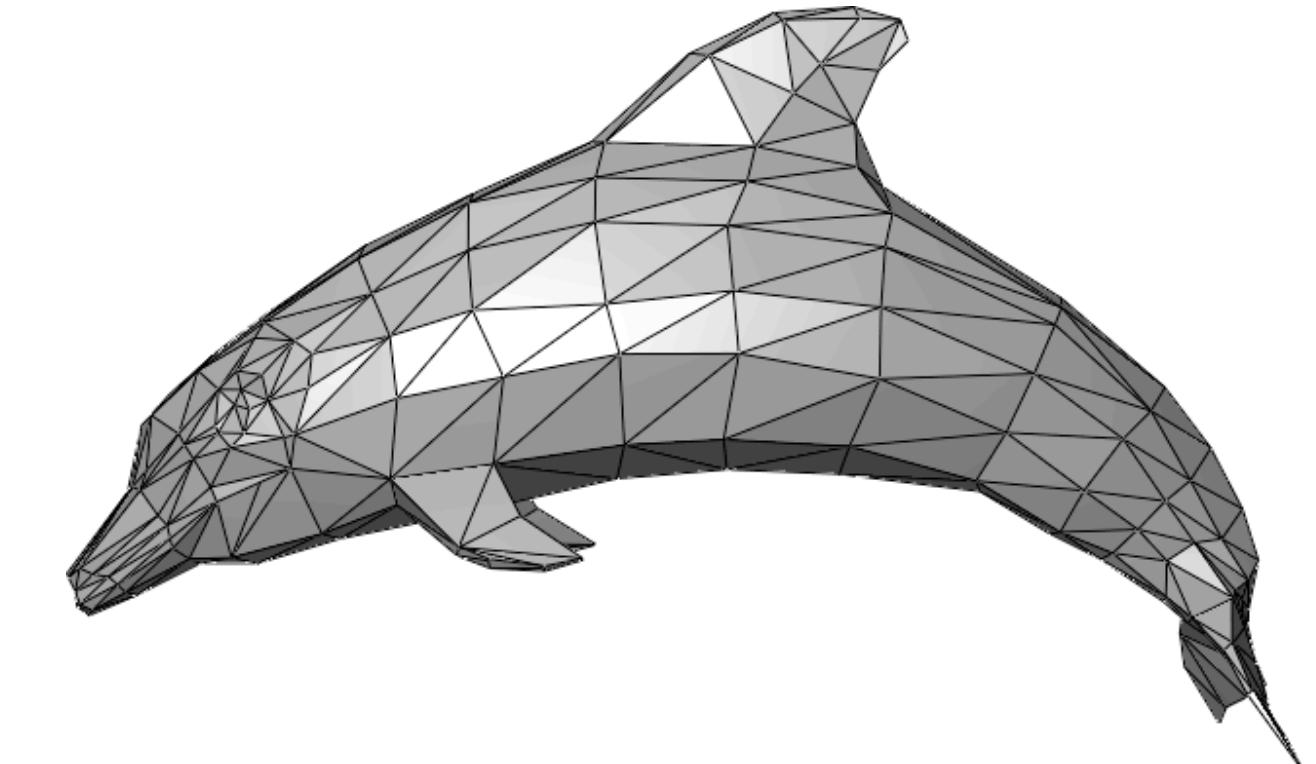
Examples:



Natural images



Medical scans



Geometric meshes

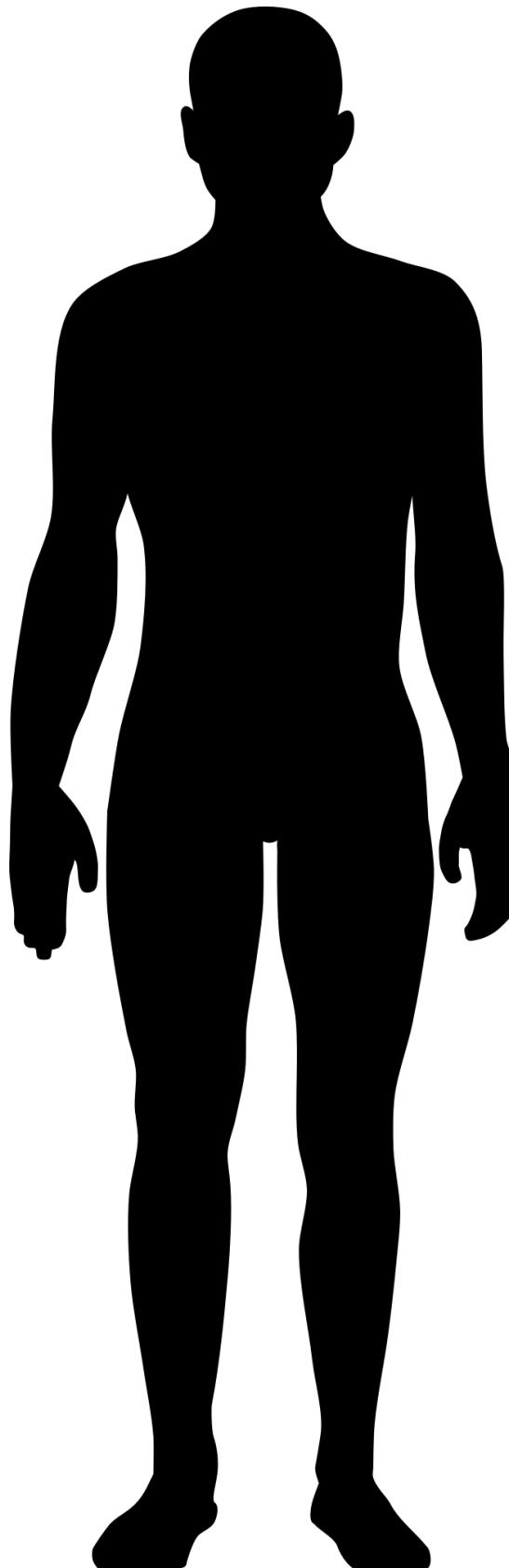
```
>Fu9D-ITS1F_D20.ab1
NNNNNNNNNNANCTCGGAAGGNNCATTATTGAATAAACATGGTTGGTTAGCTGACTCCTGGAGTATGTGACACCT
GCCGTCTTATCTATCCACCTGTGCACACATTGTAGTCTTGGGGATTGATTAGTGACAAATTGTTGCAATGTCGCC
TCCGAGGTCTATGTTATCATAAACCACTTAGTATGTCGTAGAATGAAGTATTTGGGCCTTAGTGCCCTATAAAACAAAATA
CAACTTCAGCAACGGATCTTGGCTCTCGCATCGATGAAGAACGCGAGCGAAATGCGATAAGTAATGTGAATTGAGAA
TTCAGTGAATCATCGAATCTTGAACGCACCTTGCCTGGTATTCCGAGGGAGCATGCCTGTTGAGTGTCACTAAA
TTCTCAACCCCTCCAGCTTTGTTGGCTGGCTGGCTGGATATGGGAGTTTGCTGGTCTCATTCGAGATCAGCTCTC
```

DNA Sequences

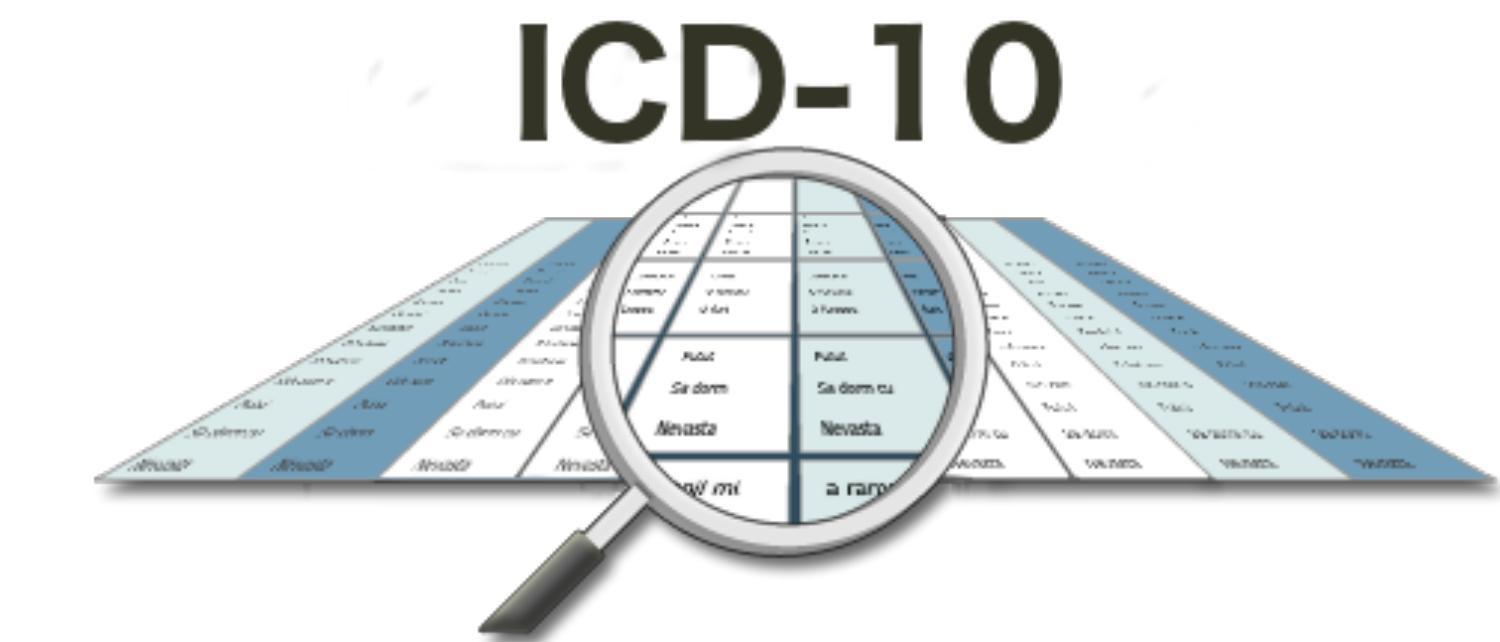
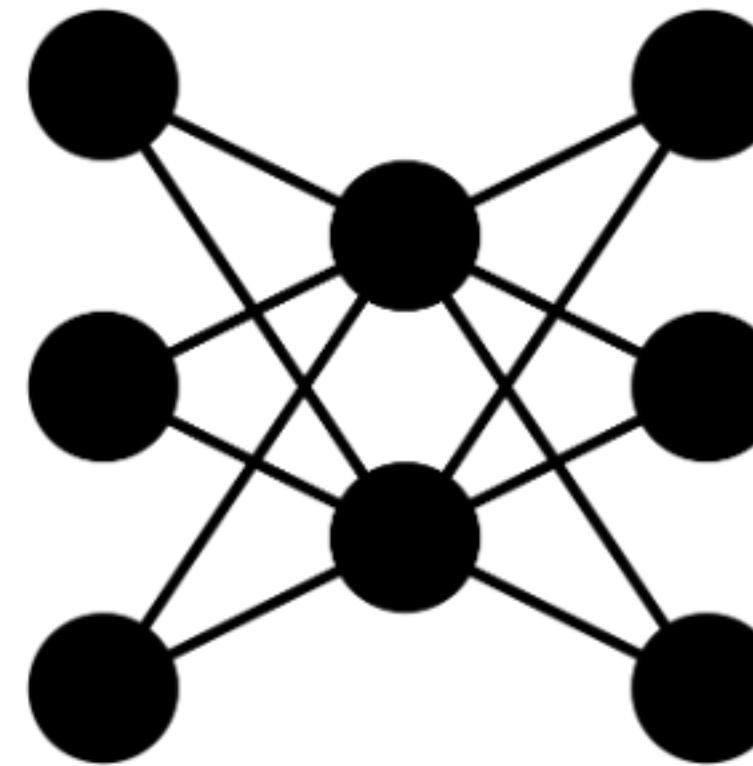
Smoker	Age	BP	BMI	Sex	Fitness	Alcohol
FALSE	62	150/90	29.2	Male	High	Moderate

Tabular data

RECAP FROM EARLIER



Object



Prediction

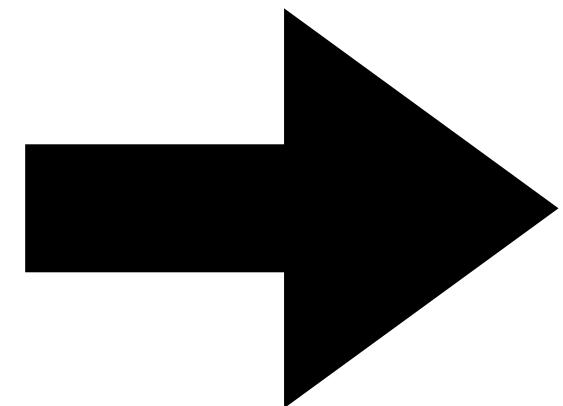
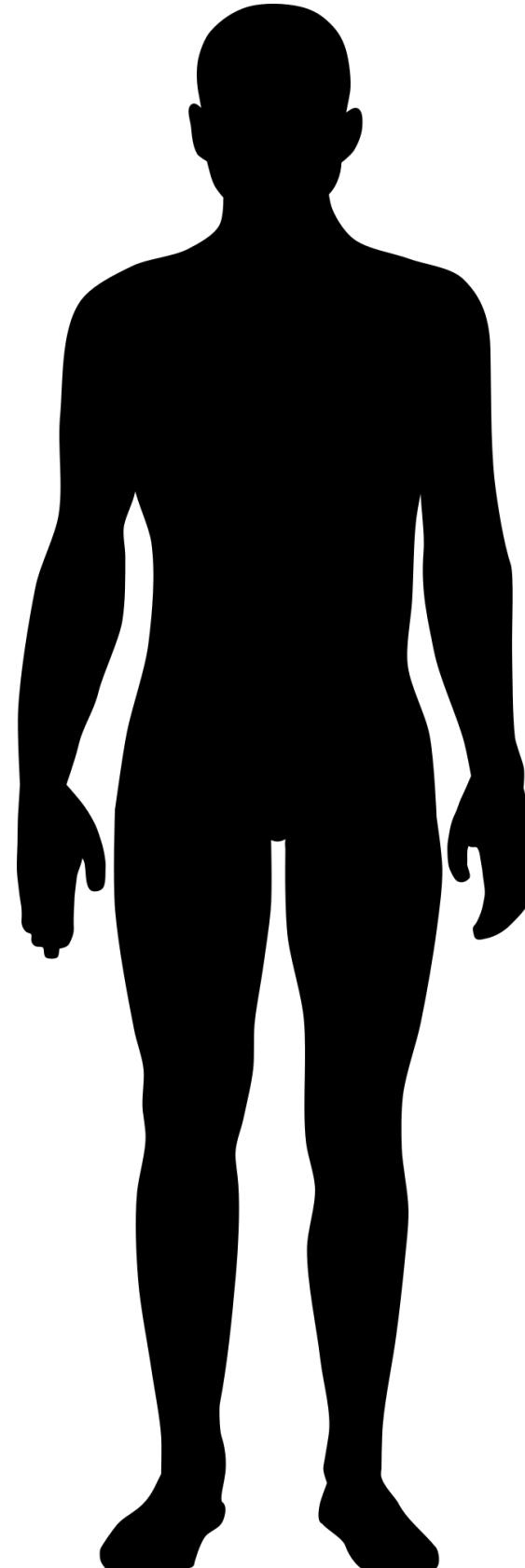
You cannot process a human with a neural network

DATA

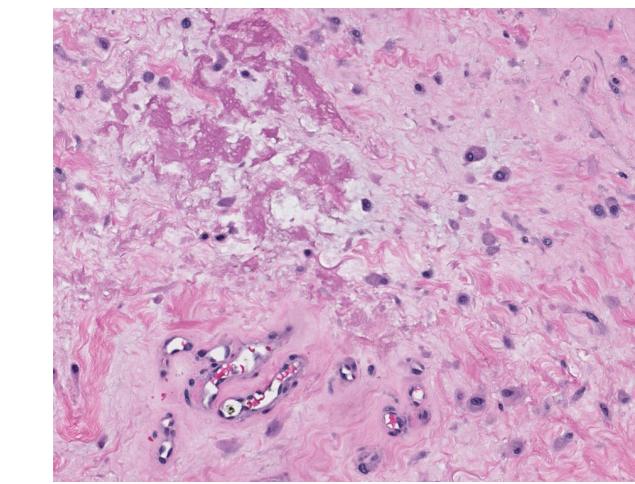
Definition from Wikipedia:

*"Data are a **collection of discrete or continuous values that convey information**, describing the quantity, quality, fact, statistics, [or] other basic units of meaning, [...] that may be further interpreted formally."*

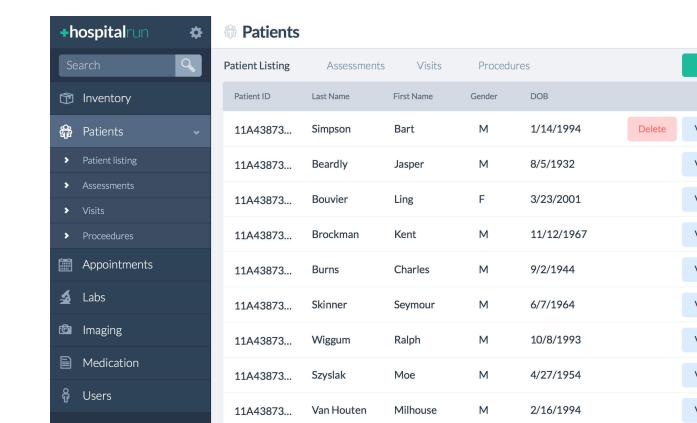
OBJECTS AND DATA



3D CT images



2D histopathology images

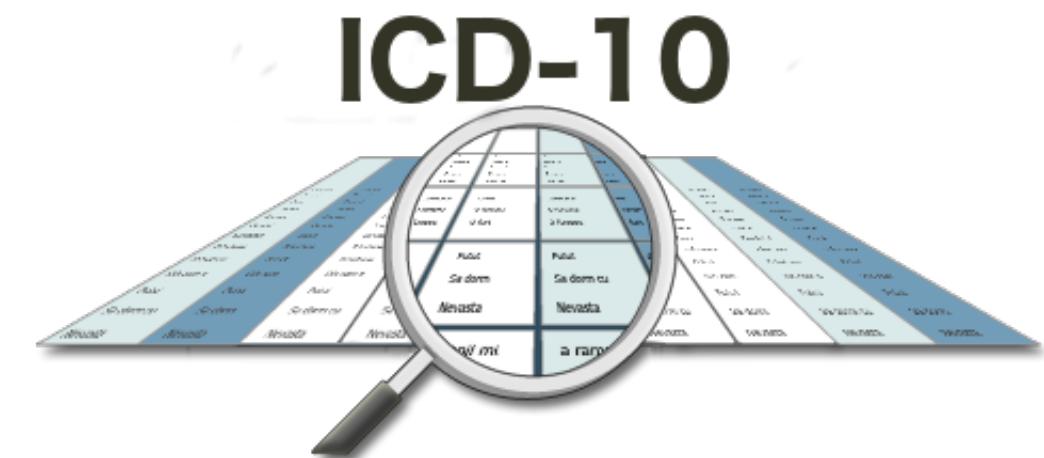
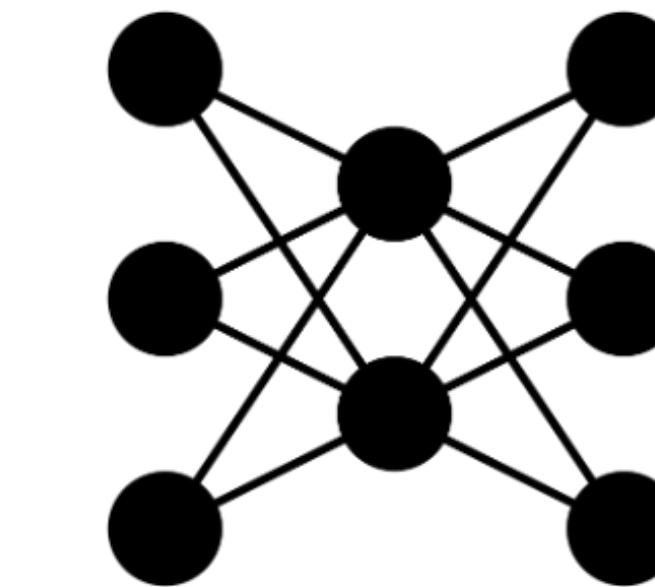


Patient ID	Last Name	First Name	Gender	DOB	Action
11A43873...	Simpson	Bart	M	1/14/1994	<button>Delete</button> <button>View</button> <button>Edit</button>
11A43873...	Bearlby	Jasper	M	8/5/1932	<button>View</button> <button>Edit</button>
11A43873...	Bouvier	Ling	F	3/23/2001	<button>View</button> <button>Edit</button>
11A43873...	Brockman	Kent	M	11/12/1967	<button>View</button> <button>Edit</button>
11A43873...	Burns	Charles	M	9/2/1944	<button>View</button> <button>Edit</button>
11A43873...	Skinner	Seymour	M	6/7/1964	<button>View</button> <button>Edit</button>
11A43873...	Wiggum	Ralph	M	10/8/1993	<button>View</button> <button>Edit</button>
11A43873...	Styliak	Moe	M	4/27/1954	<button>View</button> <button>Edit</button>
11A43873...	VanHouten	Milhouse	M	2/16/1994	<button>View</button> <button>Edit</button>

Electronic health record

Object

Data



Prediction

How is data abstracted from the underlying information or object?

