Week 11 Lecture 11
AIID + Image

Wan Fang

Southern University of Science and Technology

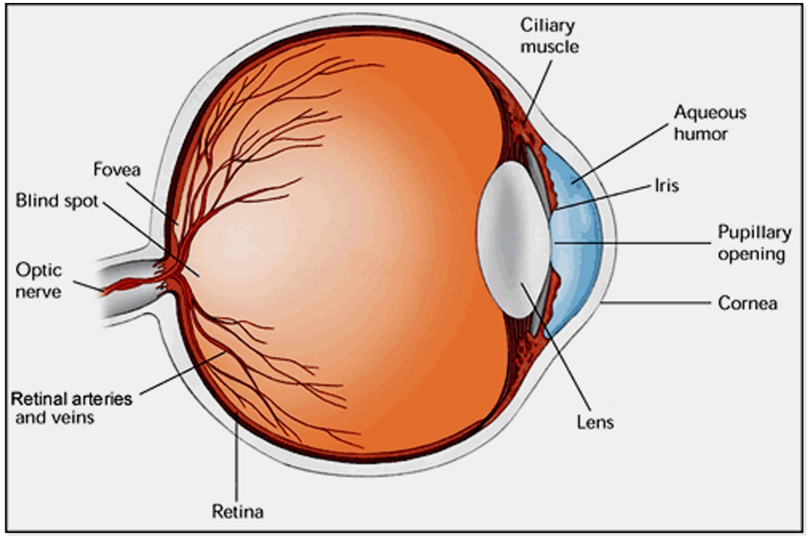
Agenda

- A very short history
- Human and computer vision
- Recent advancements
- Computer vision applications
  - Medical imaging
  - Classification with unbalanced classes

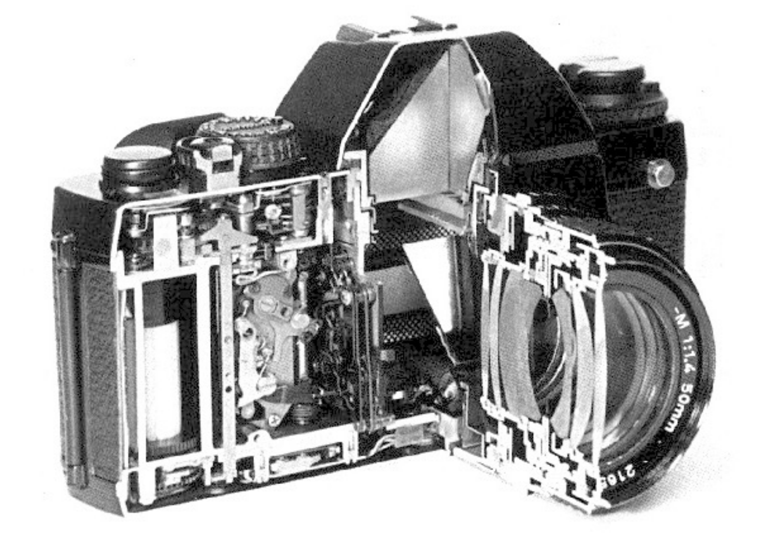
Computer vision:
A very short history

Image Formation

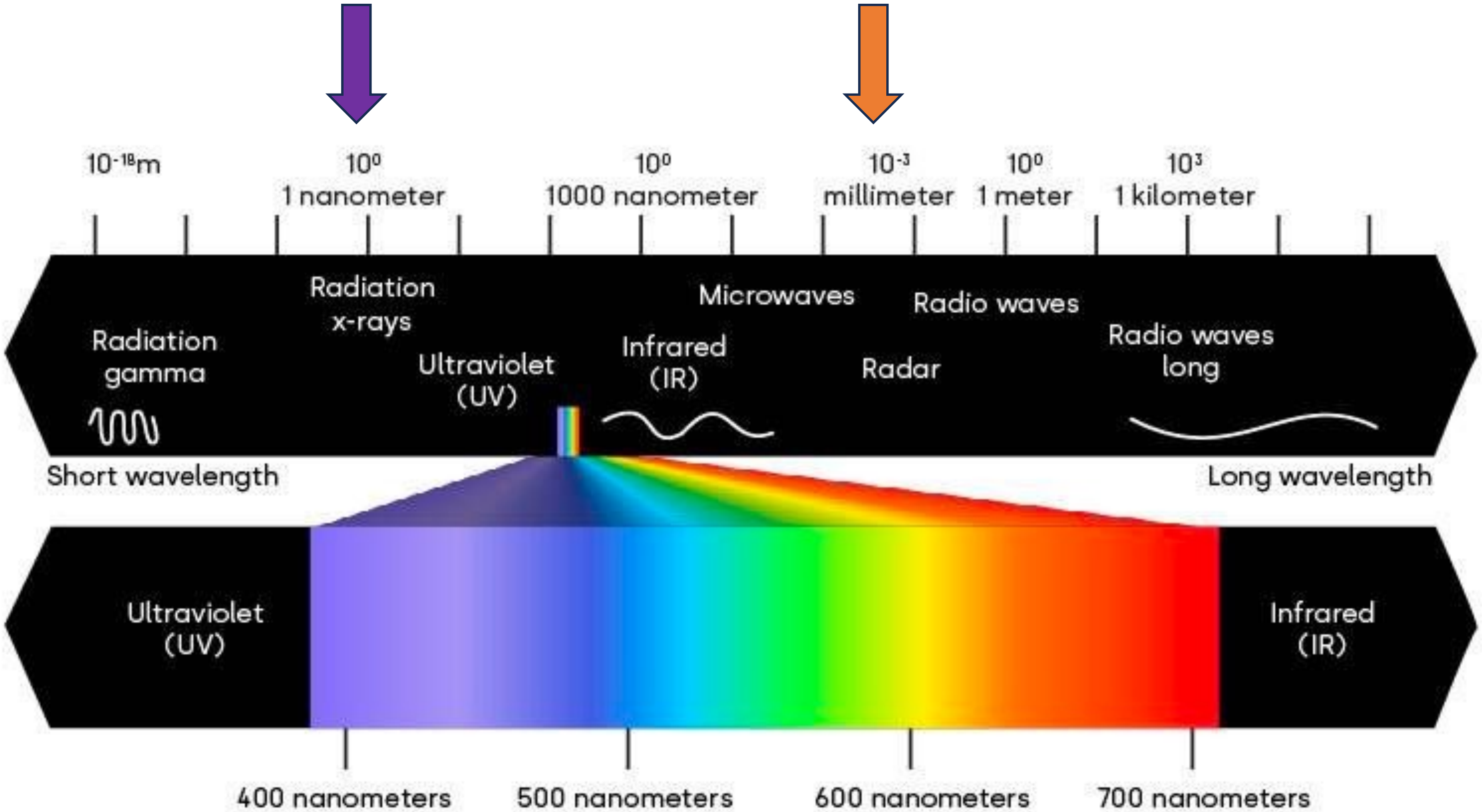
- Human: lens forms image on retina, sensors (rods and cones) respond to light