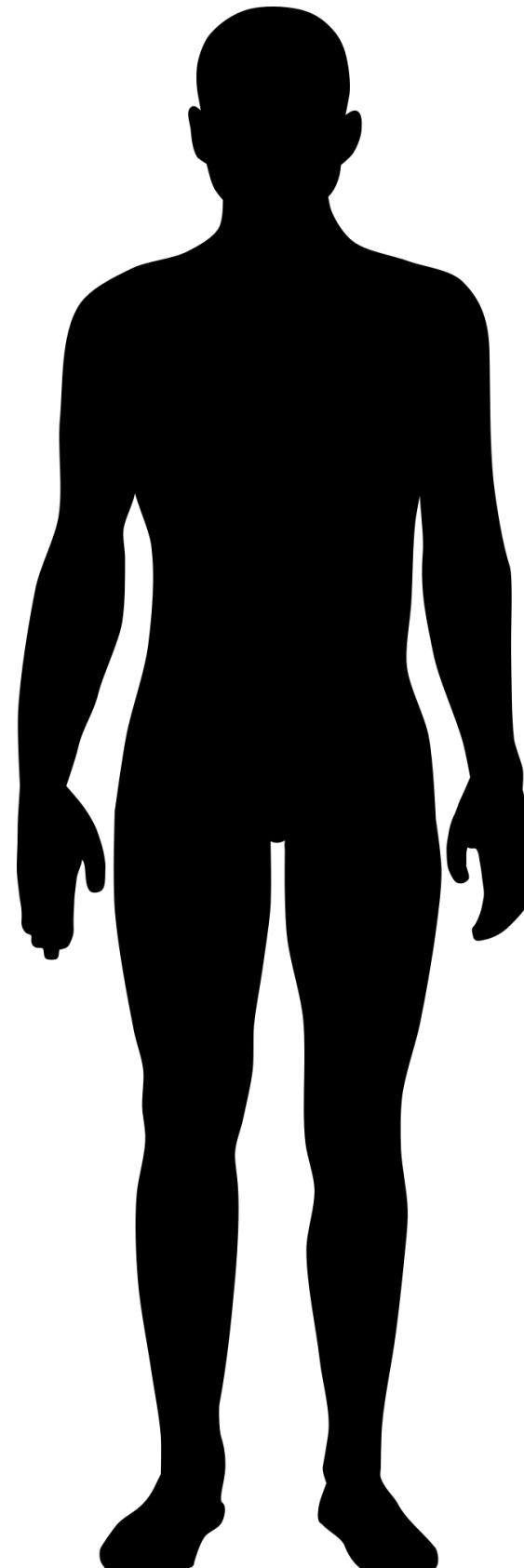
OBJECTS AND DATA

Definition from Wikipedia:

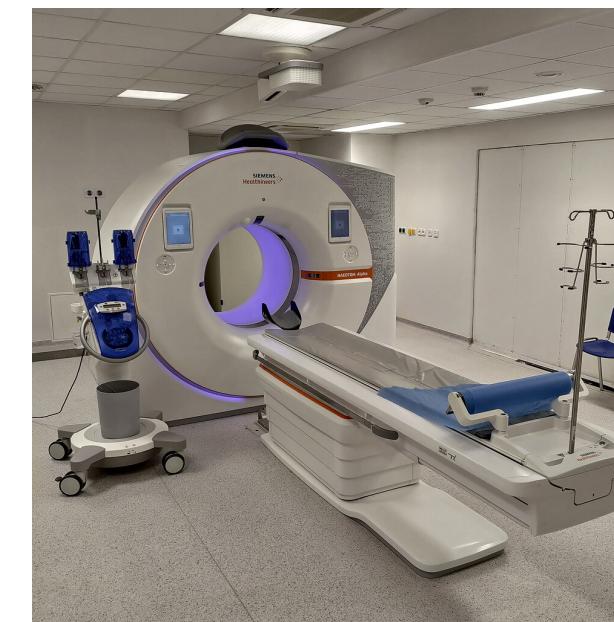
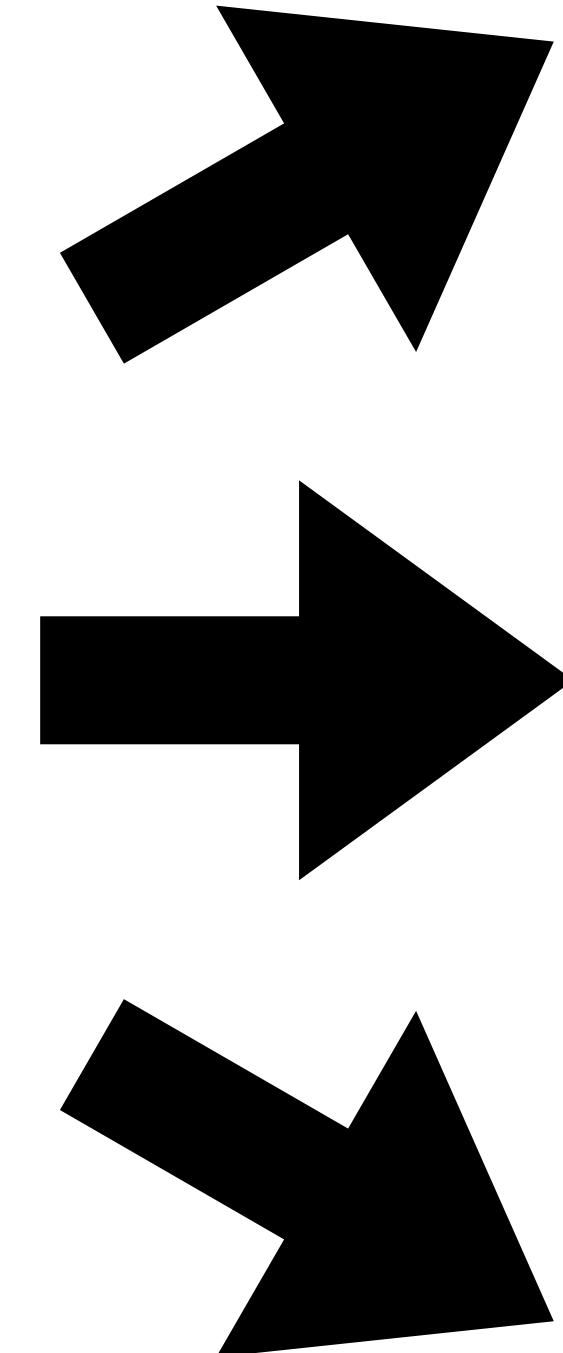
"Data are a collection of discrete or continuous values that convey information, describing the quantity, quality, fact, statistics, [or] other basic units of meaning, [...] that may be further interpreted formally."

*"Data are **collected** using techniques such as **measurement**, observation, query, or analysis."*

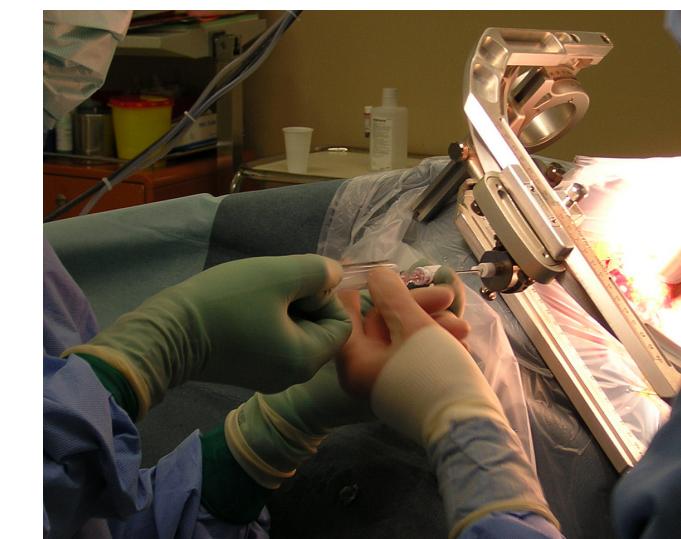
OBJECTS, MEASUREMENTS, DATA



Object



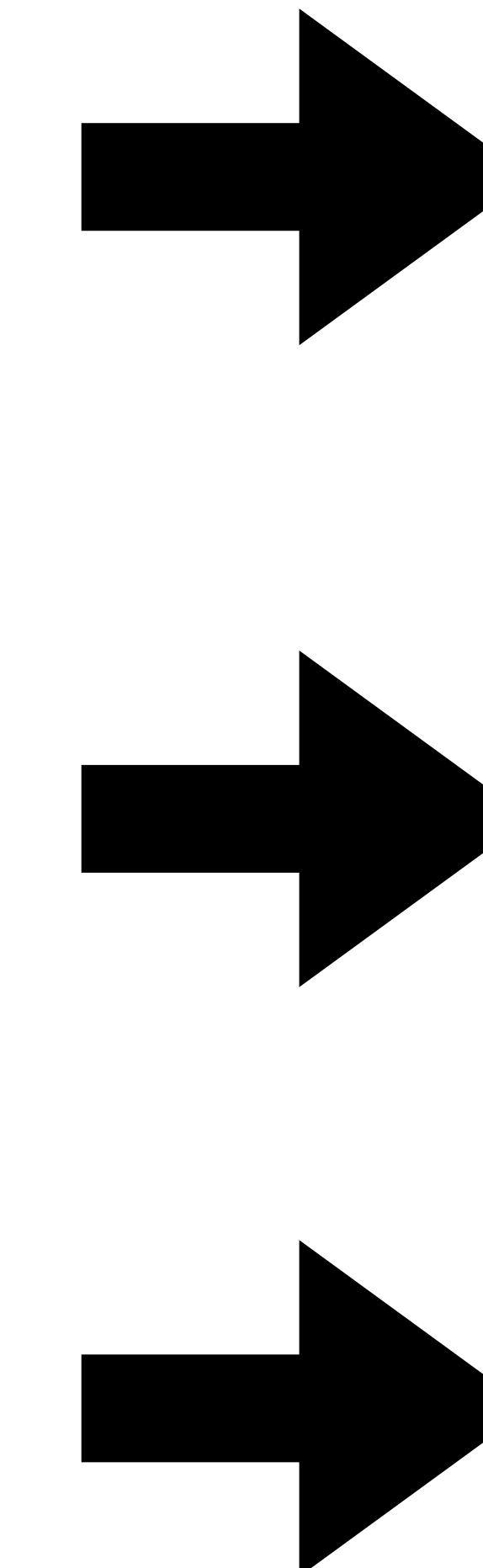
Medical imaging



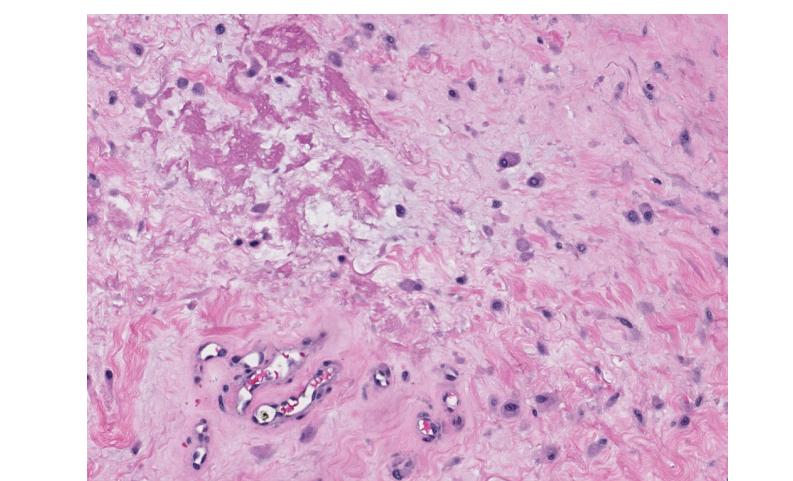
Biopsy



Patient interview
Measurement



3D CT images



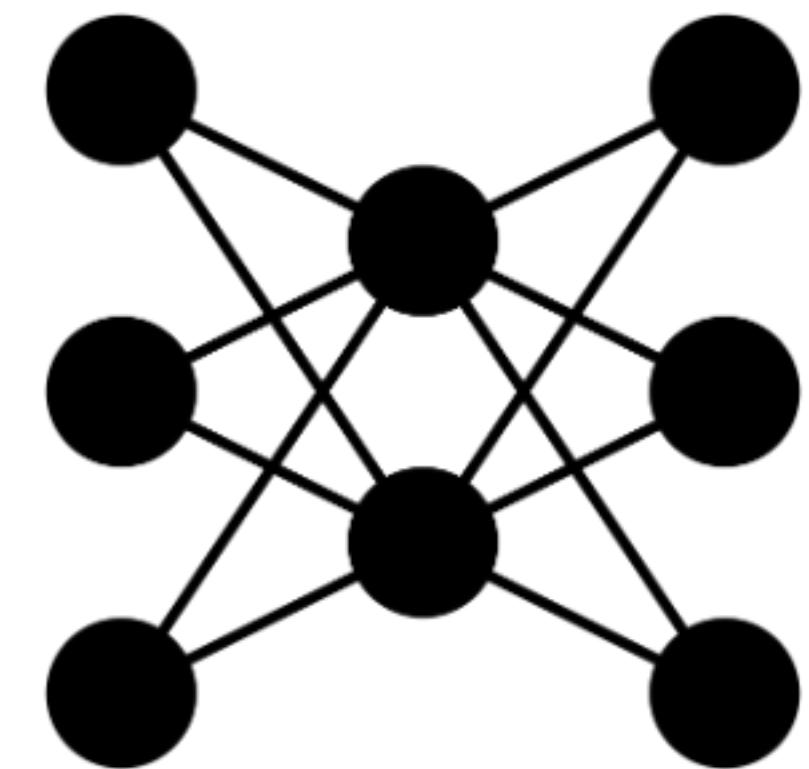
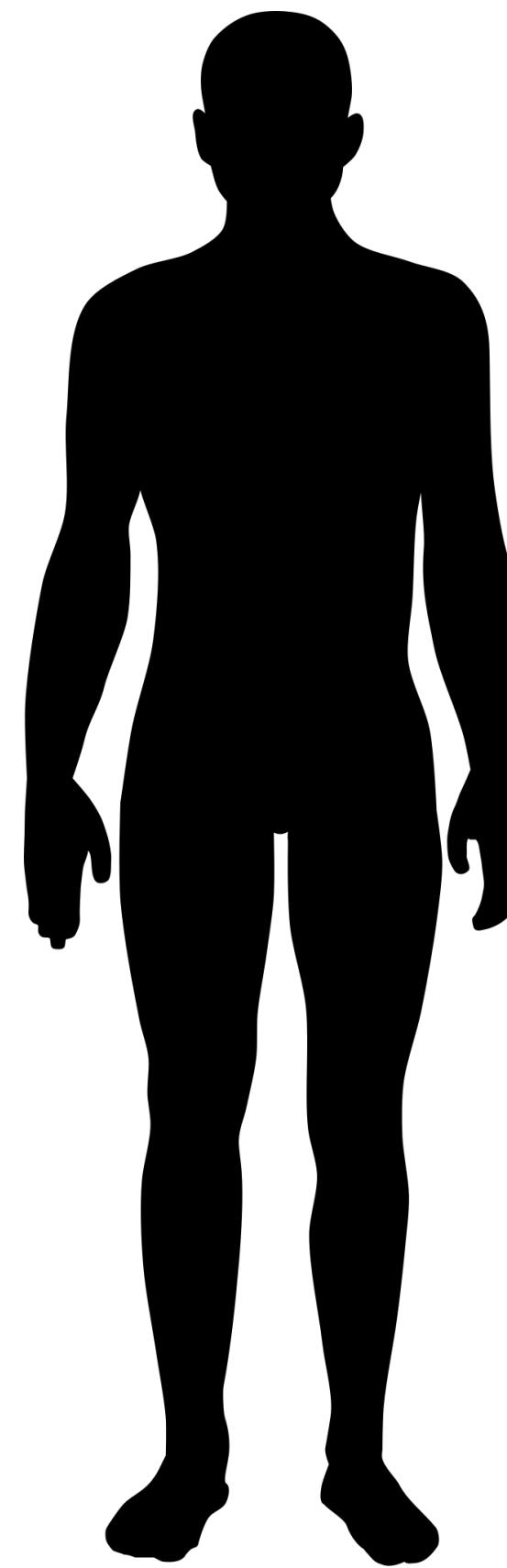
2D histopathology images

Patients				
Patient ID	Last Name	First Name	Gender	DOB
11A43873...	Simpson	Bart	M	1/14/1994
11A43873...	Beardly	Jasper	M	8/5/1932
11A43873...	Bouvier	Ling	F	3/23/2001
11A43873...	Brockman	Kent	M	11/12/1967
11A43873...	Burns	Charles	M	9/2/1944
11A43873...	Skinner	Seymour	M	6/7/1964
11A43873...	Wiggum	Ralph	M	10/8/1993
11A43873...	Szyslak	Moe	M	4/27/1954
11A43873...	Van Houten	Milhouse	M	2/16/1994

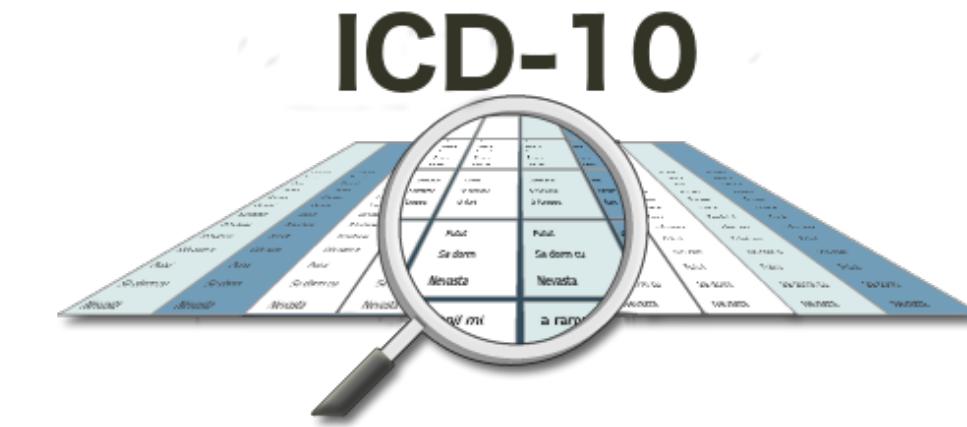
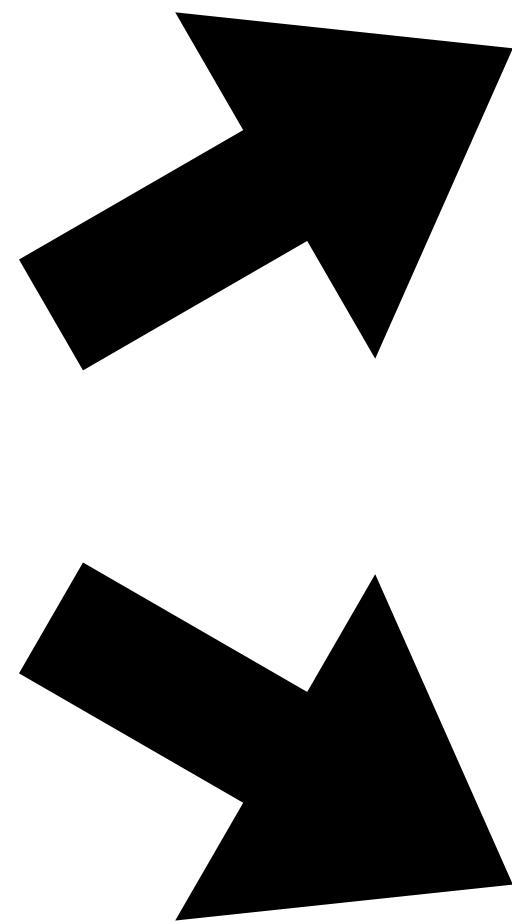
Electronic health record
Data