- Computer: lens system forms image, sensors (CCD, CMOS) respond to light



The spectrum of visible light



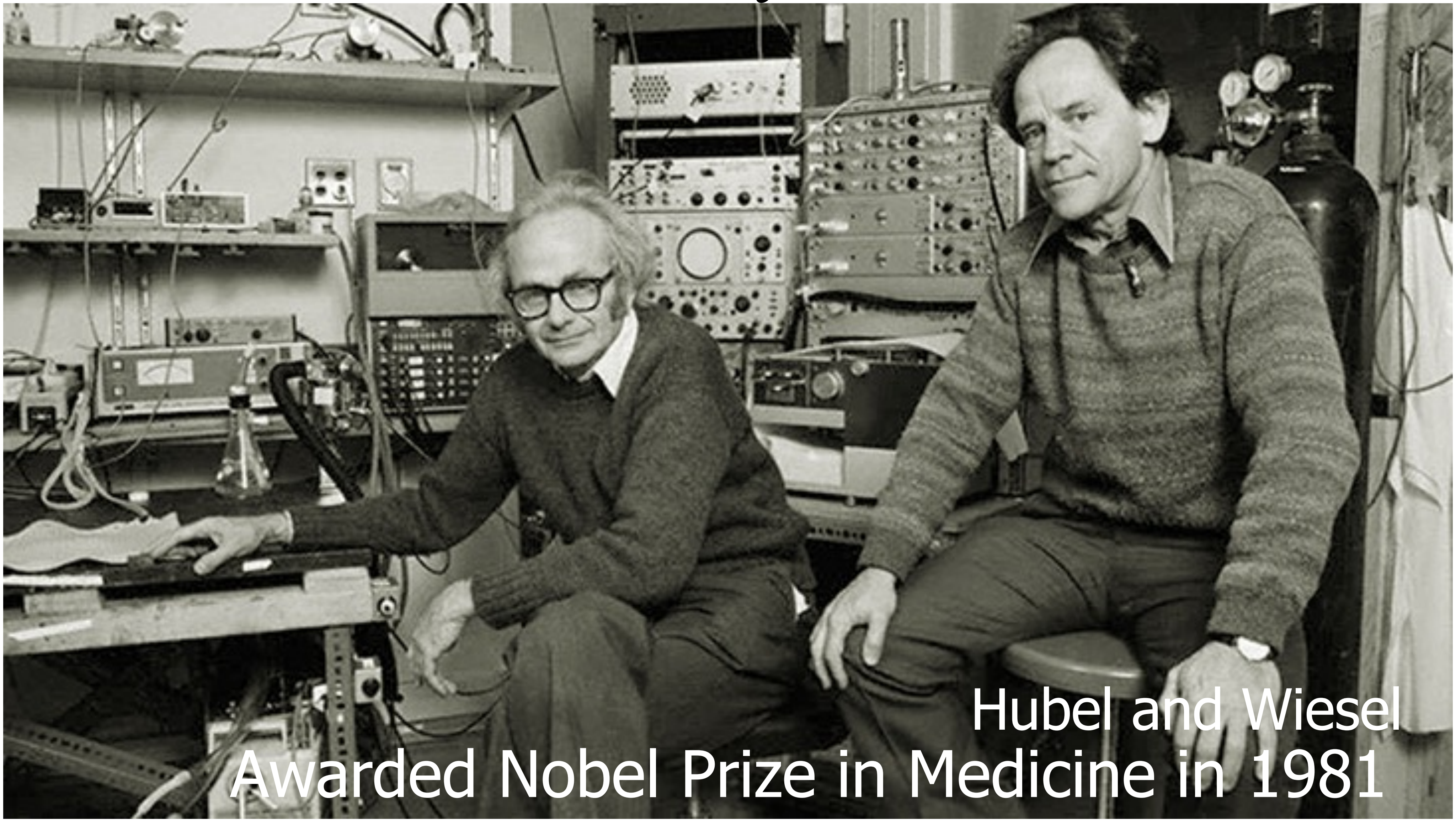
Human vision vs Computer vision

What we see

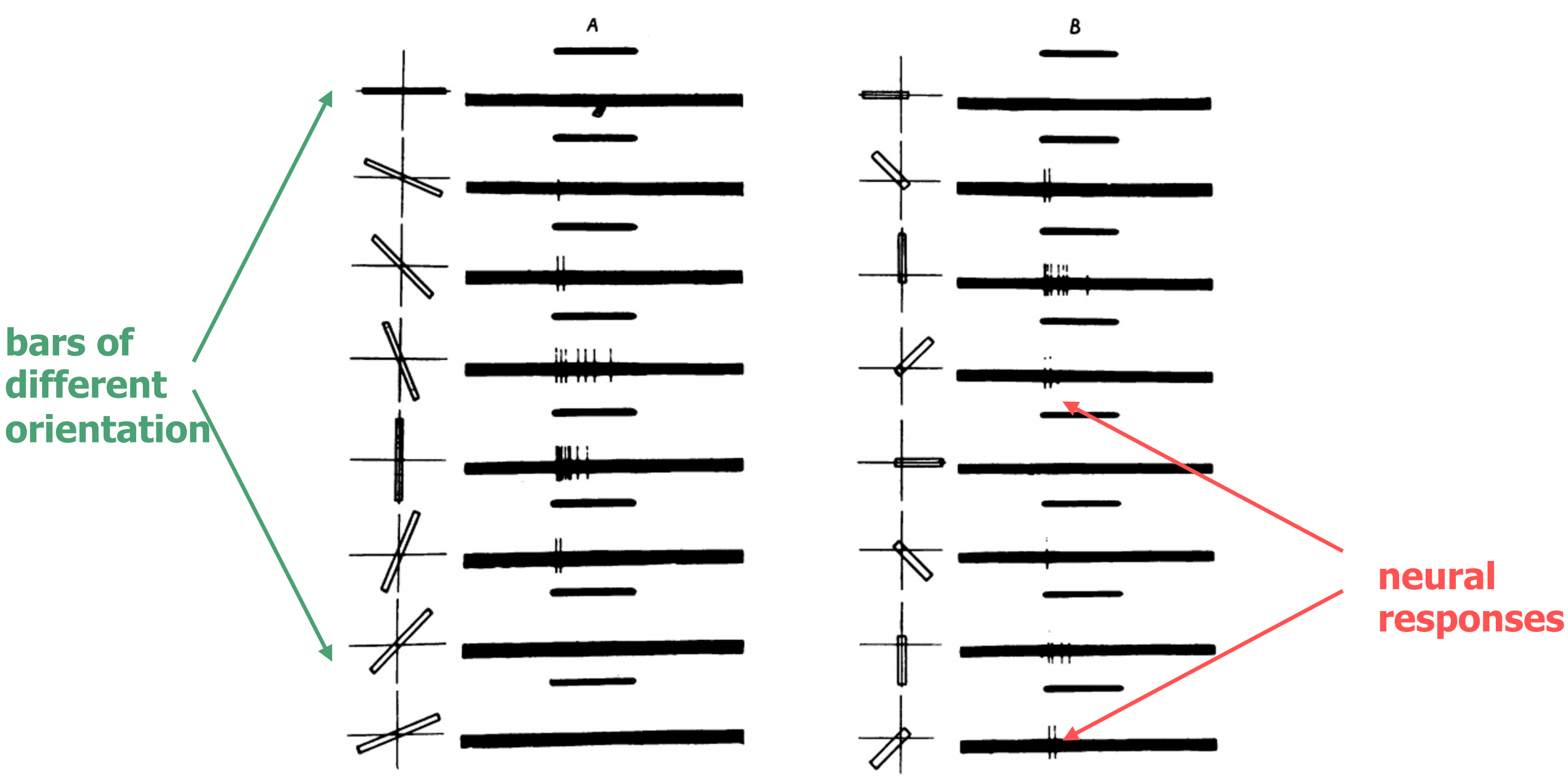
0	3	2	5	4	7	6	9	8
3	0	1	2	3	4	5	6	7
2	1	0	3	2	5	4	7	6
5	2	3	0	1	2	3	4	5
4	3	2	1	0	3	2	5	4
7	4	5	2	3	0	1	2	3
6	5	4	3	2	1	0	3	2
9	6	7	4	5	2	3	0	1
8	7	6	5	4	3	2	1	0