A modality comprises the object, measurement and data

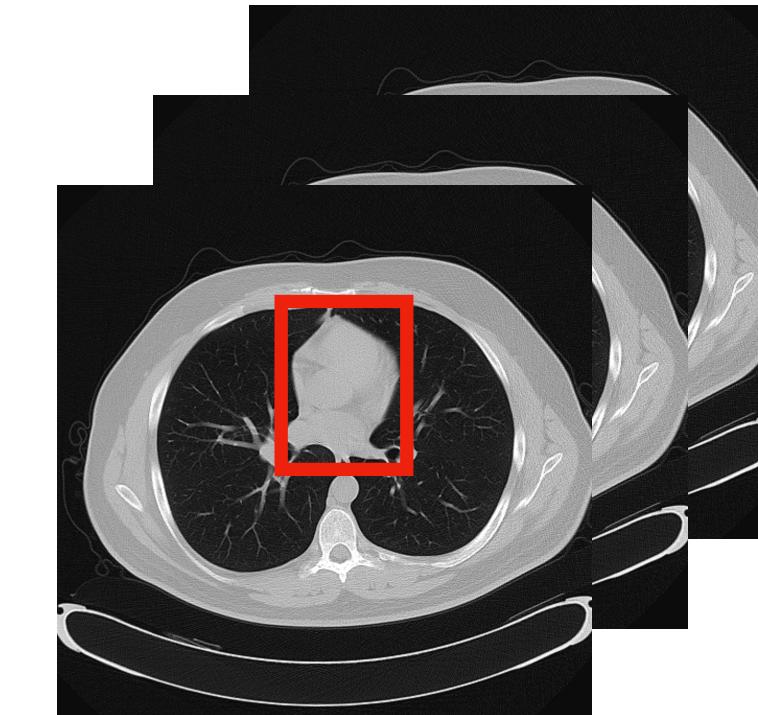
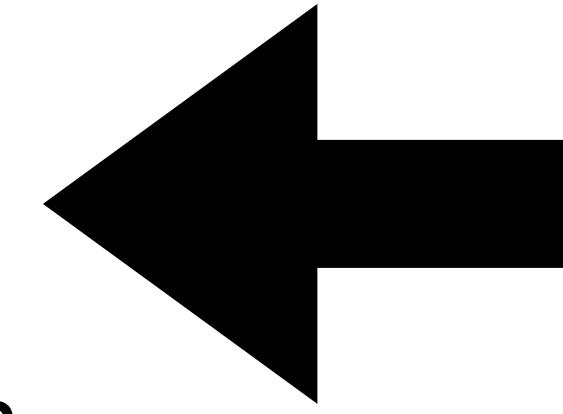
OBJECTS, MEASUREMENTS, DATA



Object

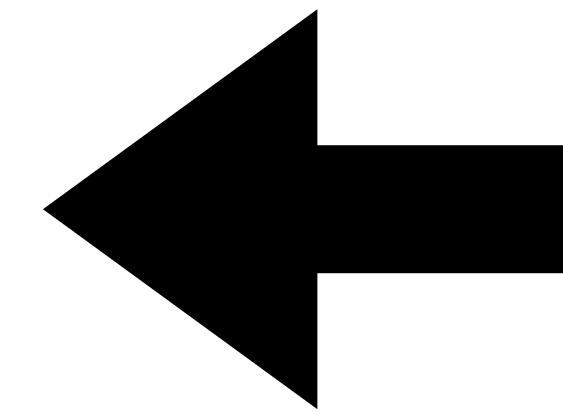


Disease classification



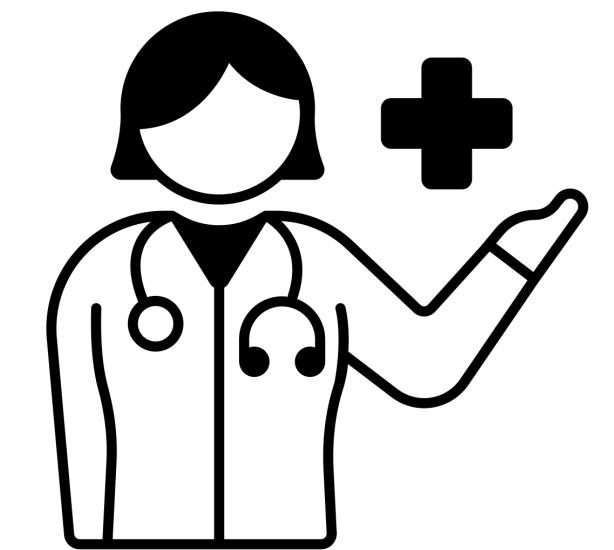
Segmentation mask

Prediction



Patients			
Patient ID	Last Name	First Name	Gender
11A43073..	Simpson	Bart	M
11A43073..	Beardly	Jasper	M
11A43073..	Bouvier	Ling	F
11A43073..	Brockman	Kent	M
11A43073..	Burns	Charles	M
11A43073..	Skiner	Seymour	M
11A43073..	Wiggum	Ralph	M
11A43073..	Sybilak	Moe	M
11A43073..	Van Houten	Milhouse	M

Electronic health record



Expert annotation

Measurement

Outputs are modalities as well

OBJECTS, MEASUREMENTS, DATA

Definition from Wikipedia:

*"Data are a collection of discrete or continuous values that convey information, describing the quantity, quality, fact, statistics, [or] other basic units of meaning, [...] that may be **further interpreted formally**."*

"Data are collected using techniques such as measurement, observation, query, or analysis."

Data types and their structure

SETS

Definition: a set is a collection of distinct elements without any particular order

$$\mathcal{S} = \{x_1, x_2, \dots, x_n\}, \quad x_i \in \mathcal{X}$$

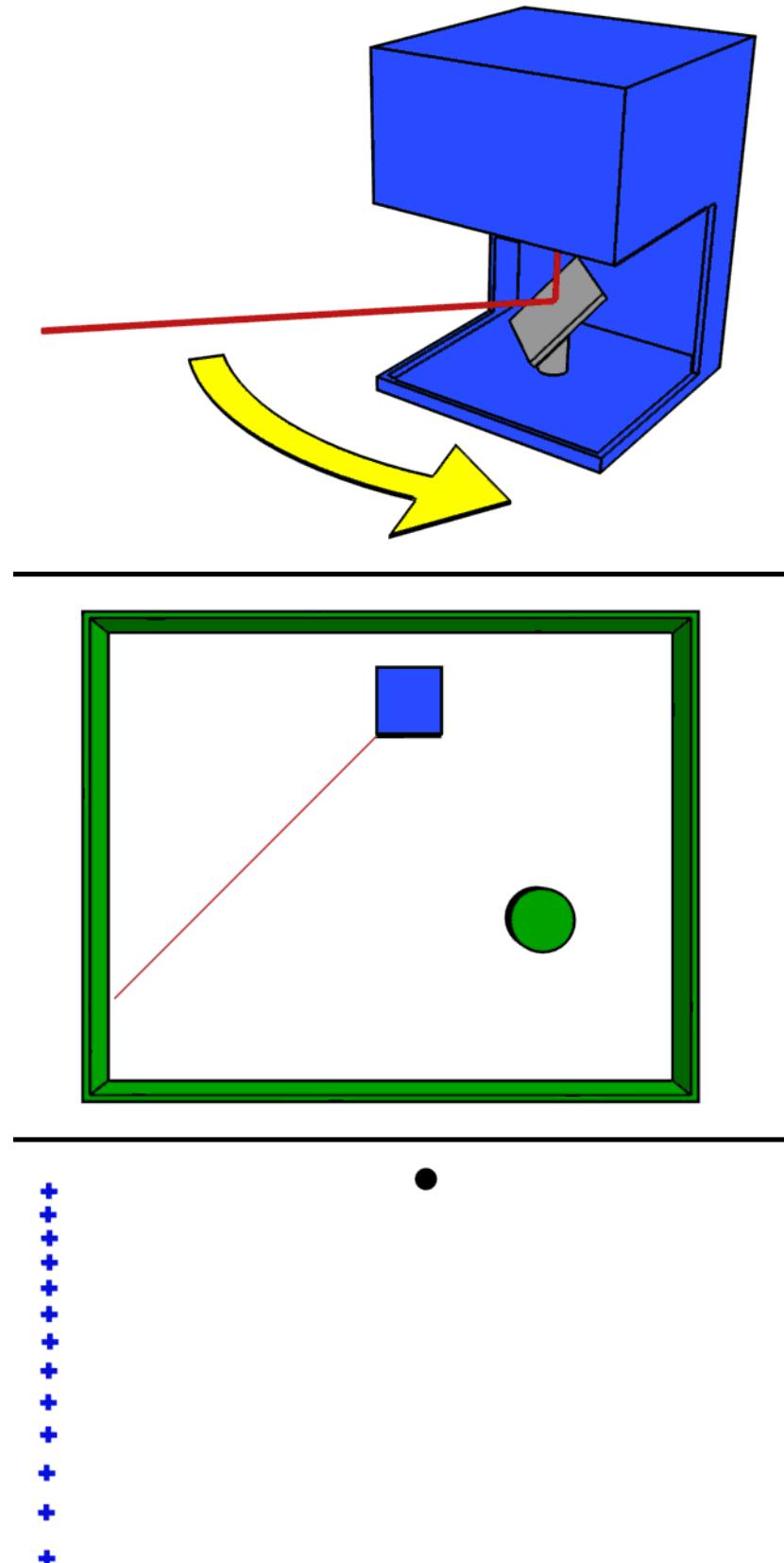
**Structure comes from space
instead of containing set**

Key properties:

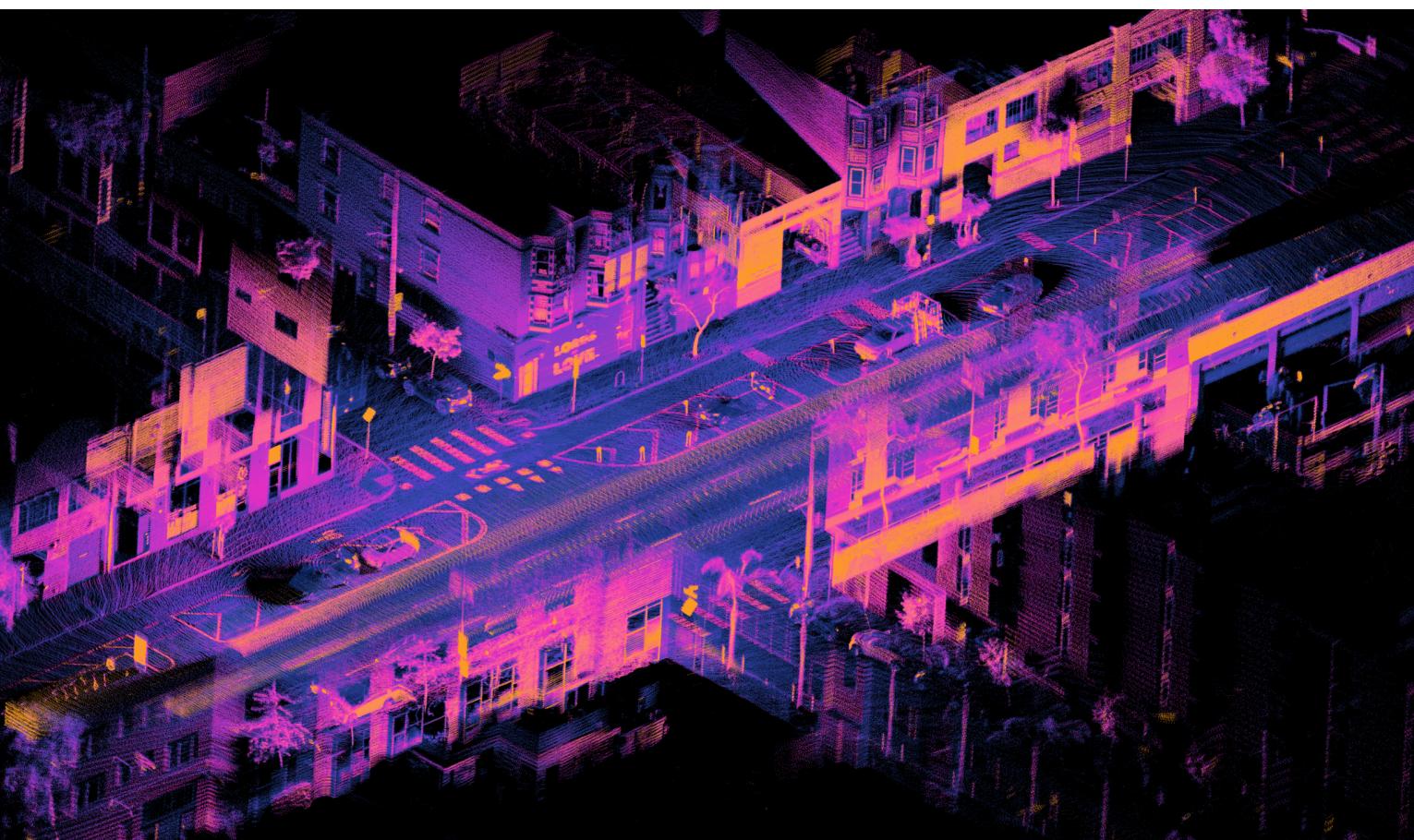
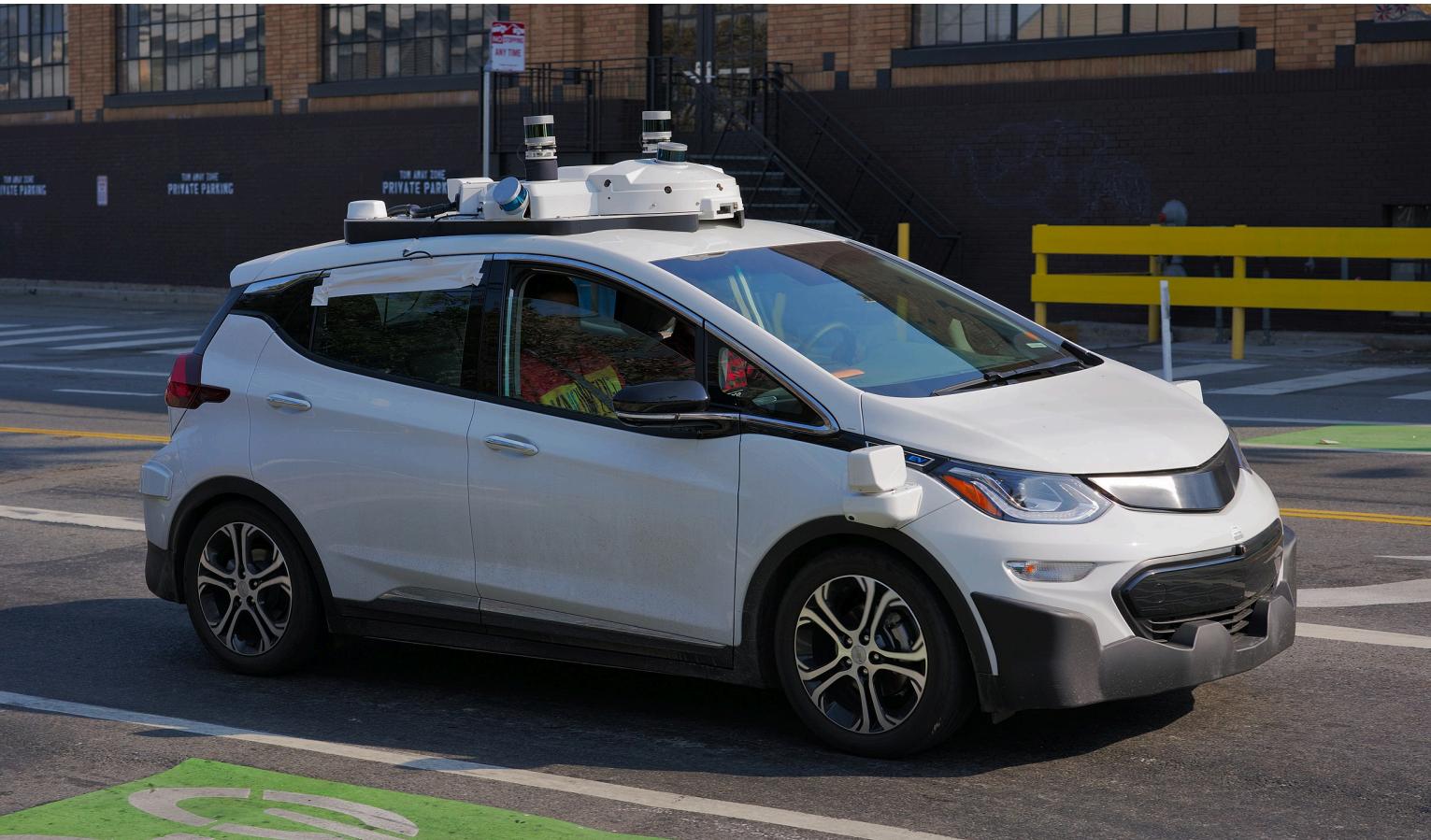
- Each element drawn x_i from a structured space \mathcal{X}
- Unordered $\{x_1, x_2\} = \{x_2, x_1\}$
- No duplicates
- Variable size

SETS

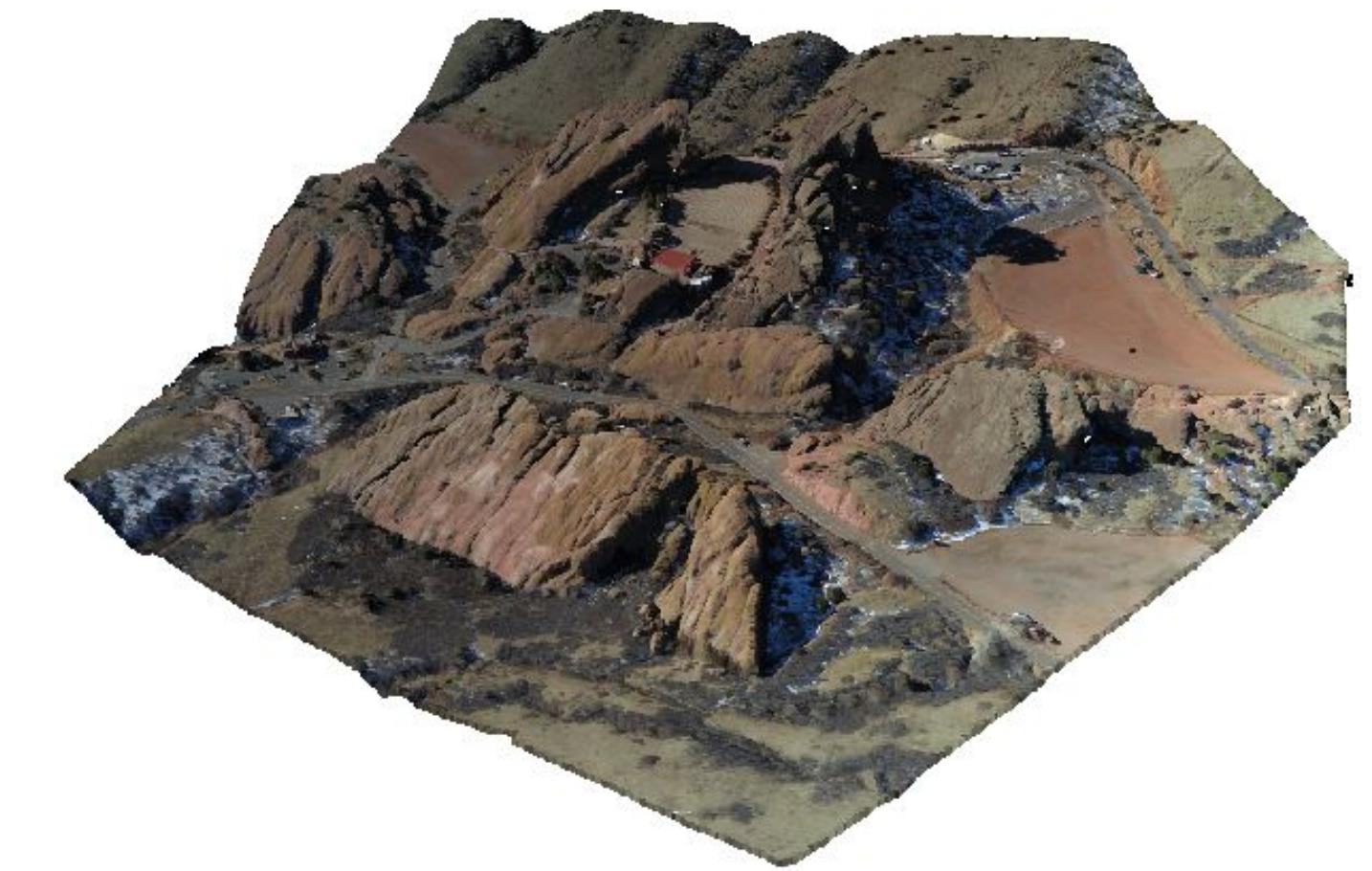
Examples:



LIDAR



Autonomous driving



Remote sensing

SEQUENCES

Definition: ordered list of elements where position carries meaning

$$\mathbf{x} = (x_1, x_2, \dots, x_T), \quad x_t \in \mathcal{X}$$

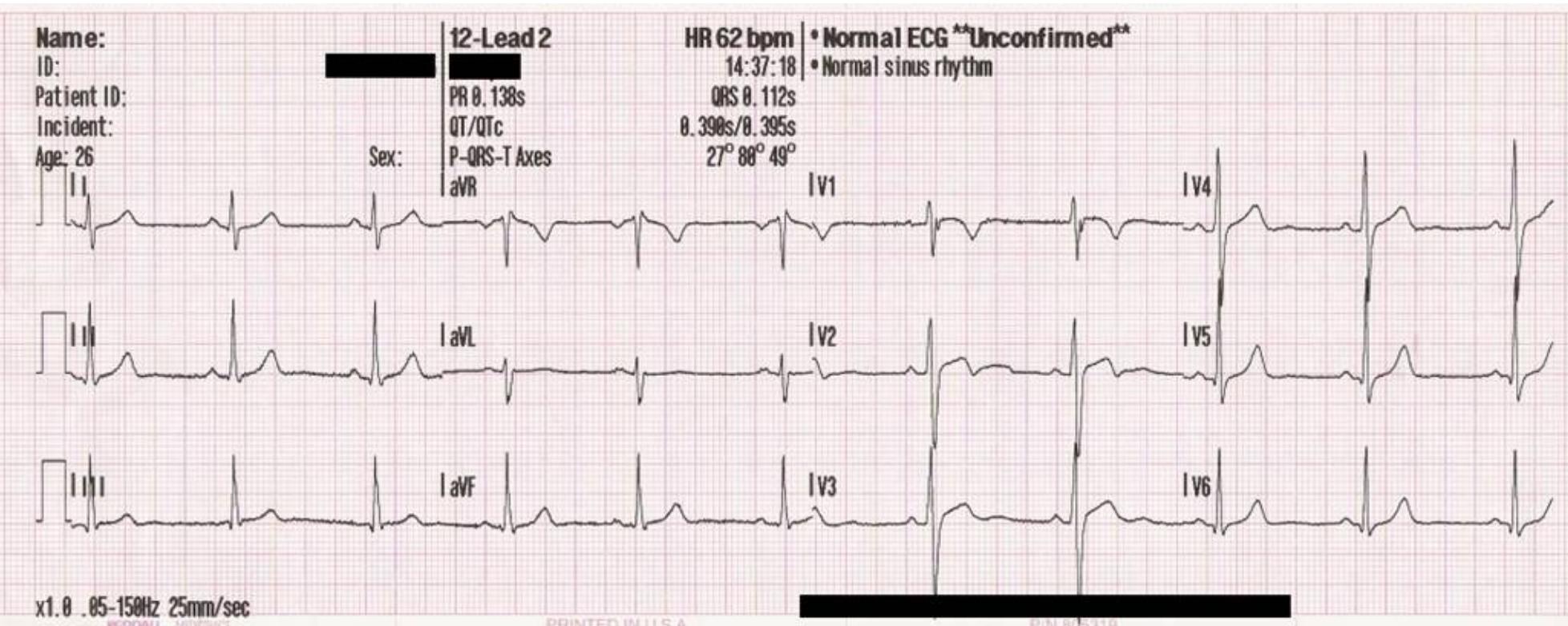
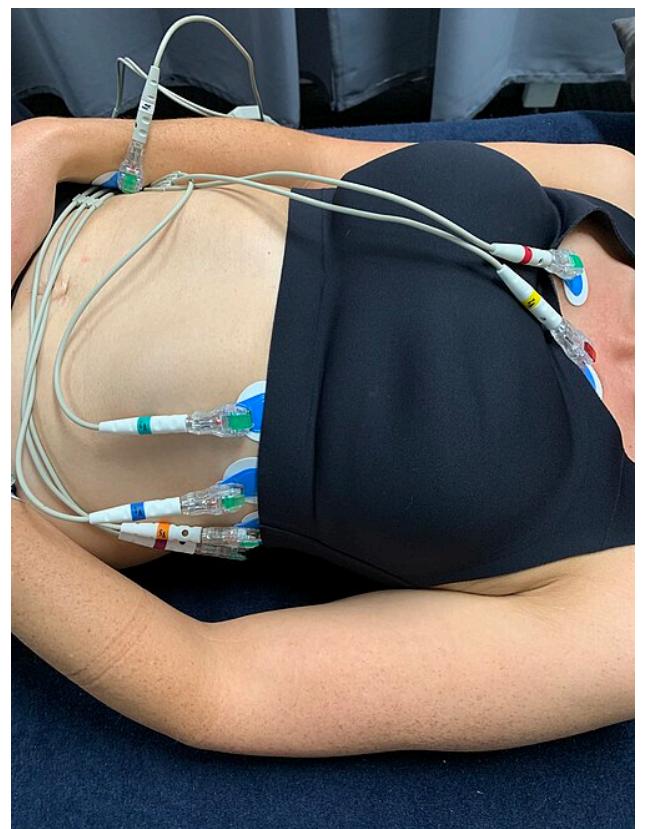
Key properties:

- Ordered $(x_1, x_2) \neq (x_2, x_1)$
- Variable length
- Difference in element index may define a distance
- Elements typically all come from same space

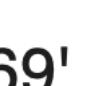
TIME SERIES

Special case of sequences where ordering has temporal meaning

Can be regularly sampled, irregularly sampled or event-driven



Electrocardiogram

Brazil		1–7	Germany
Oscar		90'	Report ↗
Müller		11'	
Klose		23'	
Kroos		24', 26'	
Khedira		29'	
Schürrle		69', 79'	

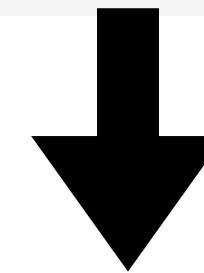
Soccer score sheet

TEXT

Sequences of characters/ words are mapped into structured token spaces

Many words map to one token, but some don't: indivisible.

Sequences of characters commonly found next to each other may be grouped together: 1234567890



```
[12488, 6391, 4014, 316, 1001, 6602, 11, 889, 1236, 4128, 25, 3862, 181386, 364, 168191, 328, 9862, 22378, 2491, 2613, 316, 2454, 1273, 1340, 413, 73263, 4717, 25, 220, 7633, 19354, 29338, 15]
```

TUPLES

Definition: A tuple is an *ordered collection* of heterogeneous elements

$$\mathbf{x} = (x_1, x_2, \dots, x_n), \quad x_i \in \mathcal{X}_i$$

Key properties:

- Each element x_i belongs to a different space
- Ordered $(x_1, x_2) \neq (x_2, x_1)$
- Fixed size

TUPLES

Patient ID	Age (years)	Blood Pressure (mmHg)	Smoking Status (Category)	Doctor's Notes (Free Text)
1	45	120/80	Former Smoker	"Patient shows good recovery, recommend exercise."
2	63	135/90	Current Smoker	"Monitor blood pressure closely, risk of stroke."
3	29	110/70	Never Smoked	"Healthy, no major concerns."

Electronic health records

GRAPHS

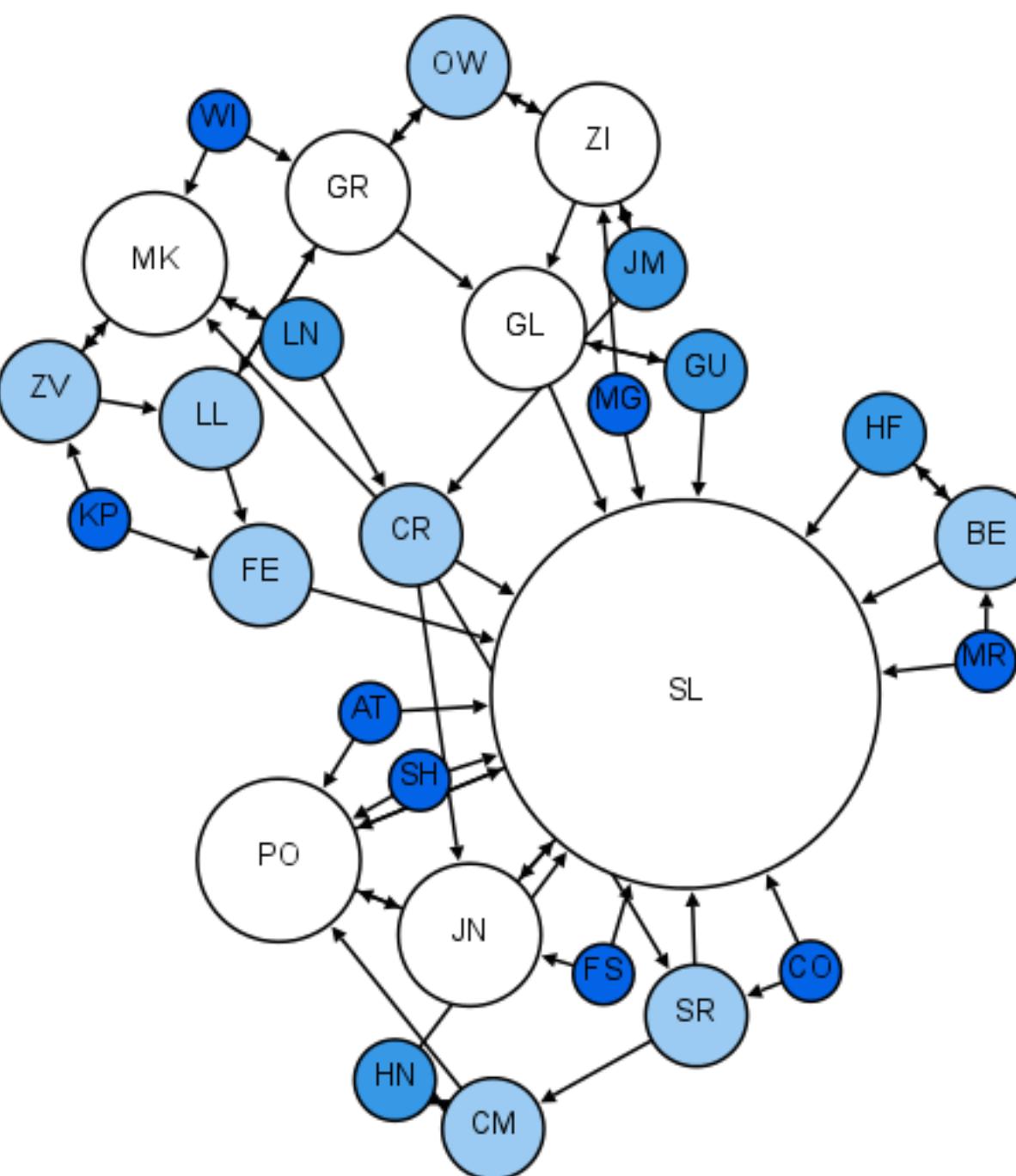
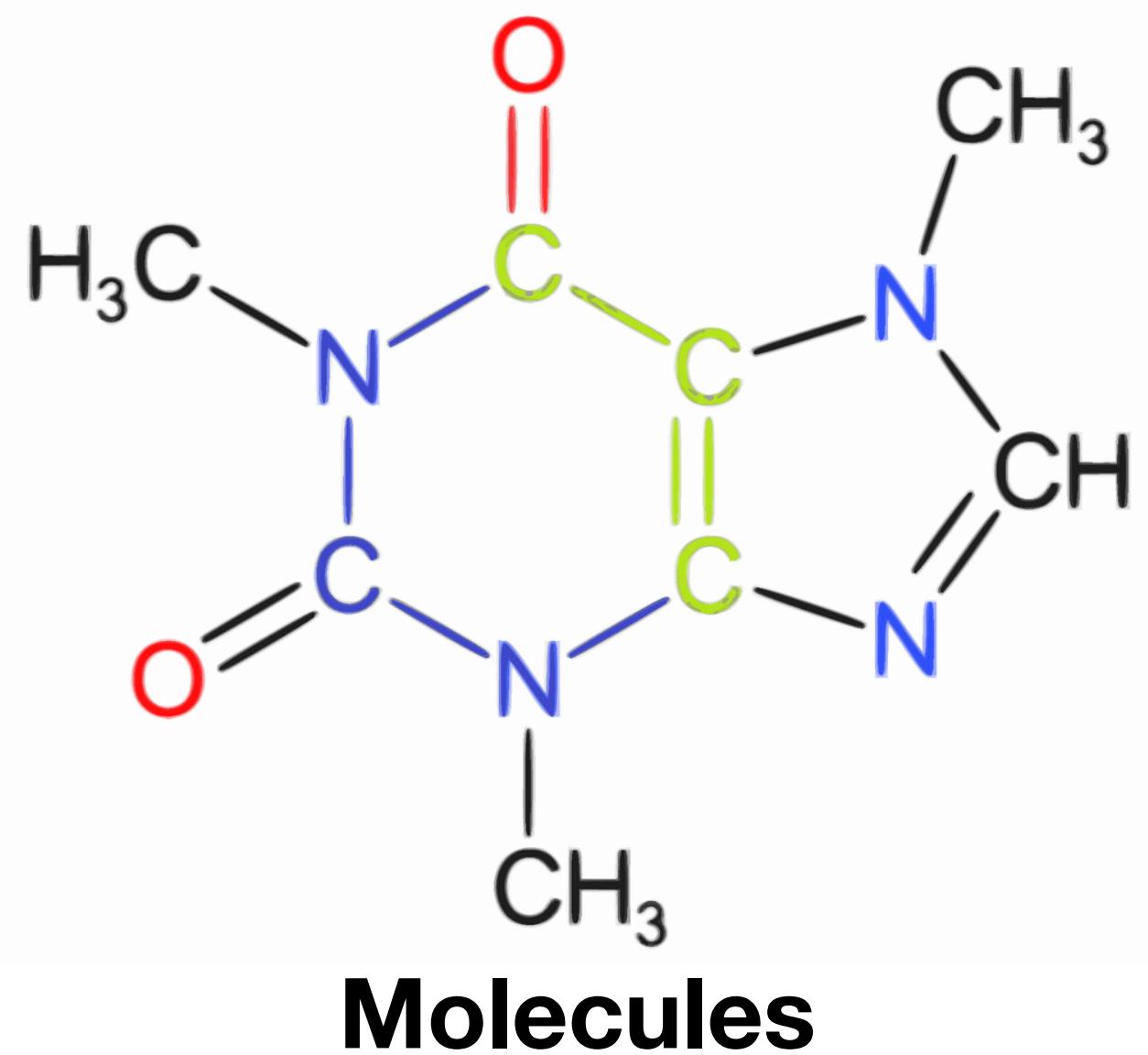
Definition: A graph is a set of nodes connected by edges

$$\mathcal{G} = (\mathcal{V}, \mathcal{E})$$

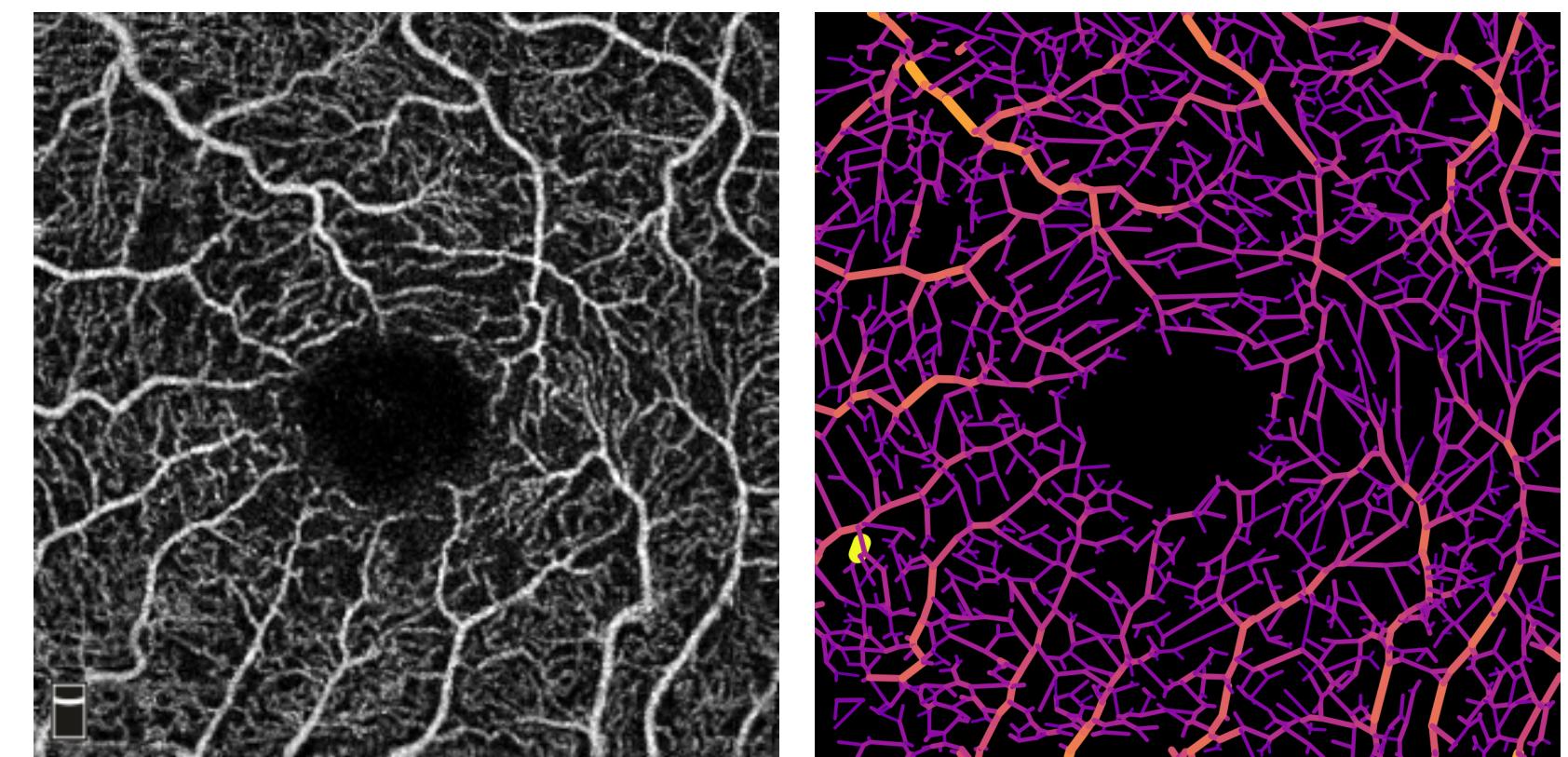
Key properties:

- Set of nodes $v_i \in \mathcal{V}$ that can also have an associated feature vector
- Each edge connects two nodes $(v_i, v_j) \in \mathcal{E}$ and can also have attributes
- Edges can have a direction or be symmetric