What a computer sees

Information processing in the visual system

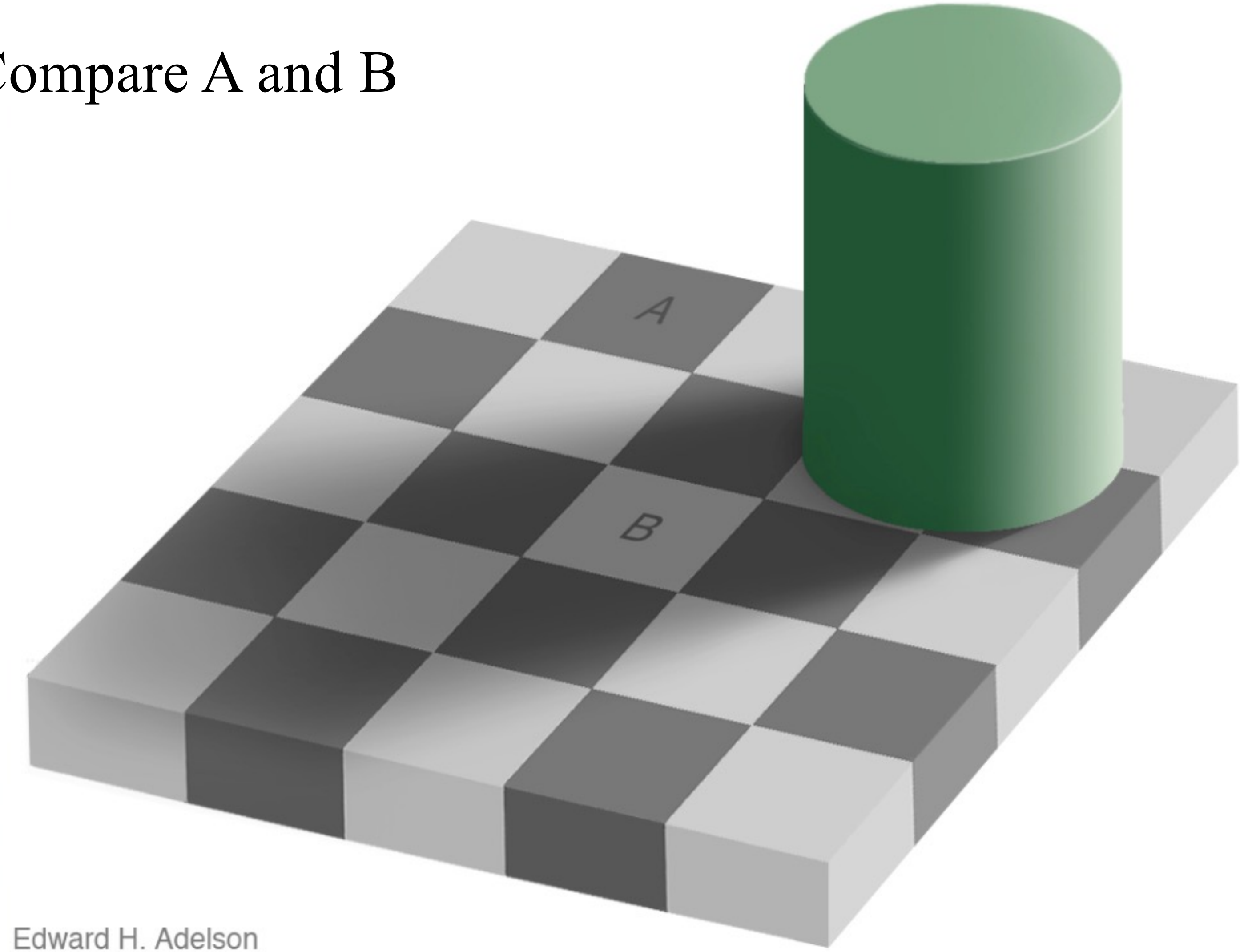


Hubel and Wiesel, 1959



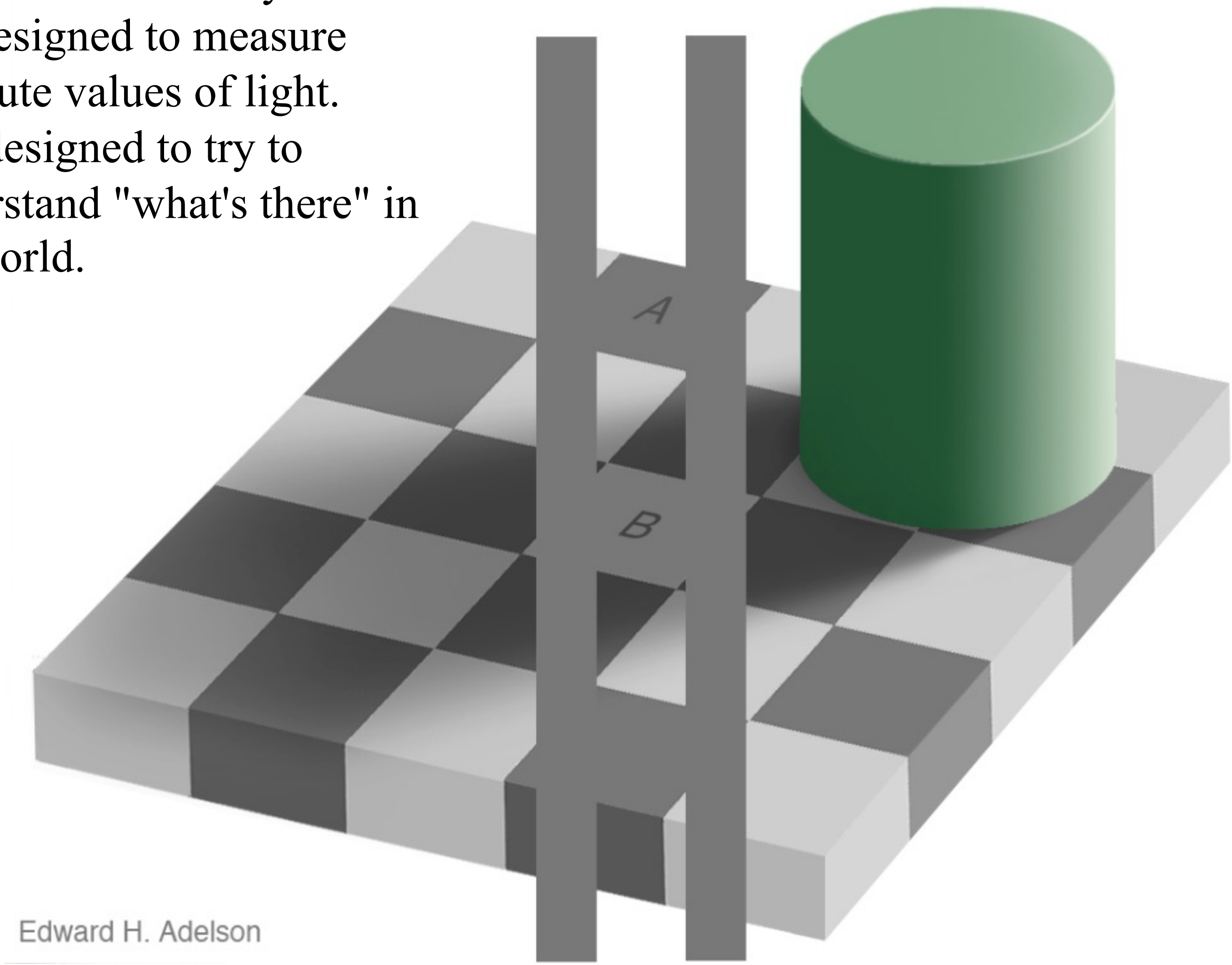
The Checker Shadow Illusion

Compare A and B

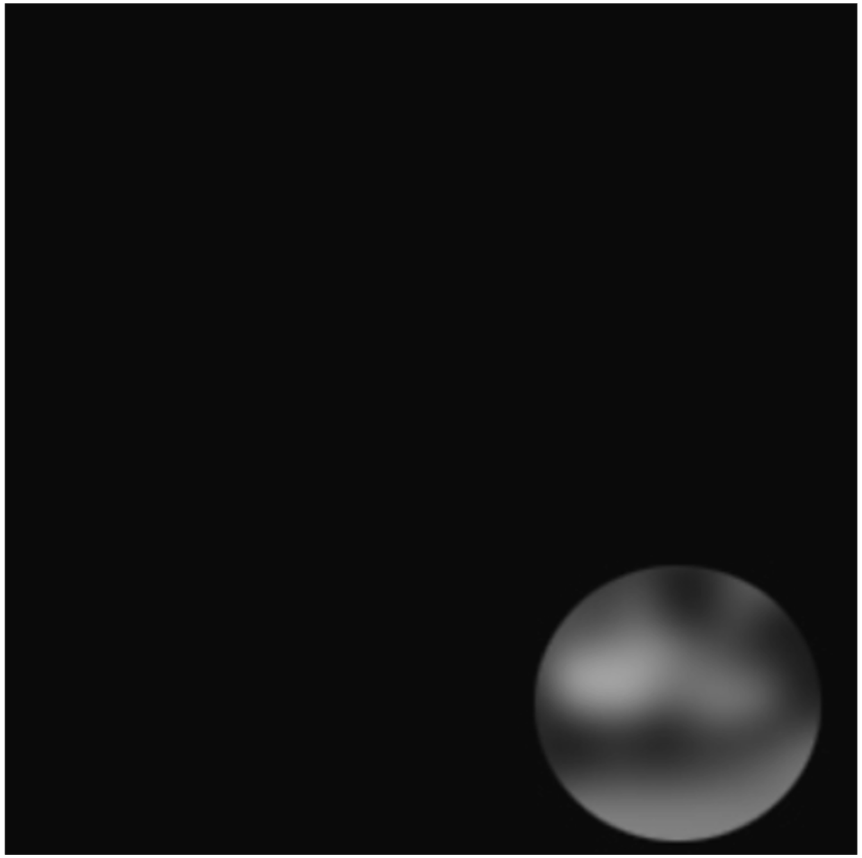


The "Proof"

- The human vision system is not designed to measure absolute values of light.
- It is designed to try to understand "what's there" in the world.



Visual context in a scene



Torralba, 2003

Visual context in a scene

Torralba, 2003

Visual context in a scene

Torralba, 2003

Takeaway

- The human vision system is not designed to measure absolute values of light.
  - It is designed to try to understand "what's there" in the world
- Images are fundamentally ambiguous:
  - Computer vision is ill-posed.
- We cannot be sure about what is there
- We use as many cues as we can to make our best guess as to what is there
- Amazingly, the human visual system usually guesses correctly.
  - Or does it?
  - When do we make a guess?

What information in the world does vision rely on?

- Objects tend to have rigid, solid surfaces
- Surfaces have constant or smoothly varying color and texture
- Surface boundaries are defined by a change in color, texture, value
- Objects tend to be opaque and occlude each other (nearer ones occlude farther ones)
- Object relationships and object part to object relationships tend to have stereotypical properties
- 3D => 2D projection is unique and computable
- Objects shapes stay constant in variable conditions (light/shadow, orientation, distance)
- ....

Is the goal of AI to replicate human intelligence?

- Computer vision does not need to be biomimetic (mimicking biology).
- What might be the pros and cons of developing AI that is based on neuroscience? On human perception?

Human and computer vision

Onto different but overlapping paths

A little story about Computer Vision

In 1966, Marvin Minsky at MIT asked his undergraduate student Gerald Jay Sussman to "spend the summer linking a camera to a computer and getting the computer to describe what it saw". We now know that the problem is slightly more difficult than that. (Szeliski 2009, Computer Vision)

A little story about Computer Vision

Founder, MIT AI Lab, 1959

In 1966, Marvin Minsky at MIT asked his undergraduate student Gerald Jay Sussman to “spend the summer linking a camera to a computer and getting the computer to describe what it saw”. We now know that the problem is slightly more difficult than that. (Szeliski 2009, Computer Vision)

MIT CSAIL

mit media lab

MIT Project MAC:
The Summer Vision Project, 1966

The final goal is OBJECT IDENTIFICATION which will actually name objects by matching them with a vocabulary of known objects.

Subgoal for July

Analysis of scenes consisting of non-overlapping objects from the following set:

balls

bricks with faces of the same or different colors or textures

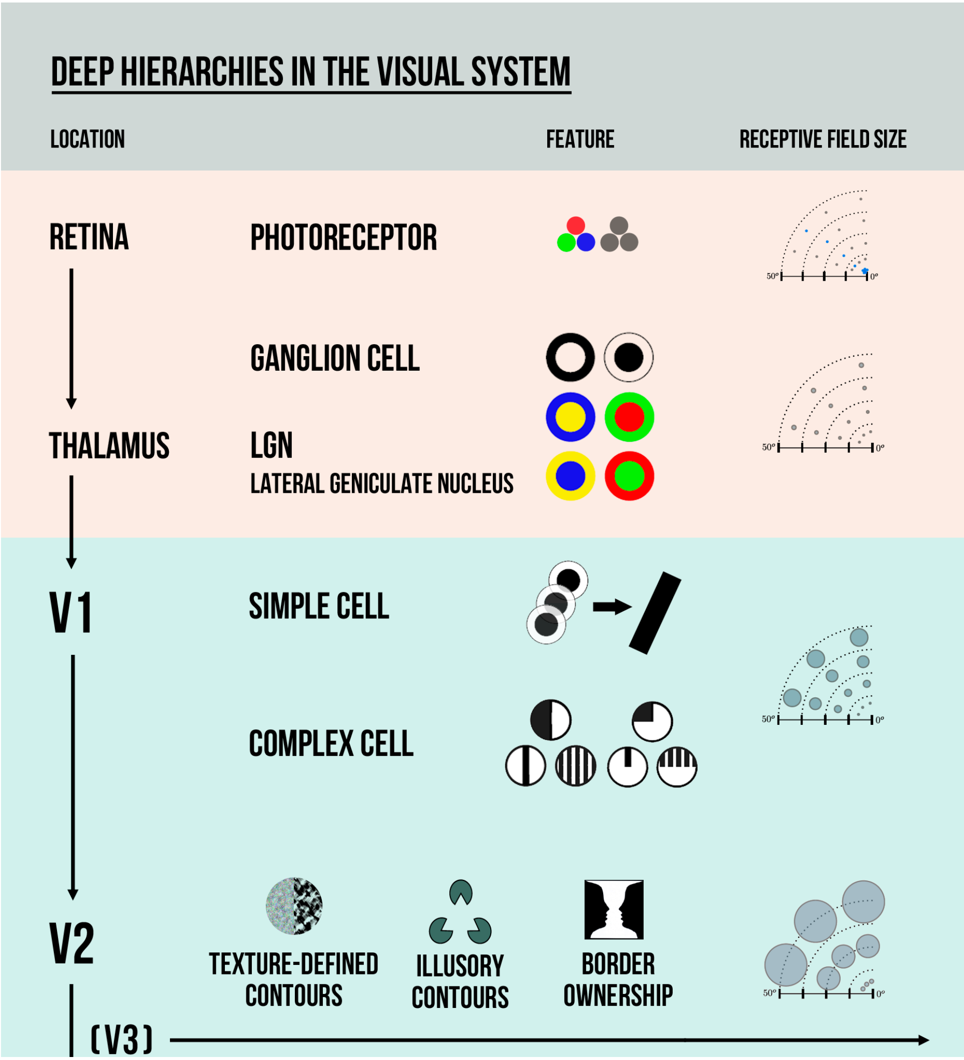
cylinders.

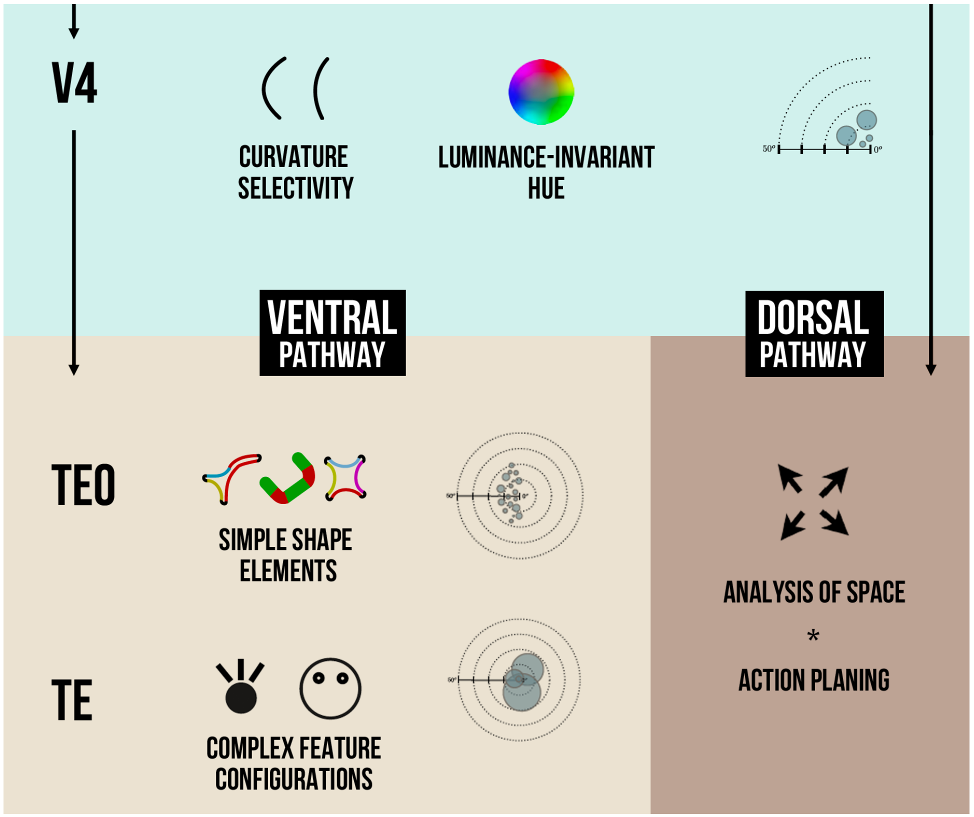
Each face will be of uniform and distinct color and/or texture.

Background will be homogeneous.

MIT Summer Vision Project

Human visual system





Retina 视网膜
Thalamus 丘脑

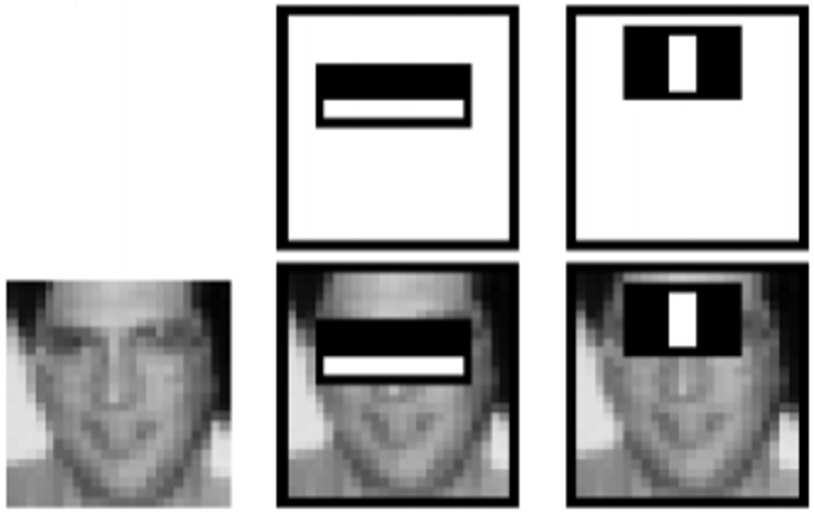
https://en.wikipedia.org/wiki/Visual_hierarchy

Computer Vision

Mimic human vision system

Low-level Computer Vision:
Feature-based algorithms

- Contrast and edges
- Points of interest
- Regions
- Contours (snakes)
- Optical flow
- Gradient-based features (e.g. HoG)
- Scale invariance (e.g. SIFT)