GRAPHS



Social graphs

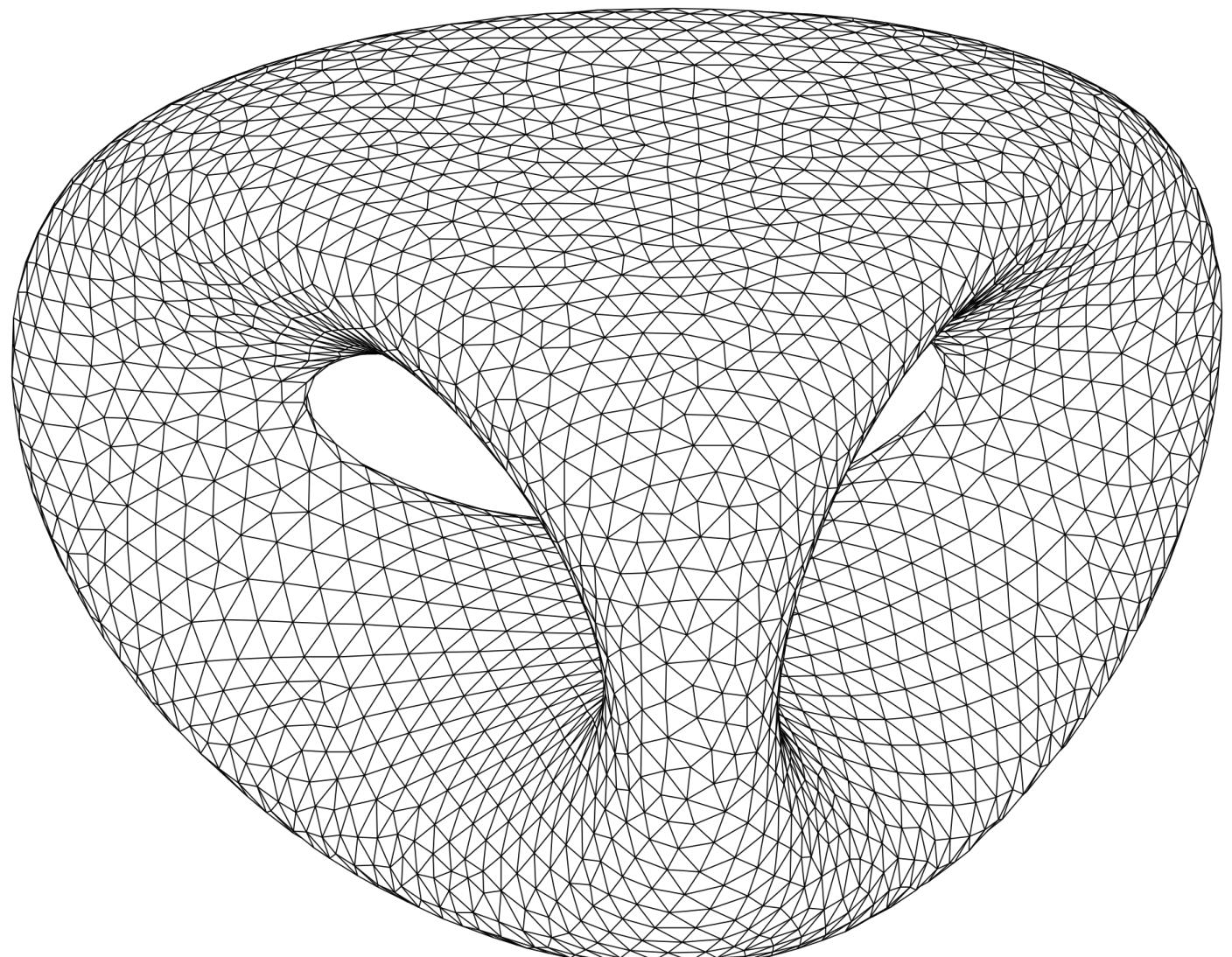


Retinal blood vessel graph

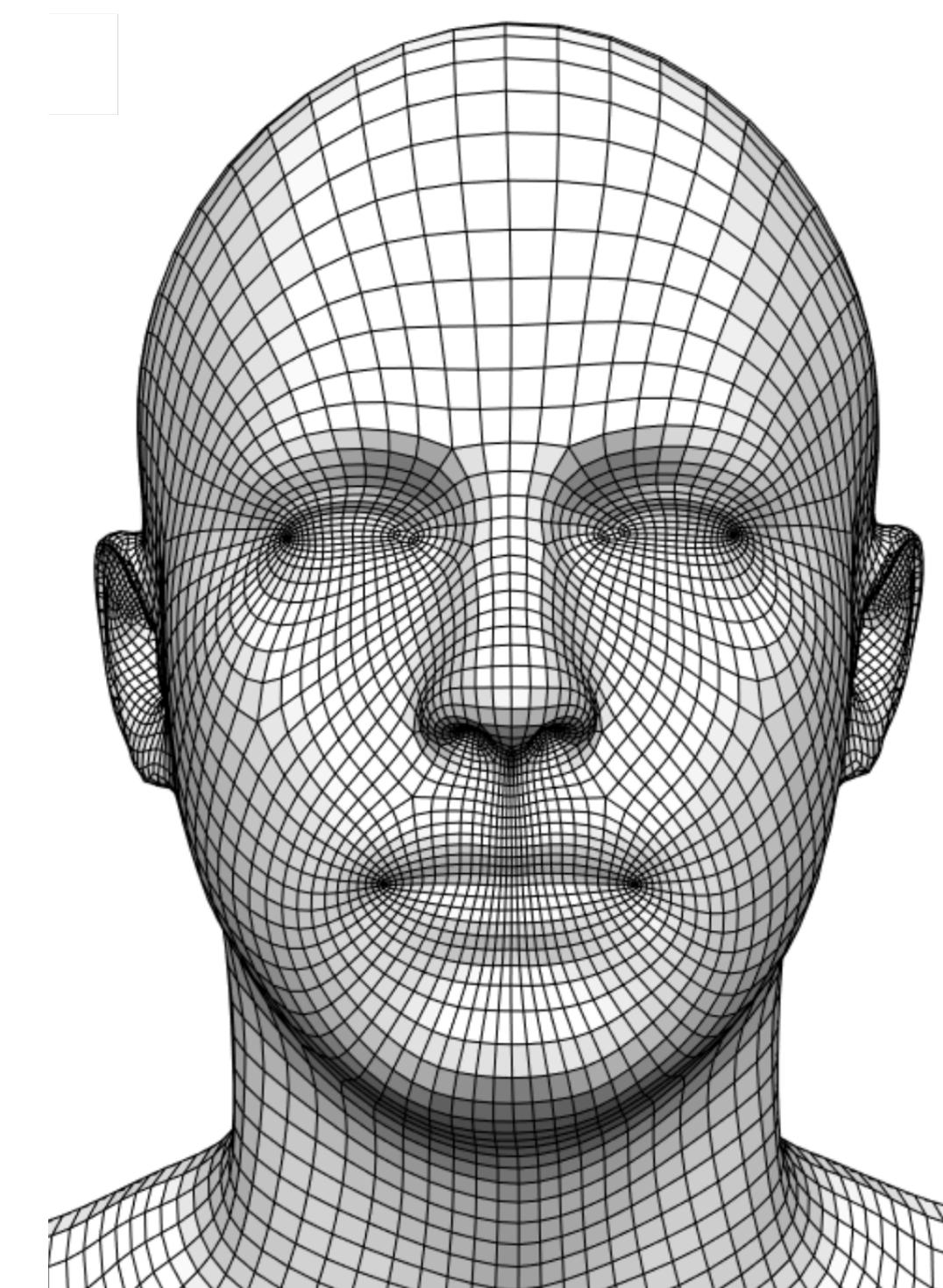
MESHS

Special graph where nodes and edges are embedded in 3D space

Usually with geometric and topological consistency



Triangle mesh



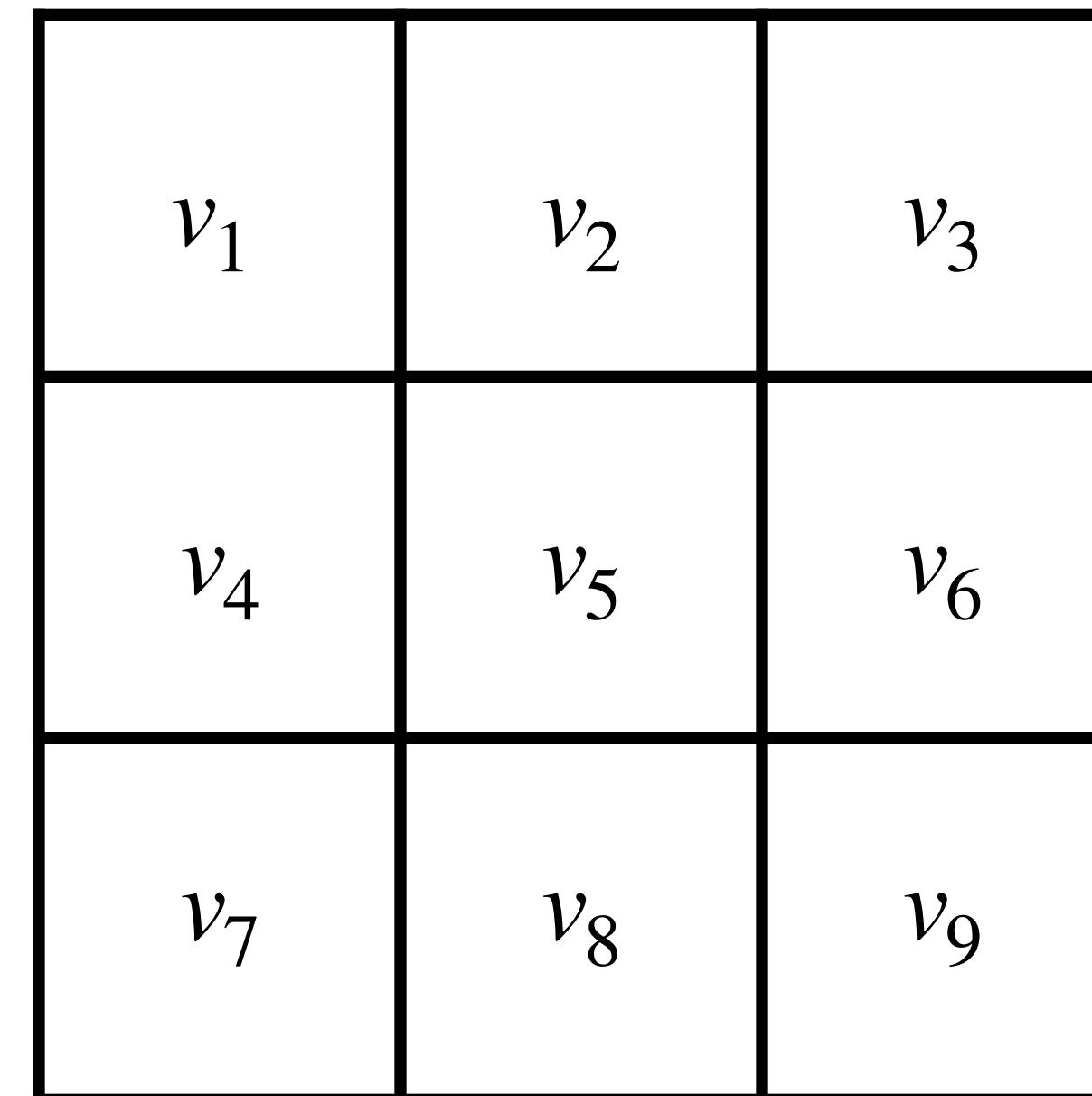
Quadrilateral facial mesh

IMAGES

Can be seen as a multidimensional case of a **sequence**:

$x_{i-1,j-1}$	$x_{i-1,j}$	$x_{i-1,j+1}$
$x_{i,j-1}$	$x_{i,j}$	$x_{i,j+1}$
$x_{i+1,j-1}$	$x_{i+1,j}$	$x_{i+1,j+1}$

Can also be viewed as special **graph** with pixels corresponding to vertices and edges representing the neighborhood:

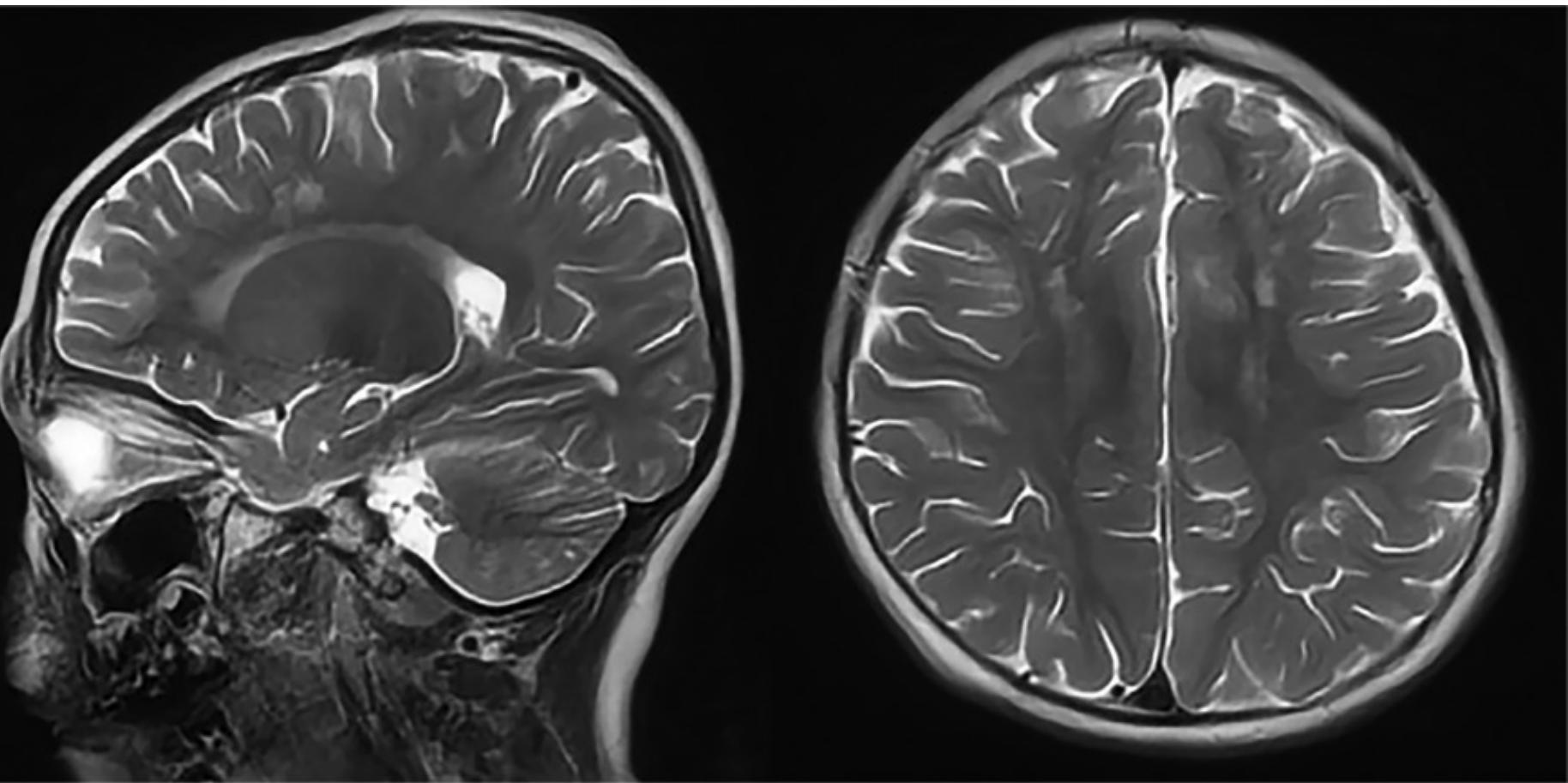


IMAGES

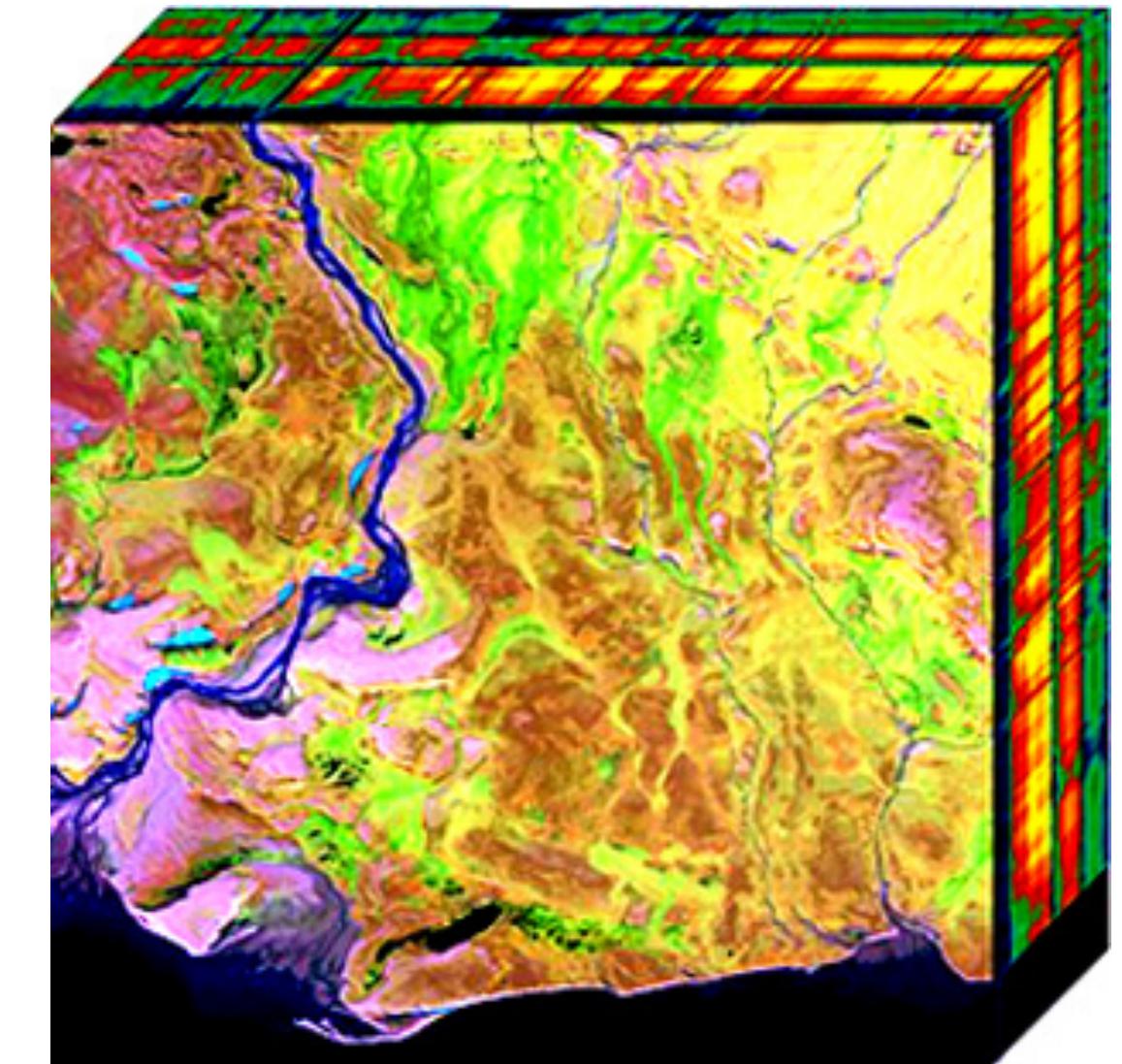
Examples:



Color photograph



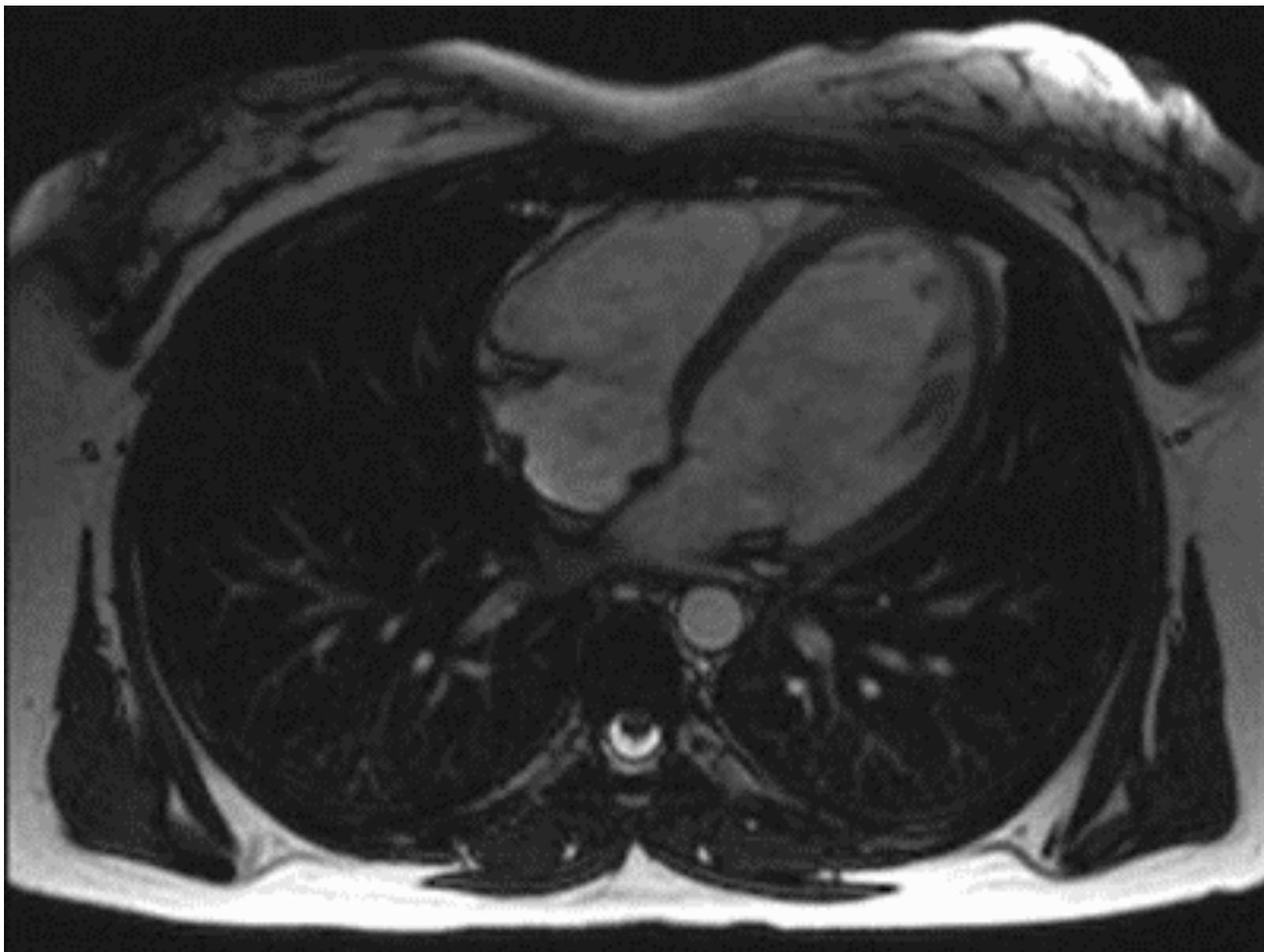
Volumetric magnetic resonance
tomography image



Hyperspectral satellite image

VIDEOS

Images nested inside a sequence



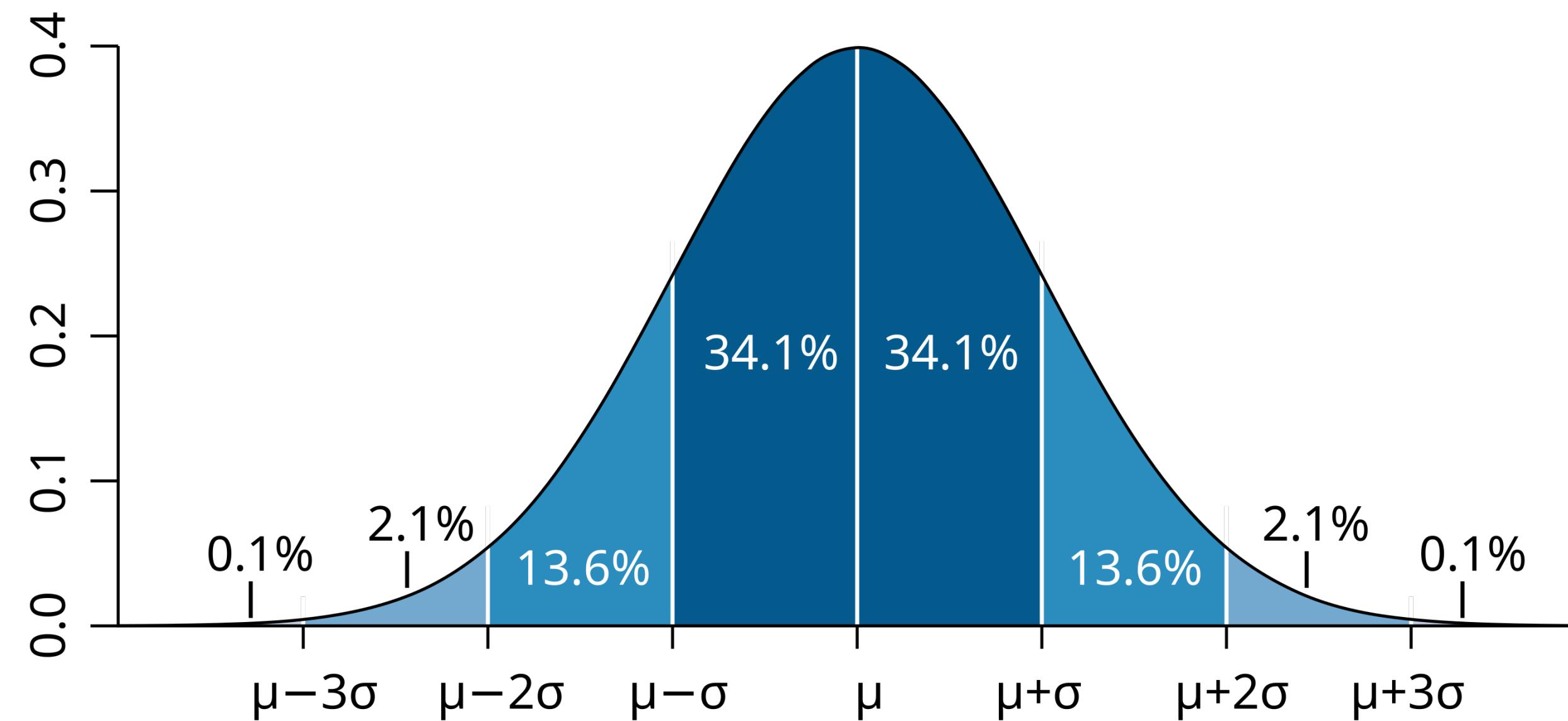
Cine cardiac MR image



Egocentric drone video

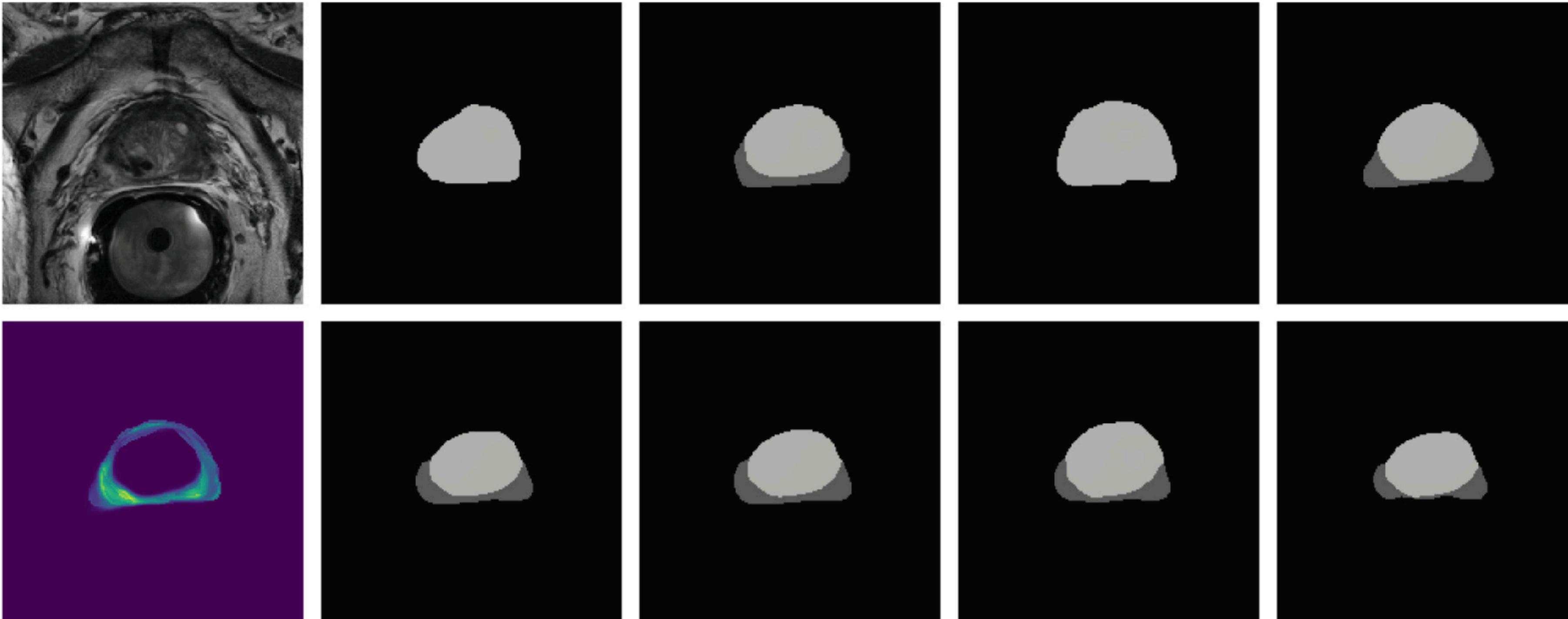
DISTRIBUTIONS

Elements of any data structure can also be parameterized and processed as statistical distribution



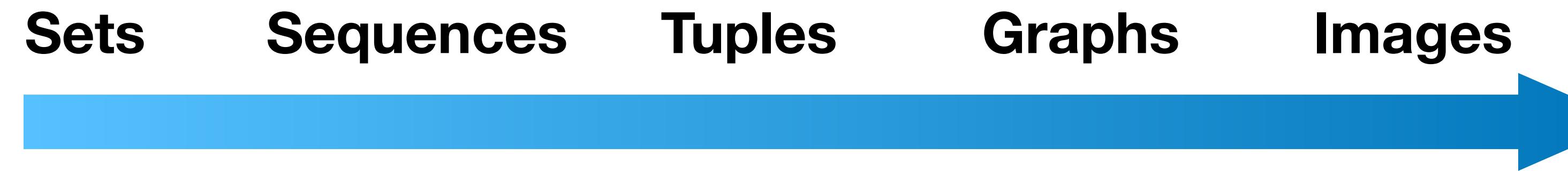
μ_1, σ_1	μ_2, σ_2	μ_3, σ_3
μ_4, σ_4	μ_5, σ_5	μ_6, σ_6
μ_7, σ_7	μ_8, σ_8	μ_9, σ_9

DISTRIBUTIONS

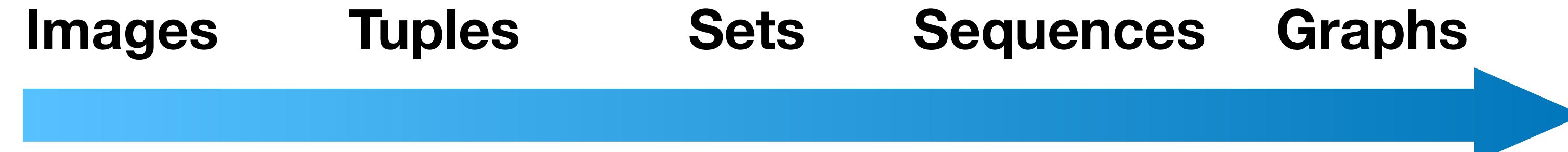


Probabilistic prostate segmentation

(ATTEMPT AT) ORDERING DATA STRUCTURES



More structured



Difficult to implement

SUMMARY

- Data can be found in diverse structures
- Shaped by data structures and its contained elements
- Choice of data highlights different aspects of the underlying object
- In future lectures: choice of data structure informs its:
 - Encoding
 - Processing
 - Combination with other data

OBJECT, MEASUREMENT, DATA

Definition from Wikipedia:

"Data are a collection of discrete or continuous values that convey information, describing the quantity, quality, fact, statistics, [or] other basic units of meaning, [...] that may be further interpreted formally."

"Data are *collected* using techniques such as *measurement*, observation, query, or analysis."

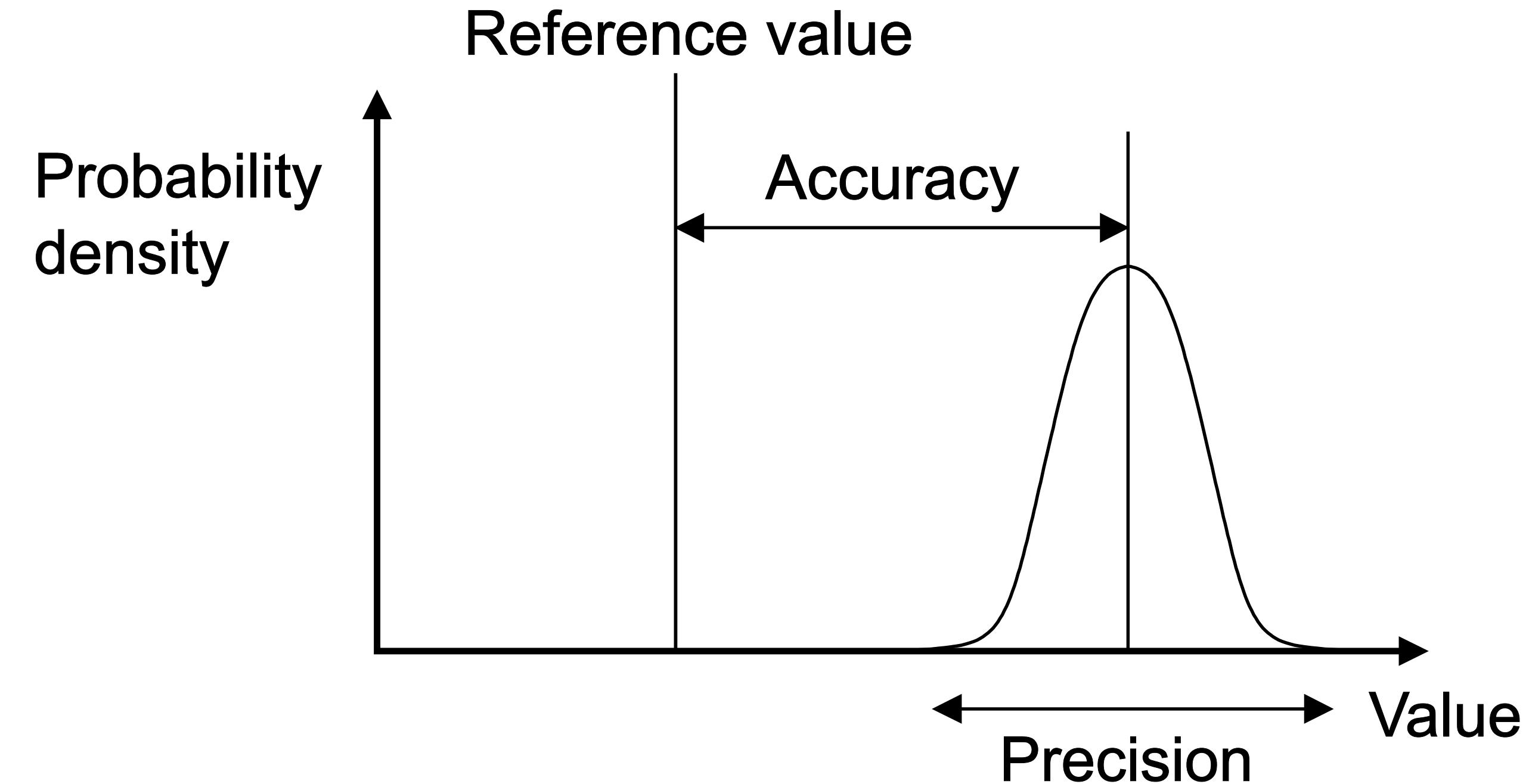
The data acquisition process

MEASUREMENT

Measurement is the quantification of attributes of an object or event.

It should be:

- Accurate
- Precise
- Unbiased
- Reproducible

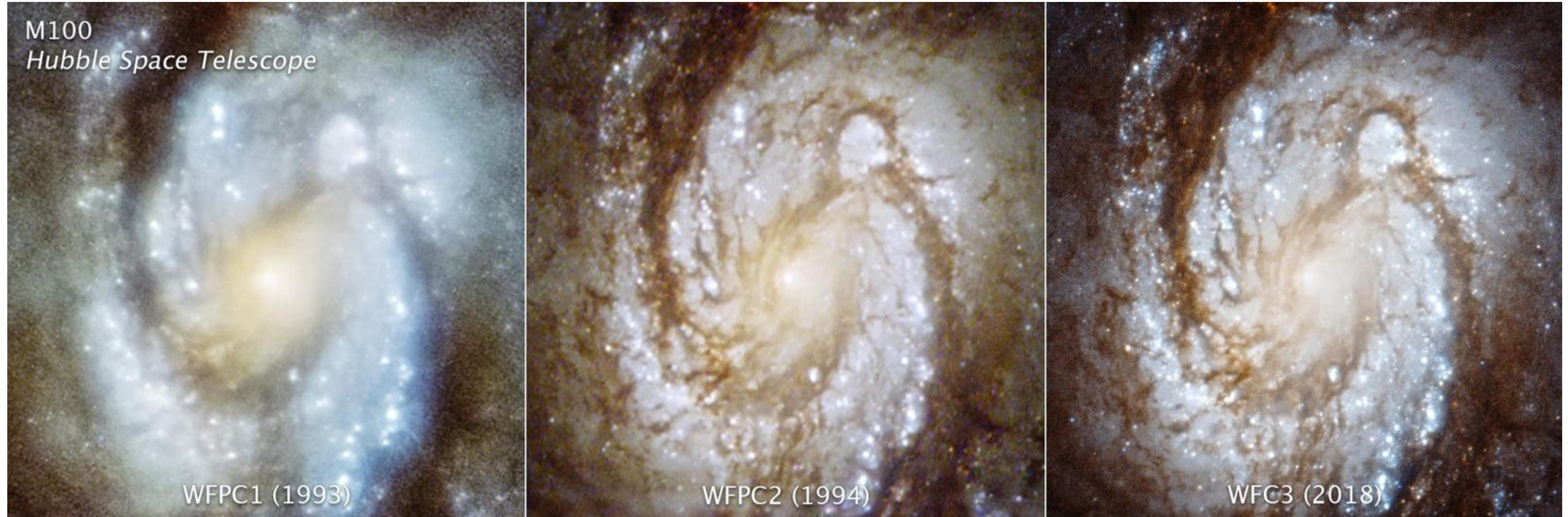


BETTER MEASUREMENT SYSTEMS ACQUIRE BETTER DATA

Data quality is often dependent on equipment



**Galileo Galilei's drawing
of the sun (1610)**



**Messier 100 galaxy, 166,000 light years from earth,
imaged by the Hubble Space Telescope**

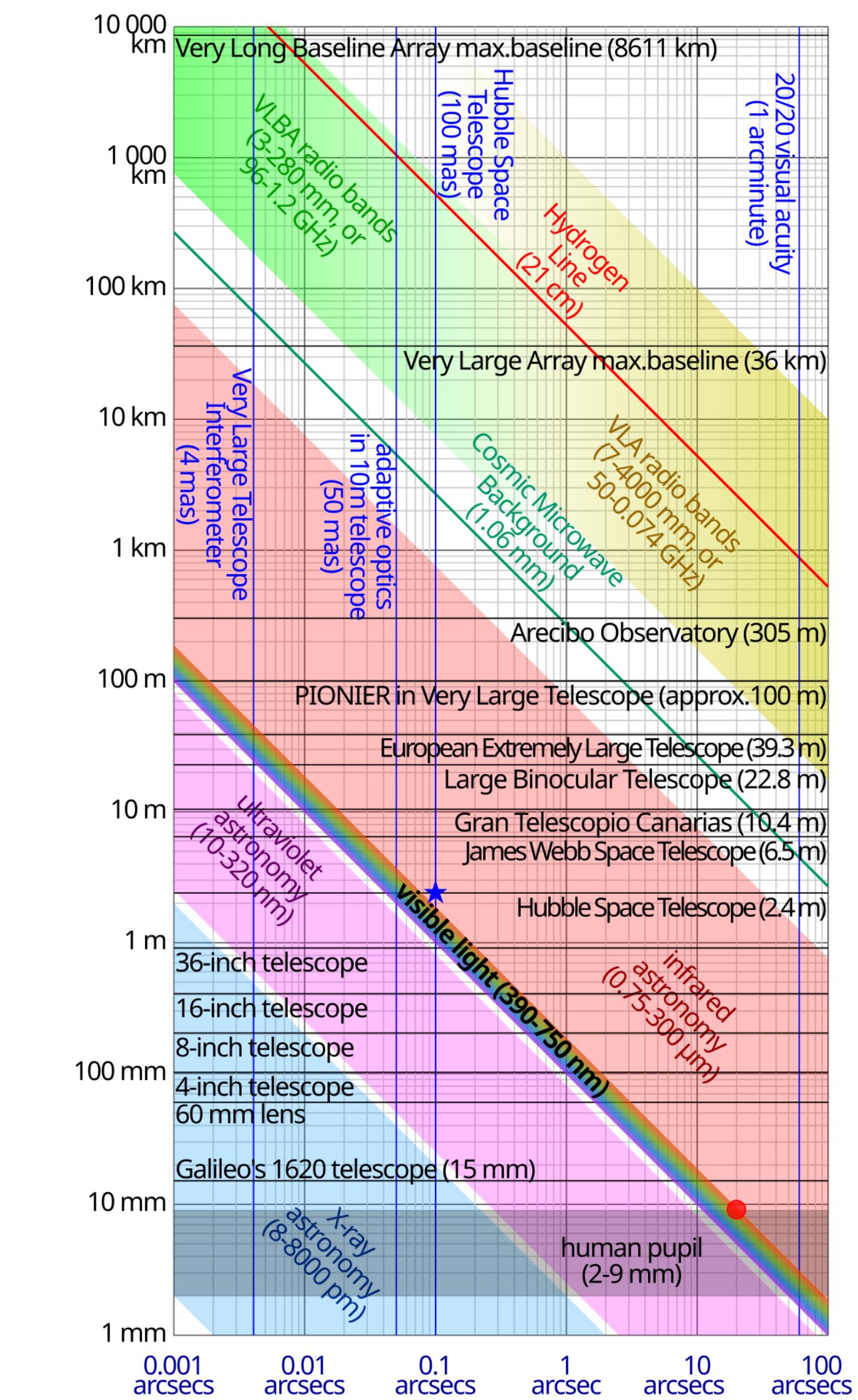
PHYSICAL LIMITS OF MEASUREMENT SYSTEMS

Resolution of optical systems is restricted by the diffraction limit:

$$d \approx \frac{\lambda}{2 \text{NA}}$$

λ = wavelength of light

NA = numerical aperture of the lens

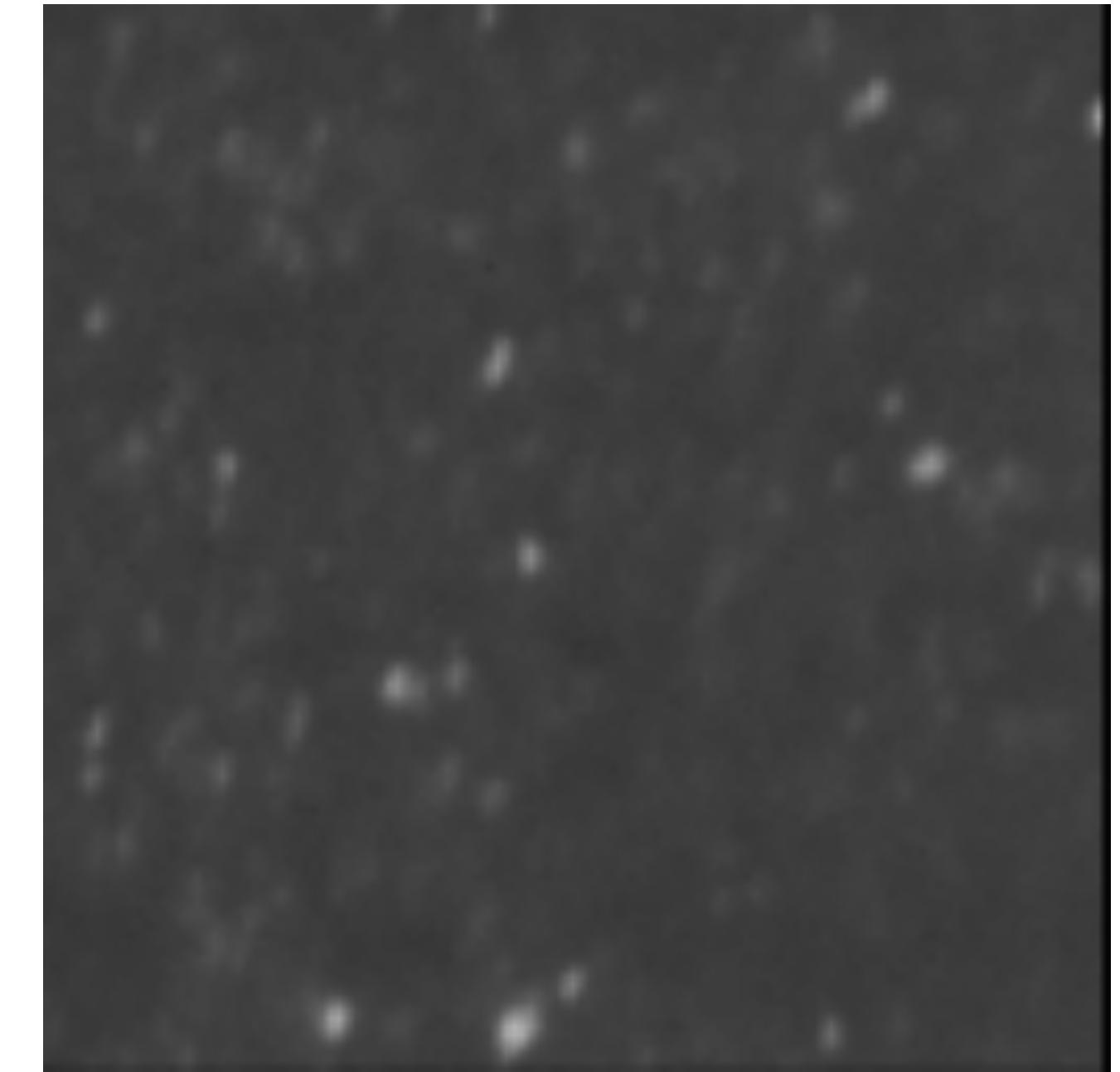


**Diffraction limit
at various wavelengths**

PHYSICAL LIMITS OF MEASUREMENT SYSTEMS

All electronic systems have a noise floor caused by:

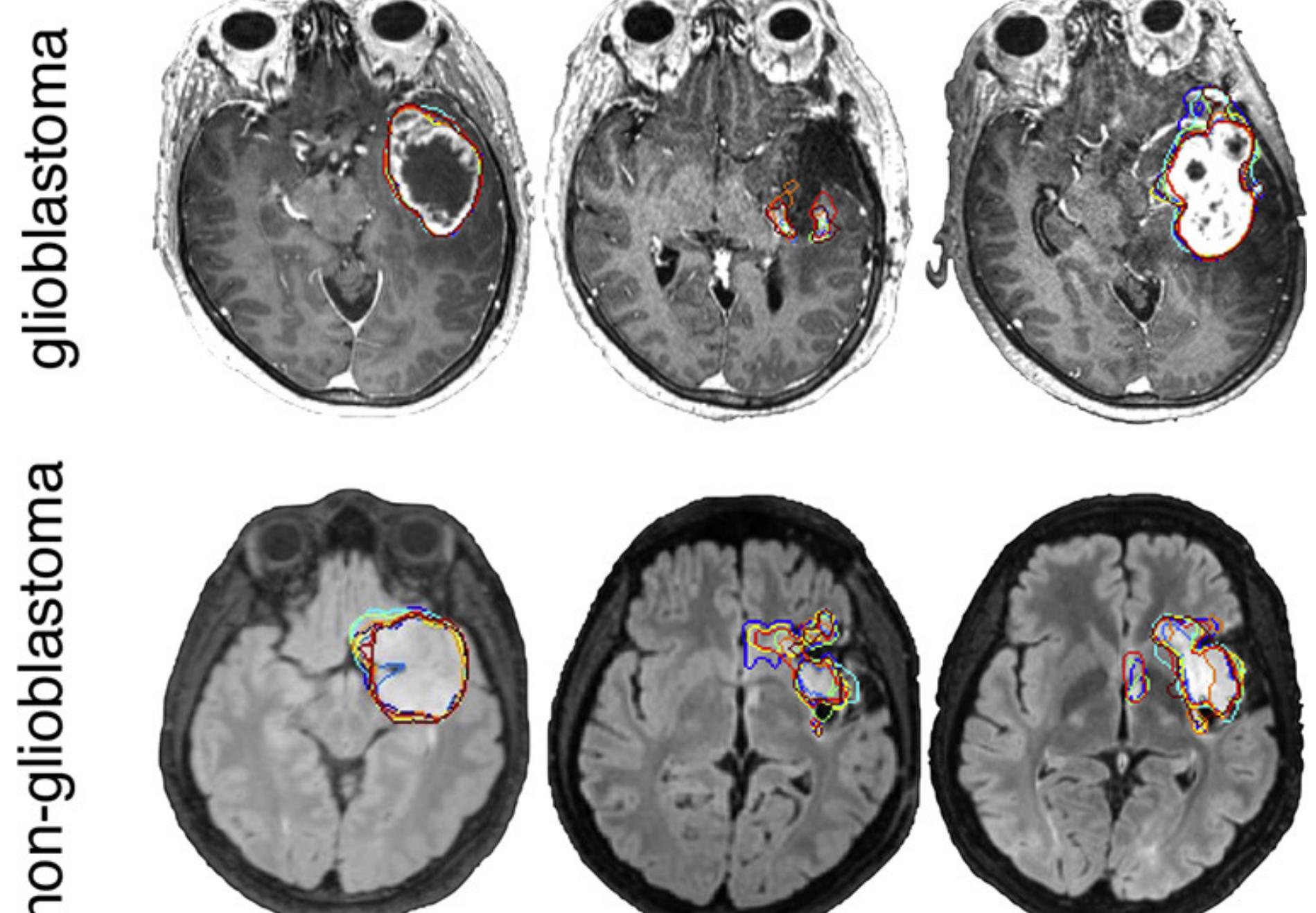
- Thermal noise
- Atmospheric noise
- Cosmic background noise



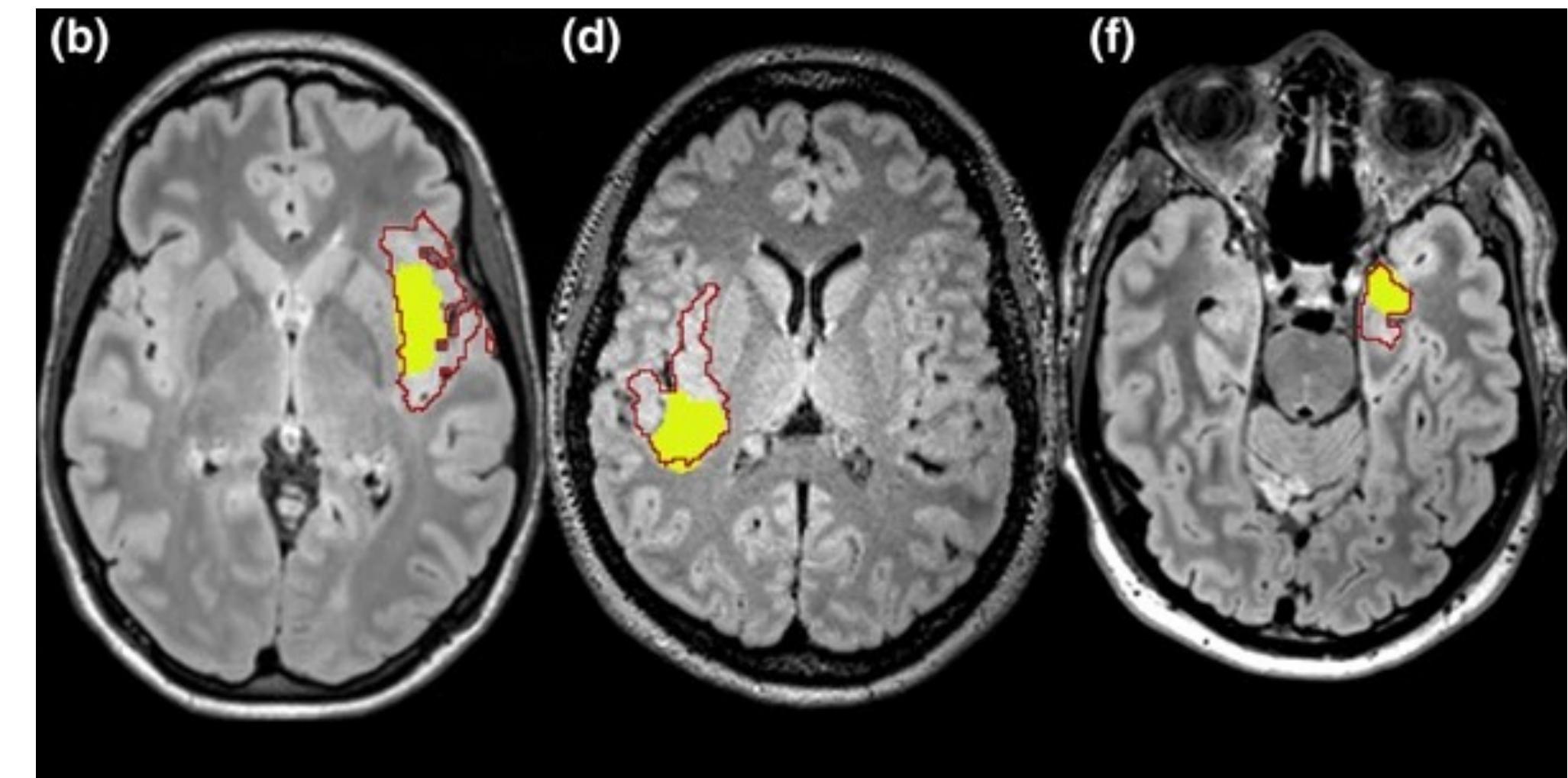
Brownian noise of small particles

PRACTICAL LIMITS OF DATA ACQUISITION

Data annotation quality is often limited by human ability



Inter-rater segmentation variability



Intra-rater segmentation variability

REAL-WORLD TRADE-OFFS

The choice of data acquisition is often decided not only by data quality, but by cost, safety and feasibility considerations.



Military reconnaissance aircraft

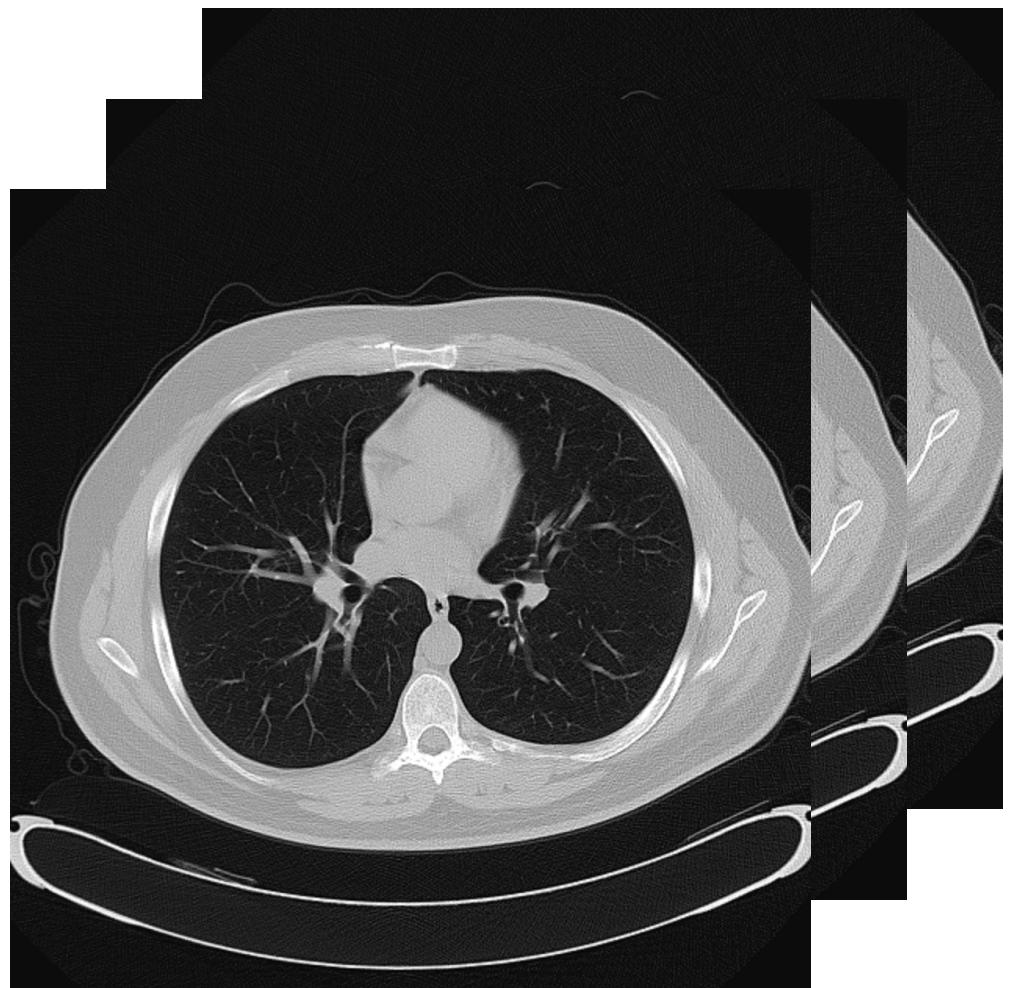


Consumer-grade aerial drone

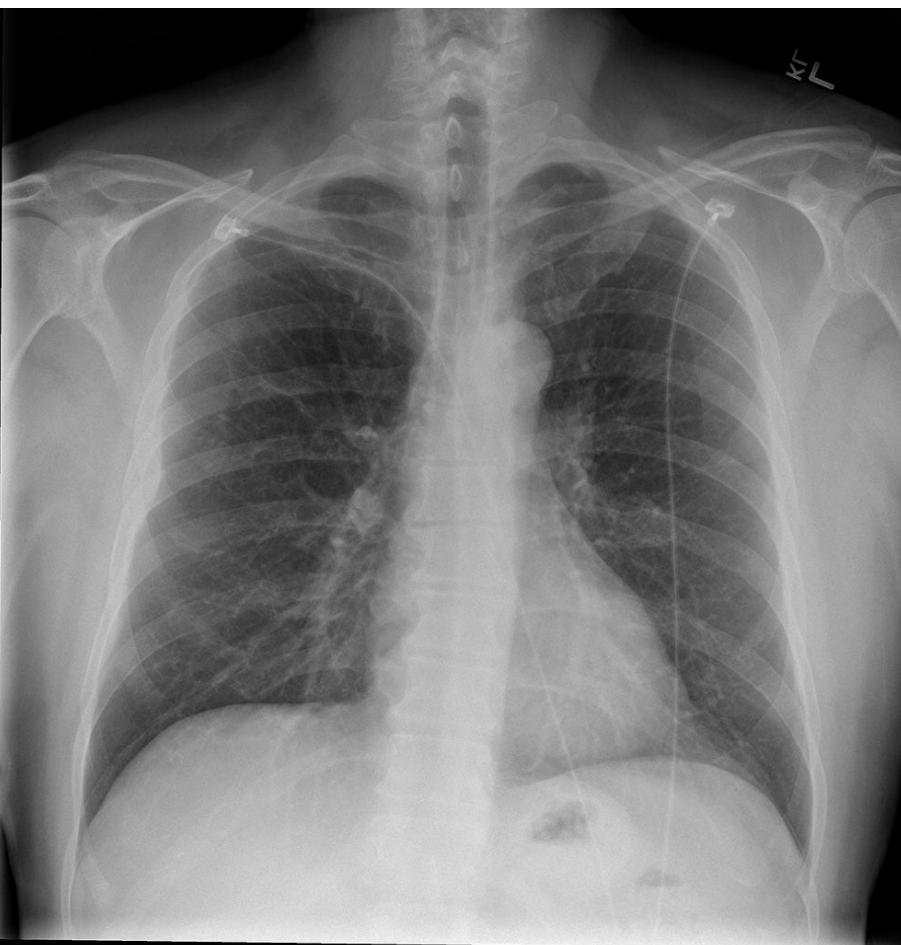


REAL-WORLD TRADE-OFFS

The choice of data acquisition is often decided not only by data quality, but by cost, safety and feasibility considerations.