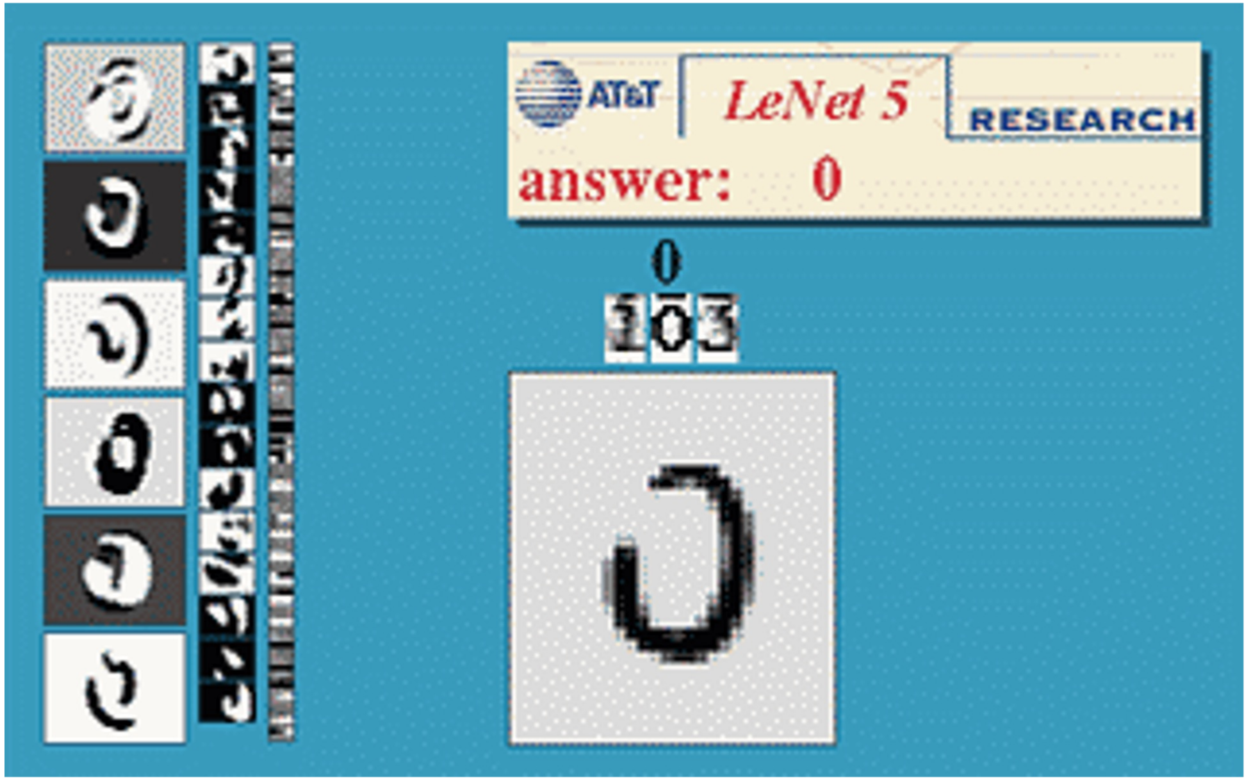
Viola-Jones object detection
(based on Haar features)

High-level Computer Vision: Applications

- Image alignment (e.g., panoramic mosaics)
- Object recognition
- 3D reconstruction (e.g., stereo)
- Motion tracking
- Indexing and content-based retrieval
- Robot navigation

High-Level Computer Vision:

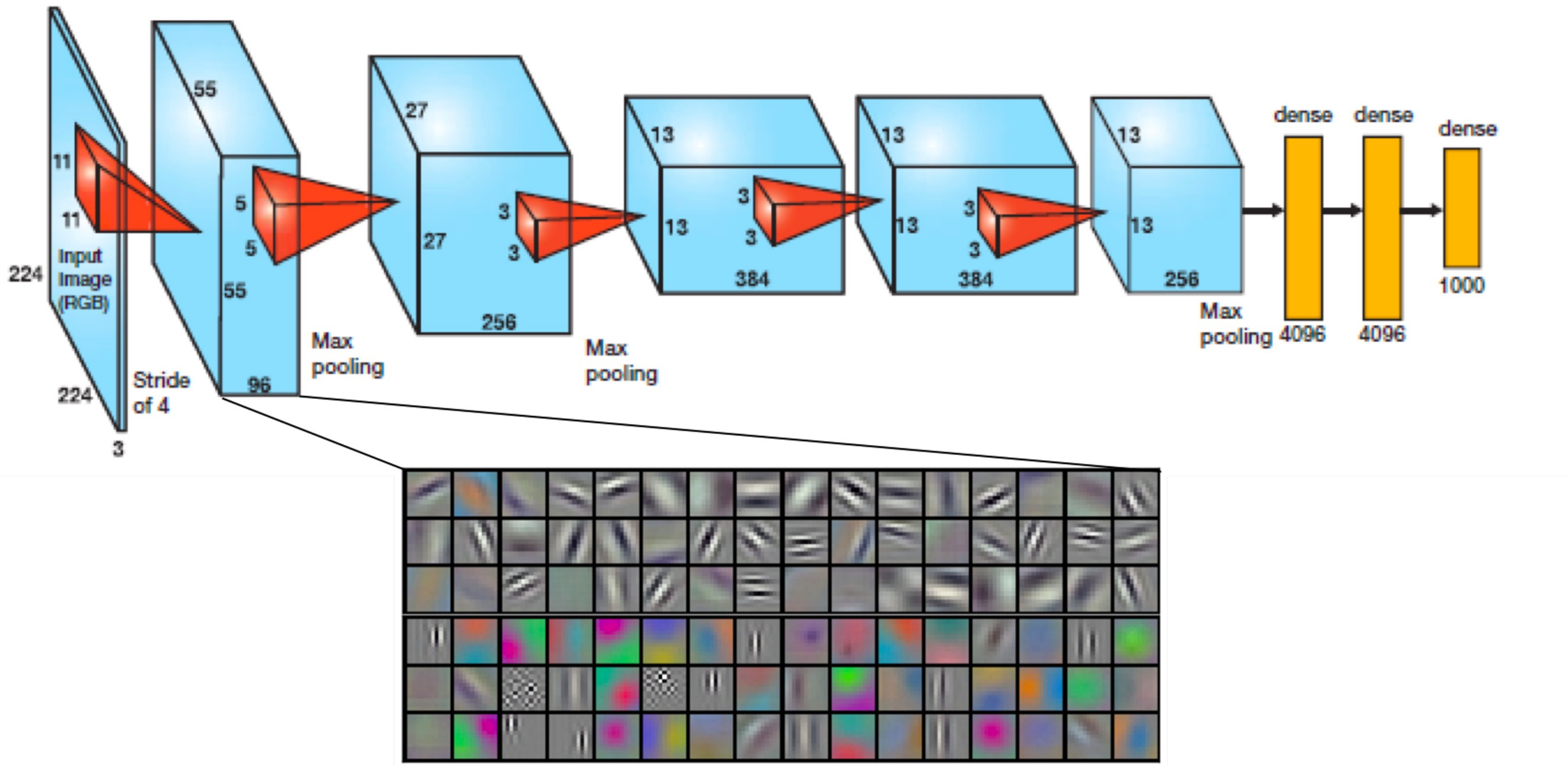
LeNet-5: **First** modern convolutional neural network



- Introduced the MNIST handwritten digit dataset, 1994
- Follow-up work led to automated zip code reading

LeCun et al, 1989

AlexNet and CNN resurgence



Will dive into more details later!

Krizhevsky et al., 2012

AlexNet and CNN resurgence

- ImageNet dataset: **14M** image database (Deng et al., 2009)
- ImageNet Challenge: **1000** categories (on abbreviated ImageNet): 2010
- 2012: AlexNet (Krizhevsky et al.) achieves 16% error. Previously, errors were around 25%!
- Every winner since 2012 has been a CNN.

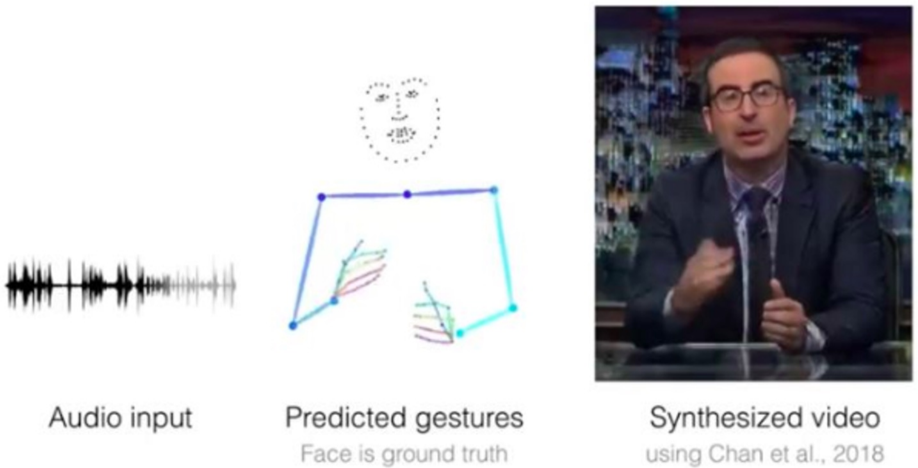
Why did it take so long for CNNs to take off?

1989 -> 2012

- Computing power (Moore's law)
- GPU development, largely thanks to the gaming industry (uniquely adept for matrix and vector operations)
- Training data availability (images and labels)

More recent advancements

Style of conversational gestures



Ginosaur et al, 2019

Inverse recipes (from images)

Title: Biscuits

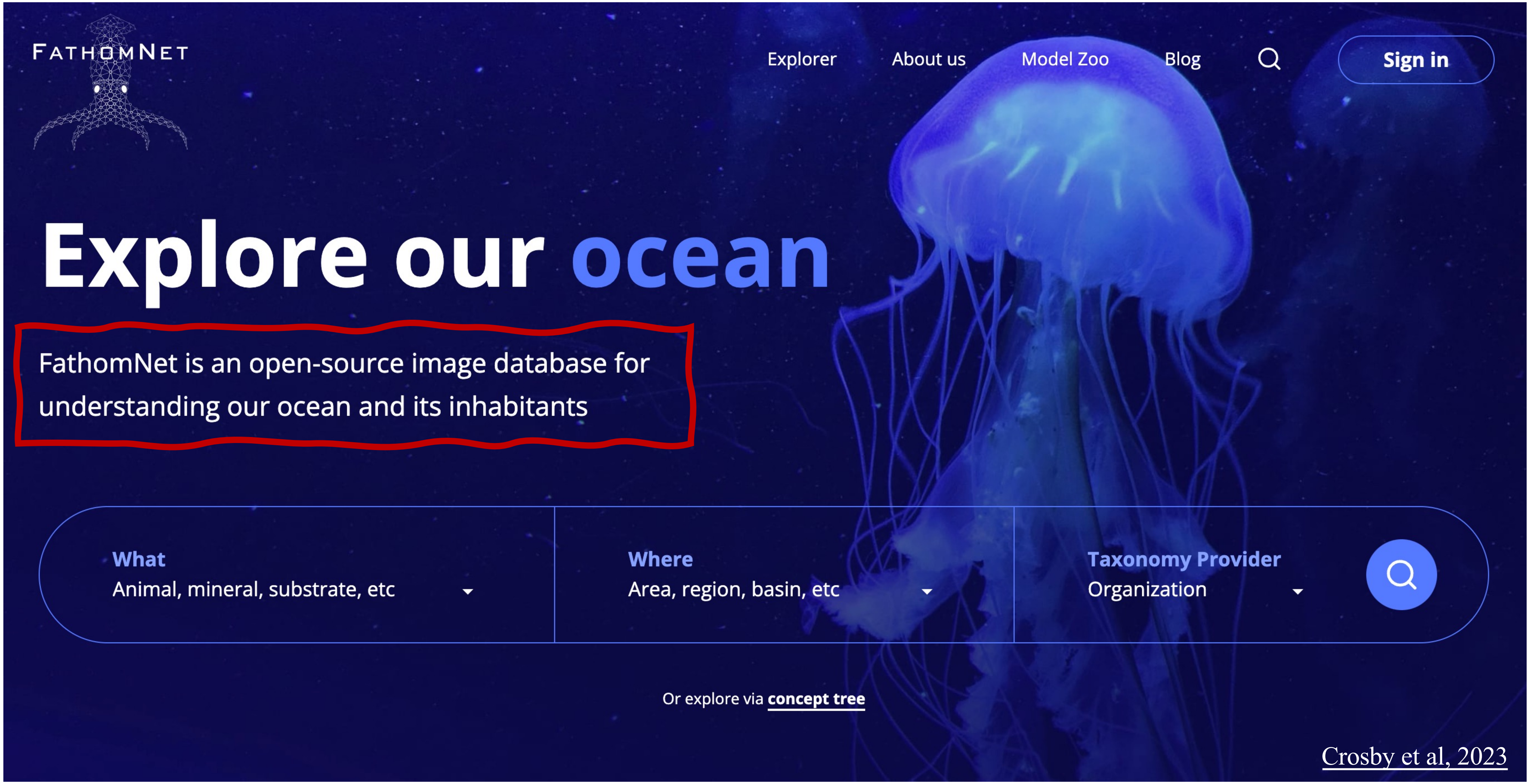
Ingredients:
Flour, butter, sugar, egg, milk, salt.

Instructions:
- Preheat oven to 450 degrees.
- Cream butter and sugar.
- Add egg and milk.
- Sift flour and salt together.
- Add to creamed mixture.
- Roll out on floured board to 1/4 inch thickness.
- Cut with biscuit cutter.
- Place on ungreased cookie sheet.
- Bake for 10 minutes.

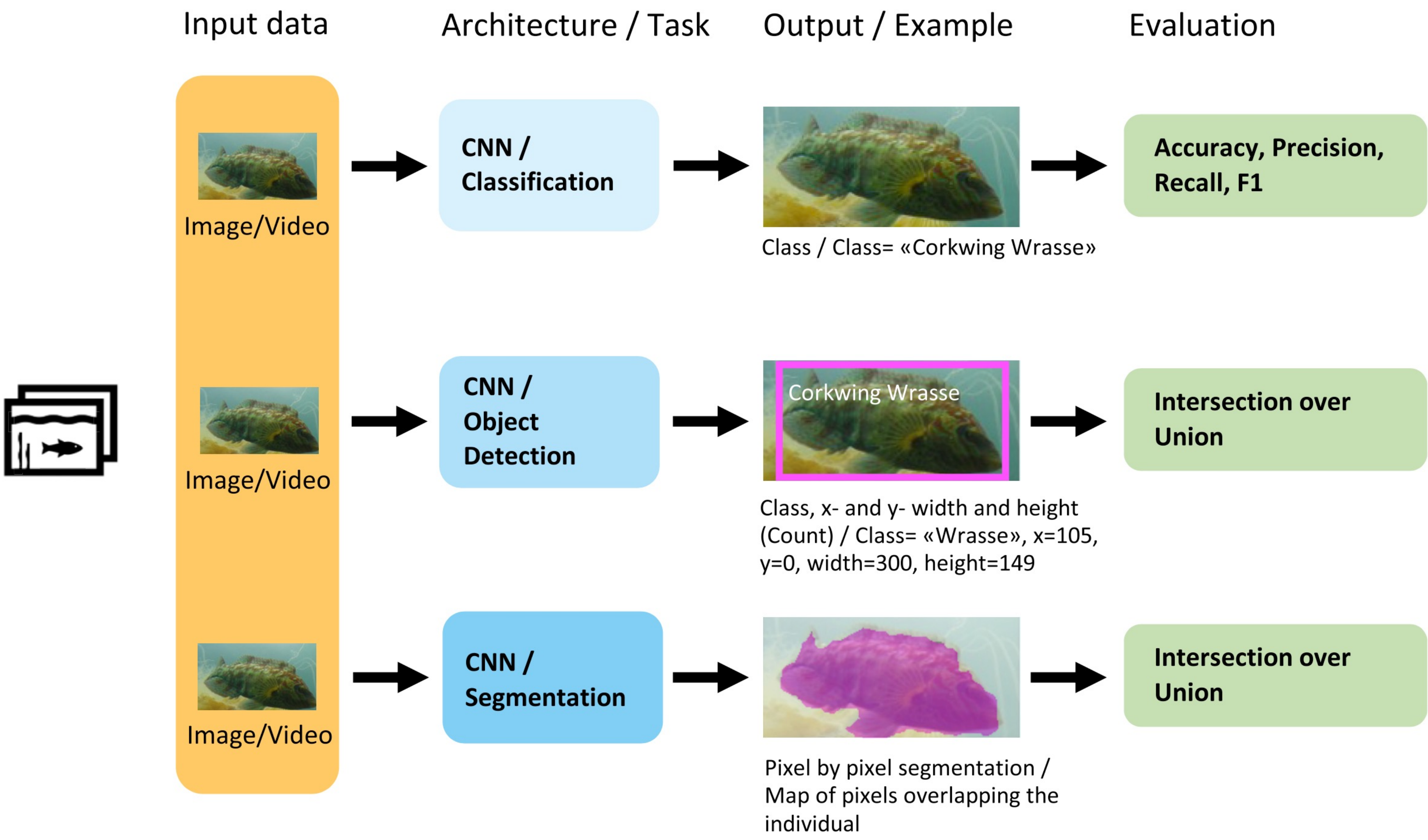
Figure 1: **Example of a generated recipe**, composed of a title, ingredients and cooking instructions.

Salvador et al, 2019


Ocean Vision AI Project



Crosby et al, 2023

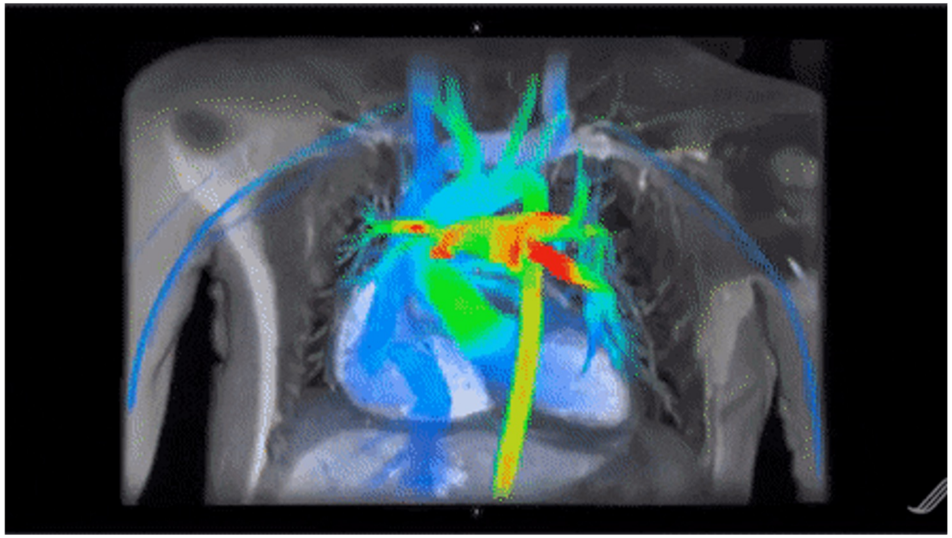

Input data
Architecture / Task
Output / Example
Evaluation
Image/Video
CNN / Classification
Class / Class= «Corkwing Wrasse»
Accuracy, Precision, Recall, F1
Image/Video
CNN / Object Detection
Corkwing Wrasse
Class, x- and y- width and height (Count) / Class= «Wrasse», x=105, y=0, width=300, height=149
Intersection over Union
Image/Video
CNN / Segmentation
Pixel by pixel segmentation / Map of pixels overlapping the individual
Intersection over Union