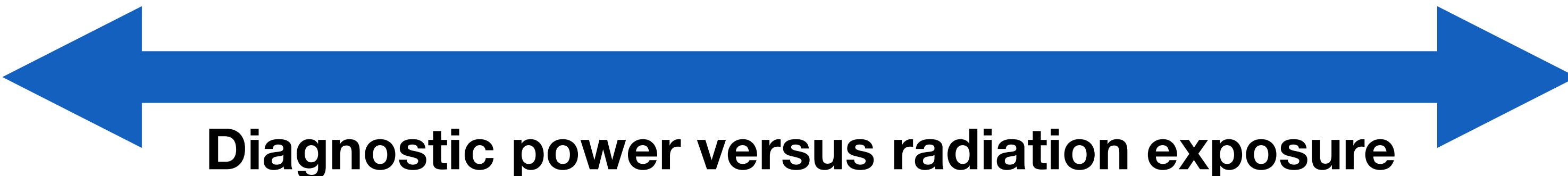
Volumetric thorax CT



Chest x-ray



Stethoscope

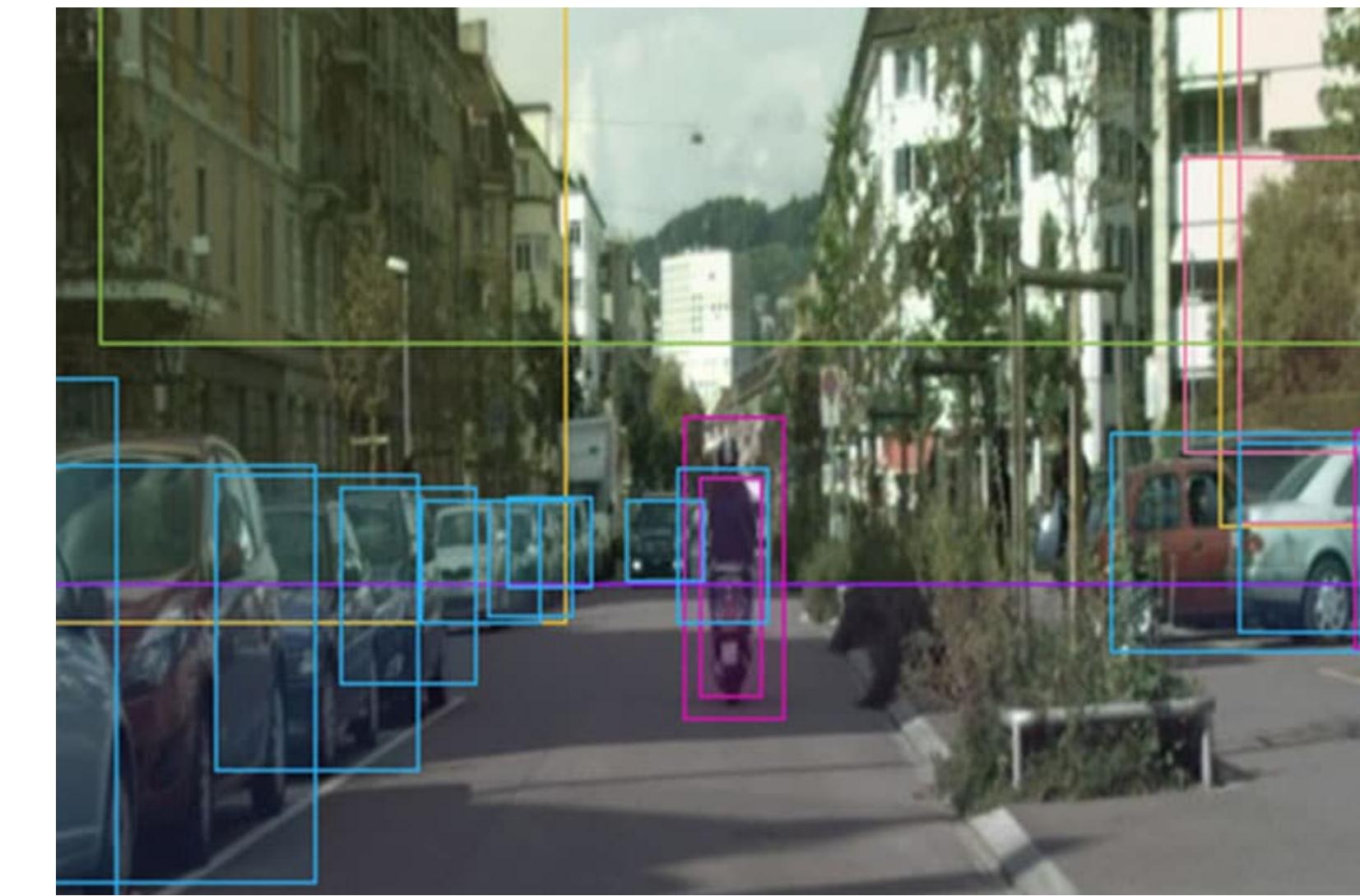


REAL-WORLD TRADE-OFFS

The choice of data acquisition is often decided not only by data quality, but by cost, safety and feasibility considerations.



Semantic segmentation masks

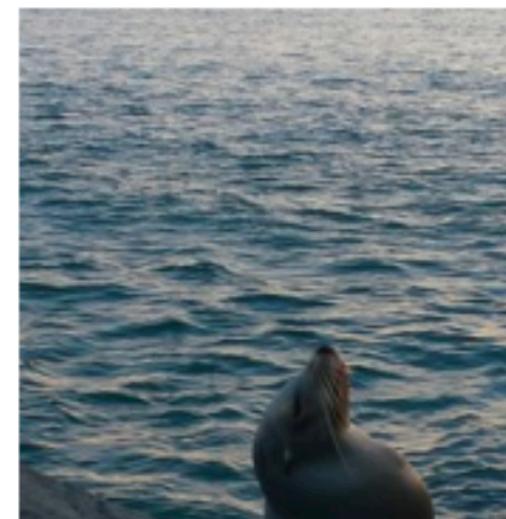
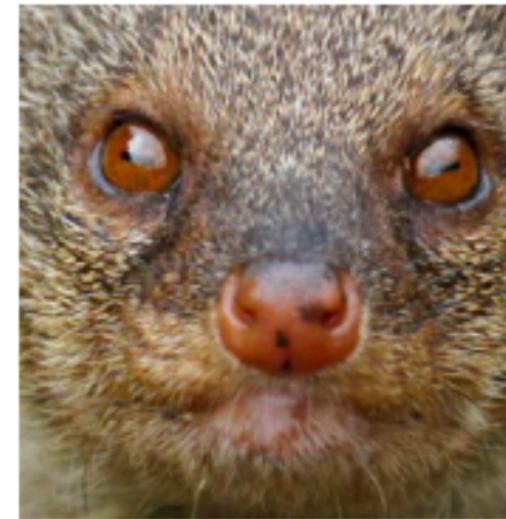
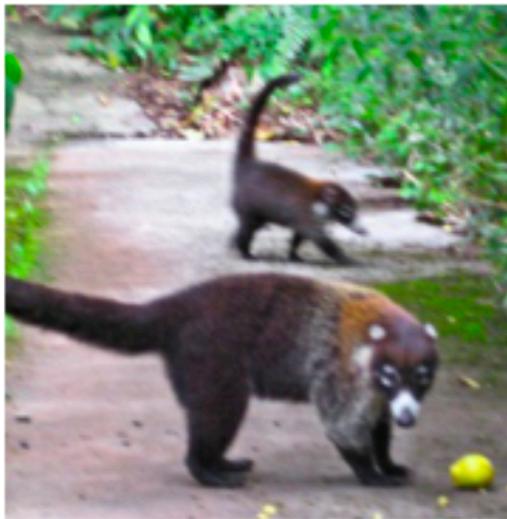


Bounding box annotations



DATA STANDARDIZATION

Data may be acquired using different measurement equipment, leading to a lack of data standardization



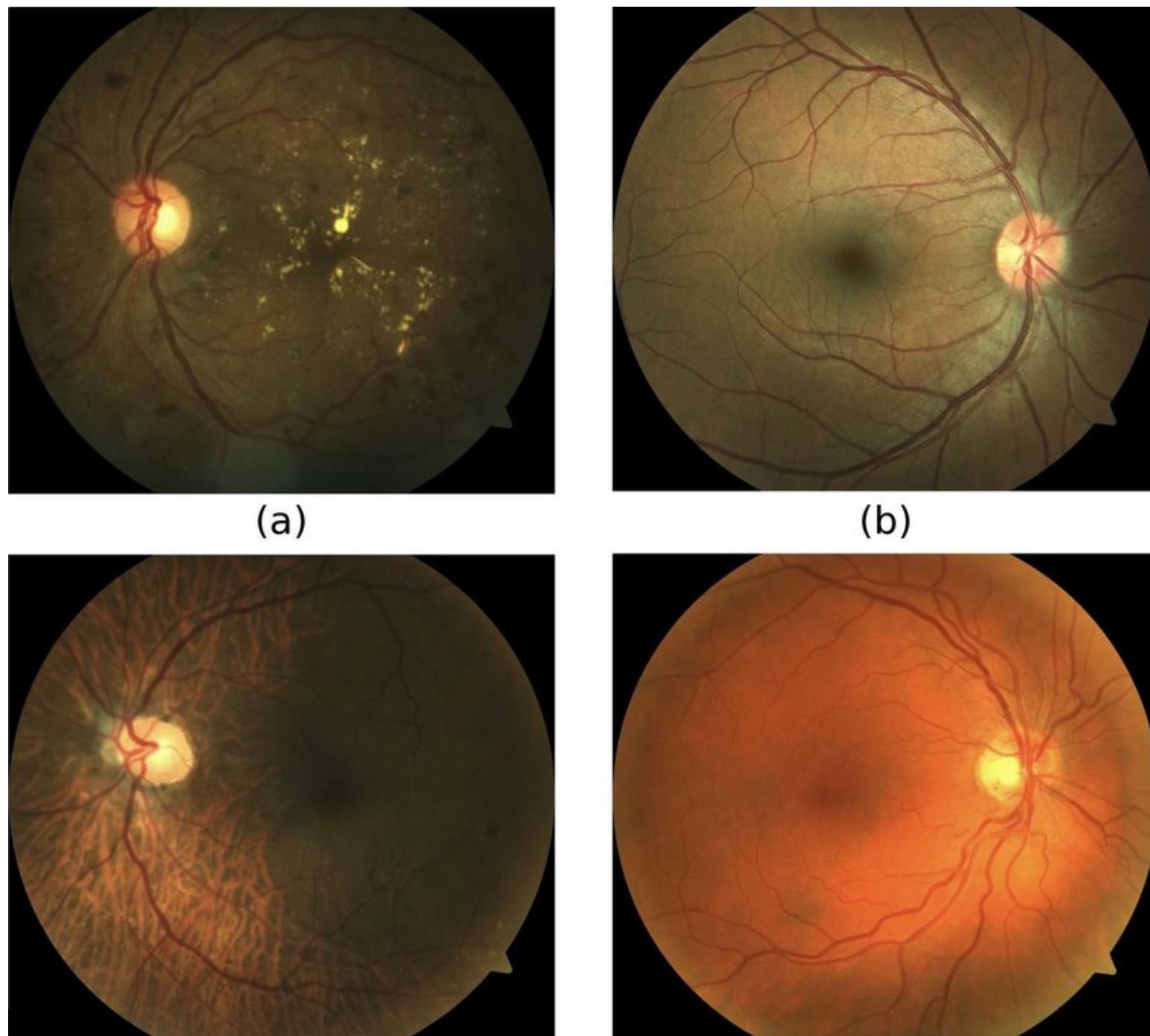
Natural image datasets are typically collected from various smart phone photographs

DATA BIAS

Sampling bias may result in domain shift between acquired (multimodal) data and real-world object



Quality difference in natural images



Retinal images of patients
of different ethnicities

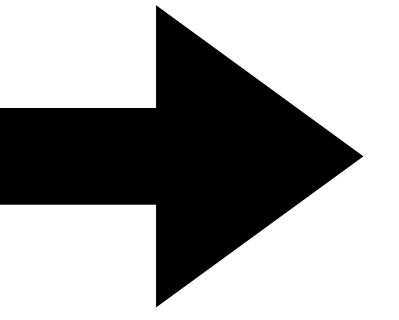
MULTI-STEP DATA ACQUISITION

Many data acquisition processes involve multiple steps

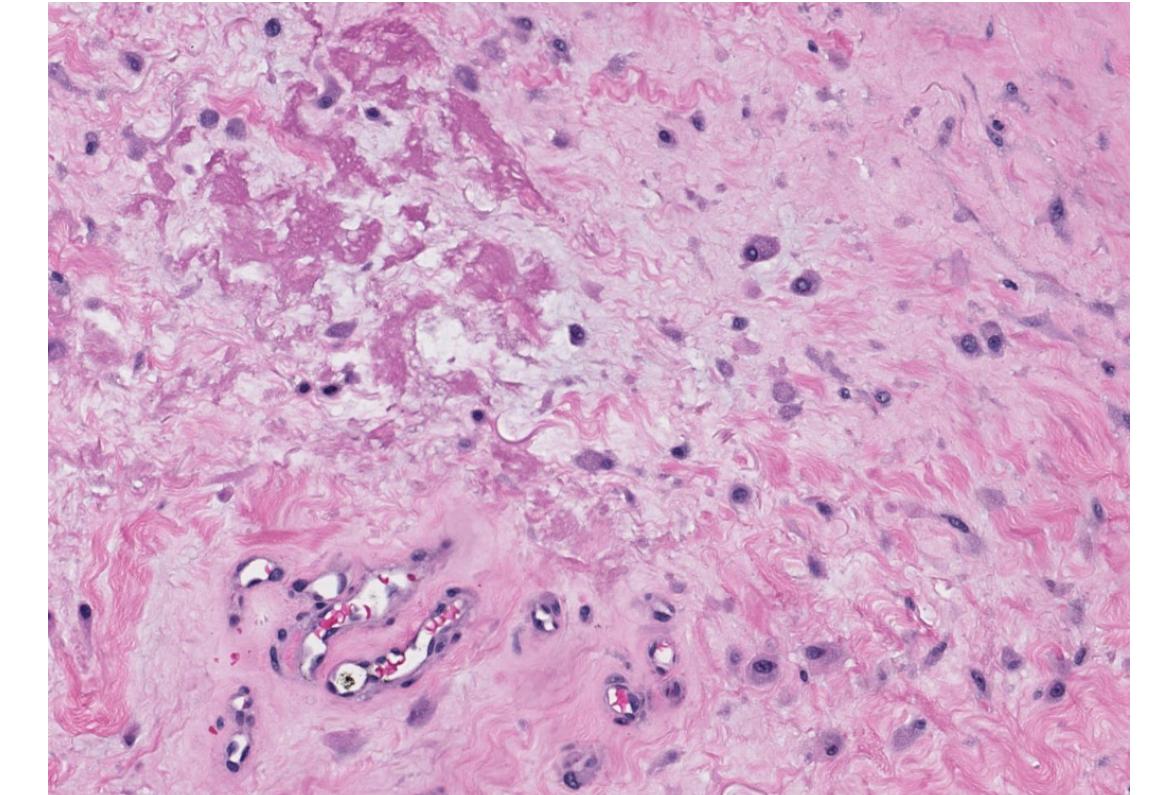
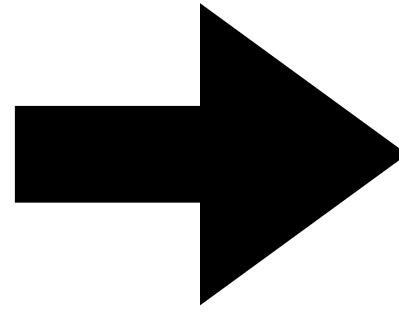
Individual steps may involve deep learning



Biopsy



Sample preparation

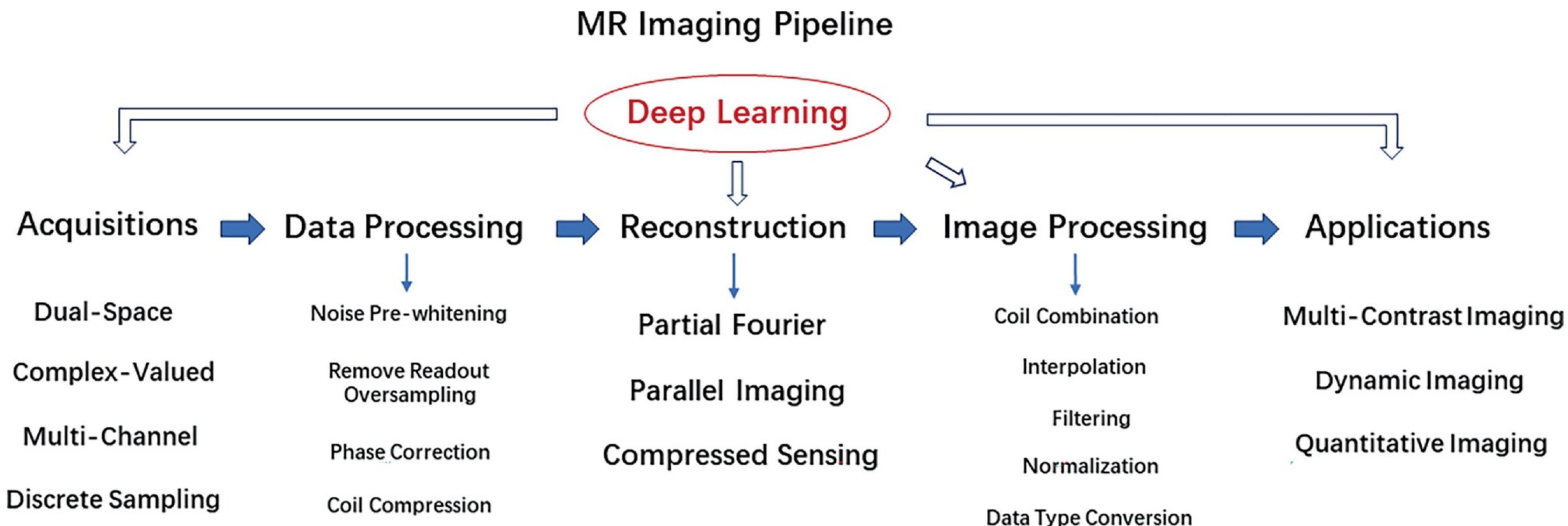


Histopathology images

MULTI-STEP DATA ACQUISITION

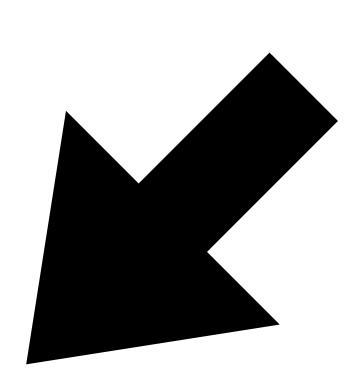
Many data acquisition processes involve multiple steps

Individual steps may involve deep learning



ASYNCHRONOUS MULTIMODAL DATA

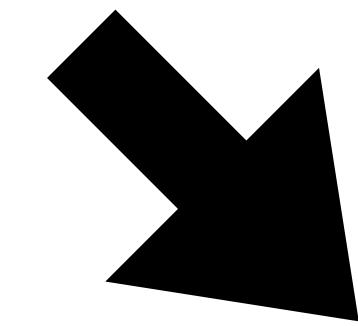
Even when concerning the same object, multimodal data can be asynchronous



LIDAR @ 10 Hz



Radar @ irregular 2 Hz



Video @ 25 Hz

MISSING MODALITIES

In most (real-world) datasets only a limited amount of modalities is available for each sample

Patient ID	Age	MR Image	Laboratory Tests	Patient Interview
P001	68	/data/mri/ P001.nii	WBC: 7.2, Glucose: 102 mg/dL	"Patient reports occasional
P002	74		WBC: 6.8, Glucose: 95 mg/dL	"Complains of persistent headache."
P003	59	/data/mri/ P003.nii		"Denies any recent symptoms."
P004	82	/data/mri/ P004.nii	WBC: 8.1, Glucose: 110 mg/dL	
P005	65			

SUMMARY

Data has usually been acquired before deep learning practitioners get involved

Still, thinking about the underlying data generation process is important:

- Is the data suited for my application?
- Is the data representative or biased?
- Is the data represented in the most suitable format?

These questions inform how the data will be further processed

FINAL THOUGHTS

Unimodal data is more diverse than you may think

Multimodal is more similar than you may think