SAM 2: Segment Anything in Images and Videos

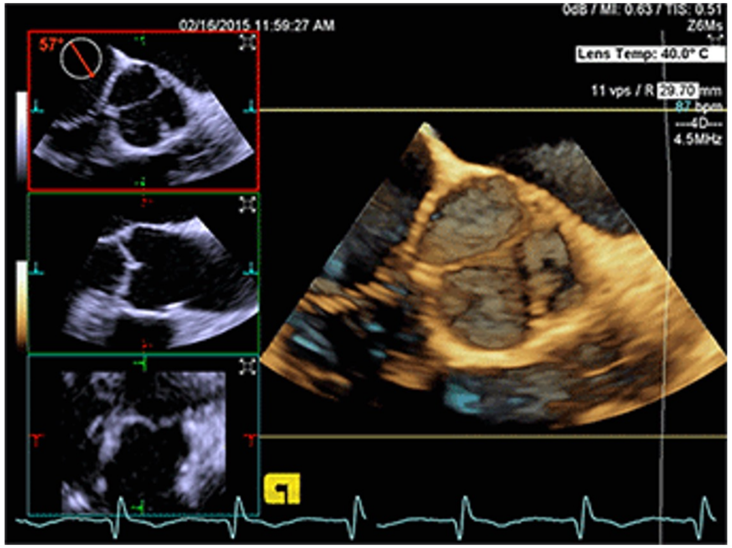
https://github.com/facebookresearch/sam2 Ravi et al, 2024

Computer vision applications： Medical Imaging

Computer vision for cardiac imaging



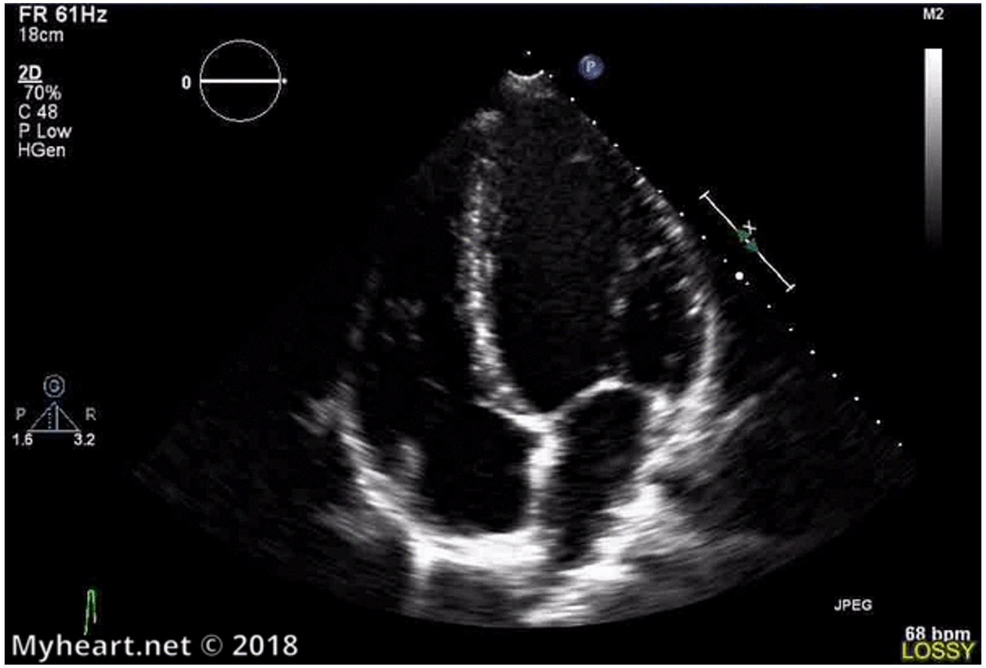
Arterys



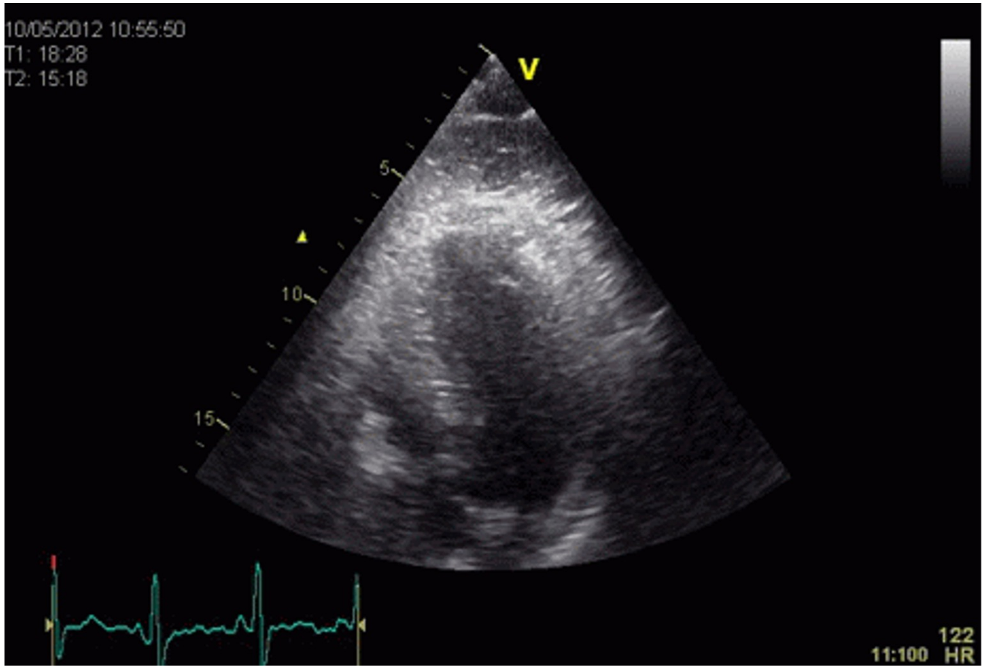
Siemens

Ultrasound image quality is a major issue

Good quality

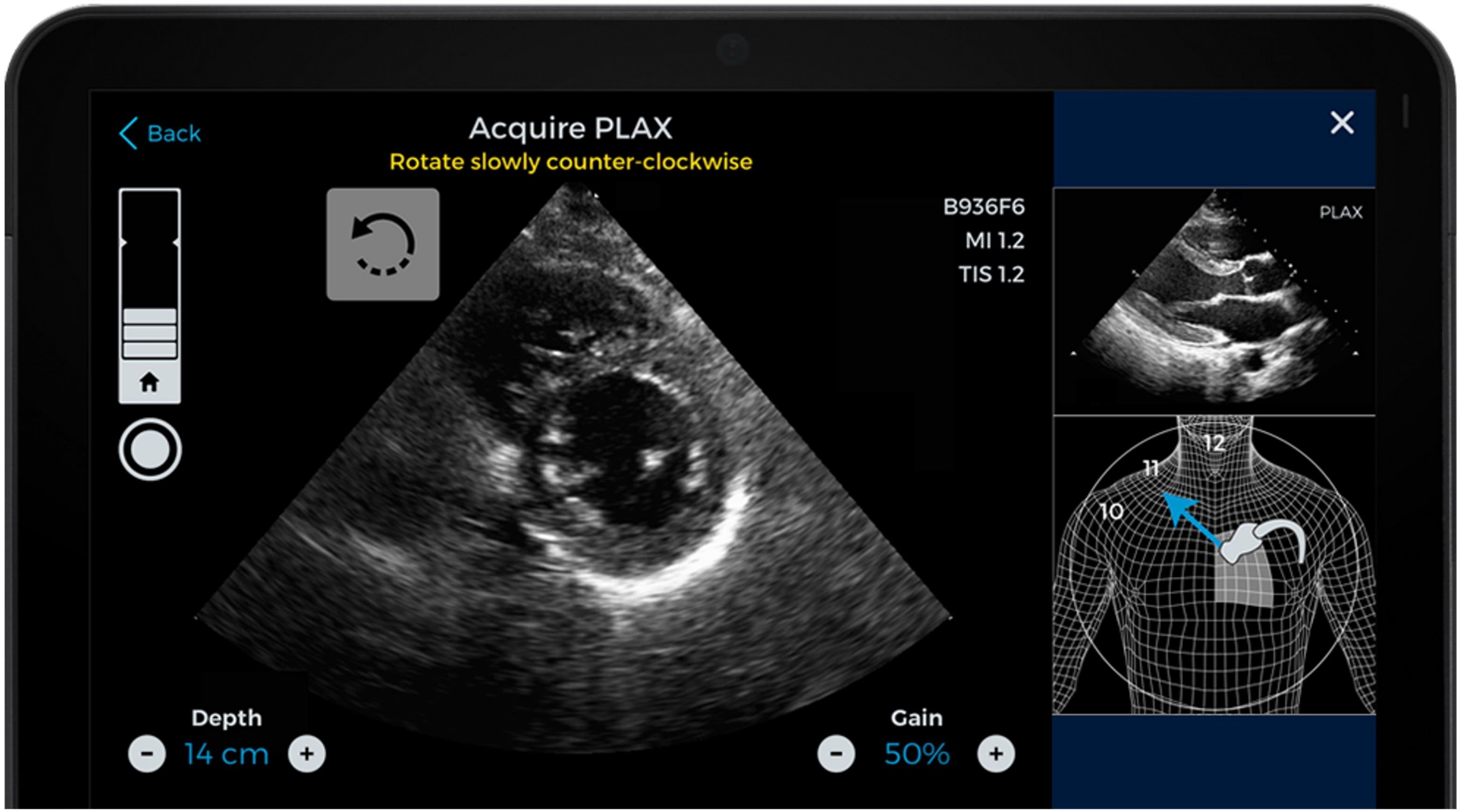


Poor quality



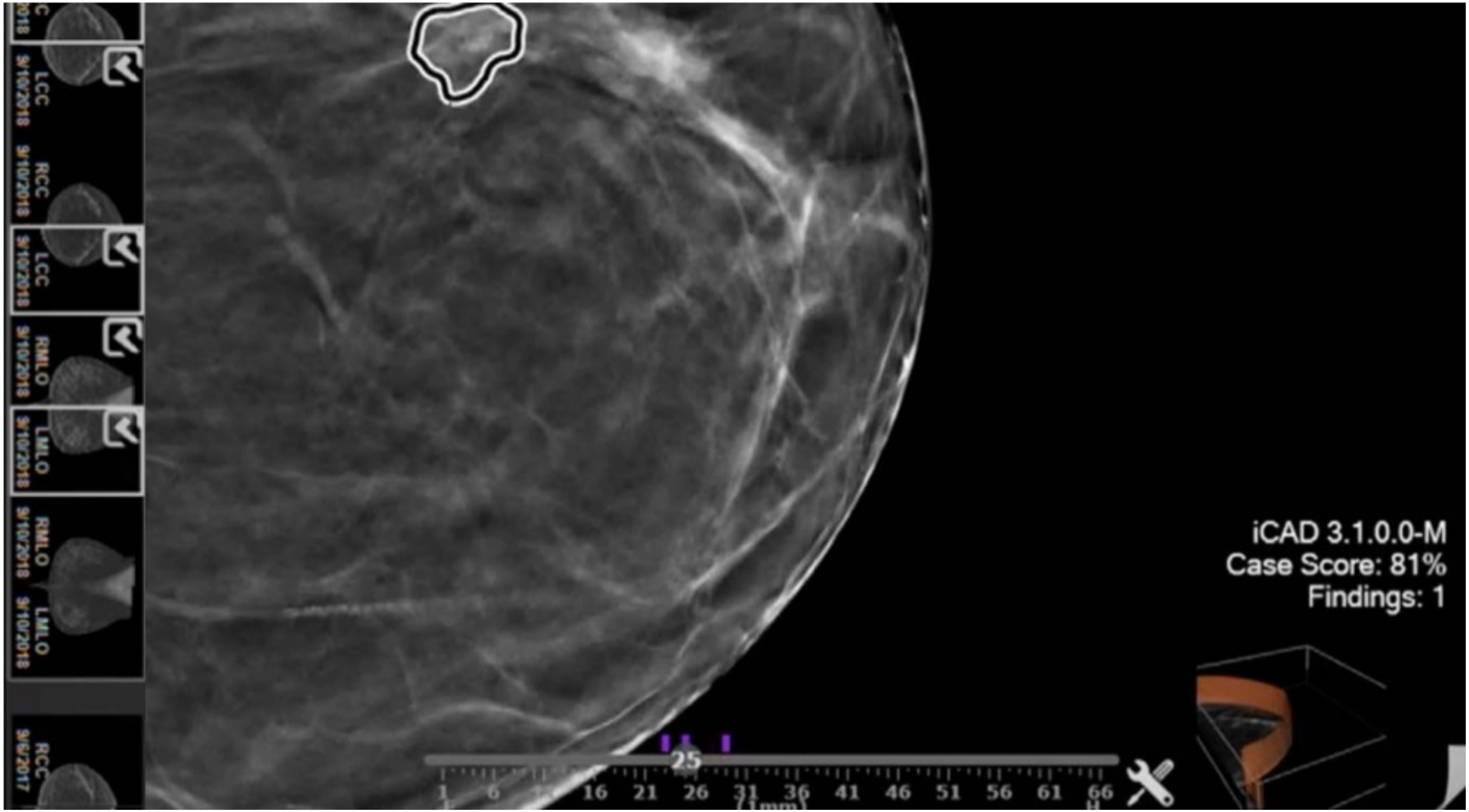
Ultrasound image quality is a major issue

Imaging guidance via machine learning



Caption Health

Categorizing a detected tumor



Breast cancer image courtesy of iCAD.

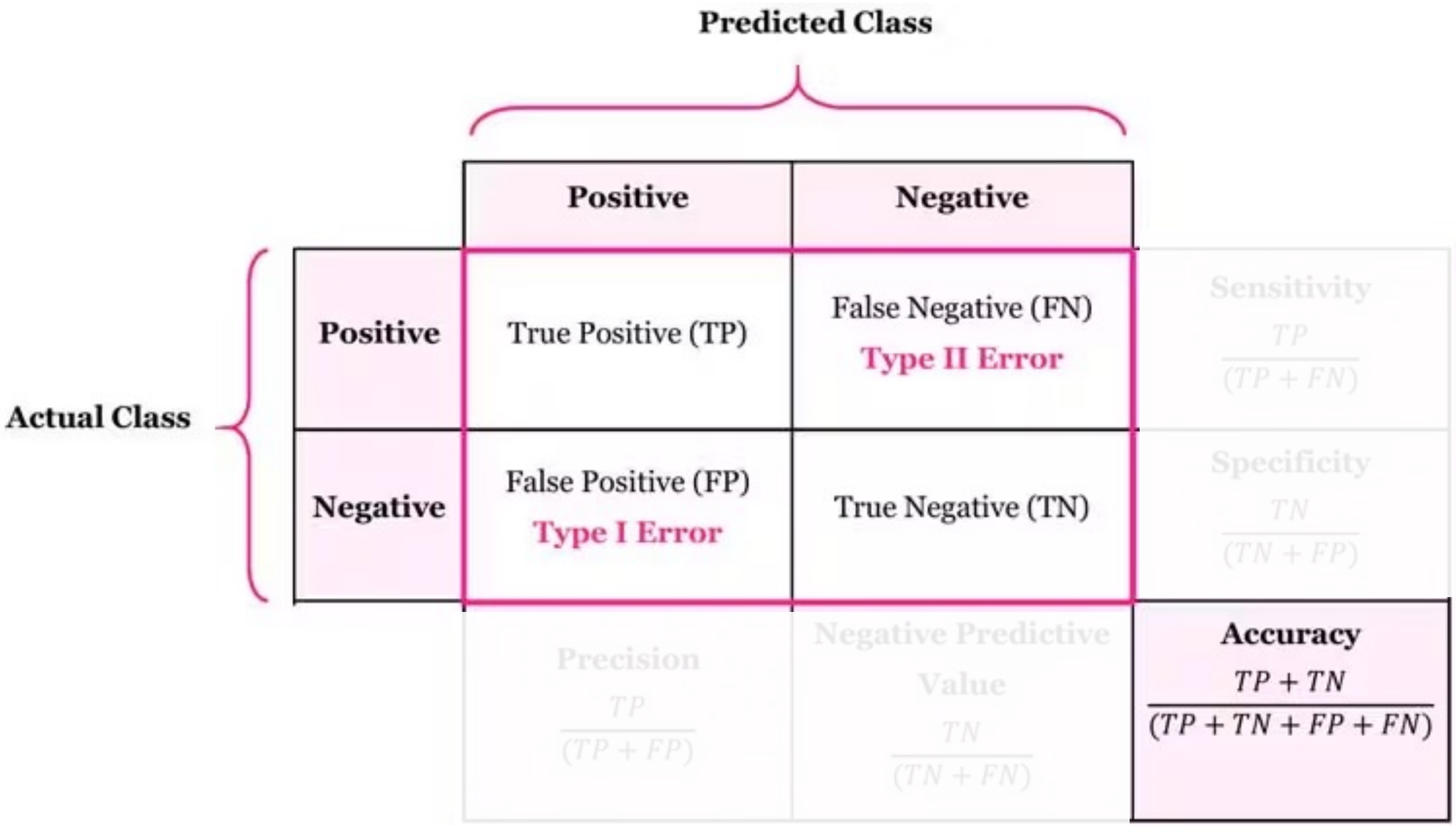
Machine learning for diagnosis: concerns

- **Metrics**: Appropriate evaluation isn't always used or reported.
- **Data**: Patient data are unbalanced. The most vulnerable patients are highly underrepresented. Algorithms are poor at generalizing to out-of-set cases.
- **Isolation**: Algorithms are often developed and evaluated without clinical experts, without regard for how they might integrate in a clinical workflow, and without appropriate clinical testing.
- **Privacy**: regulations are often insufficient for protecting patient data from re-identification, and at the same time complicate data sharing for verification.
- **Explanation**: Algorithms are often black boxes. Interpretation techniques and confidence estimation in deep learning are new and active research areas.
- **Hype**: Trust in AI among both laypeople and medical professionals may be inflated by hype and overly marketed results.

Classification with unbalanced classes

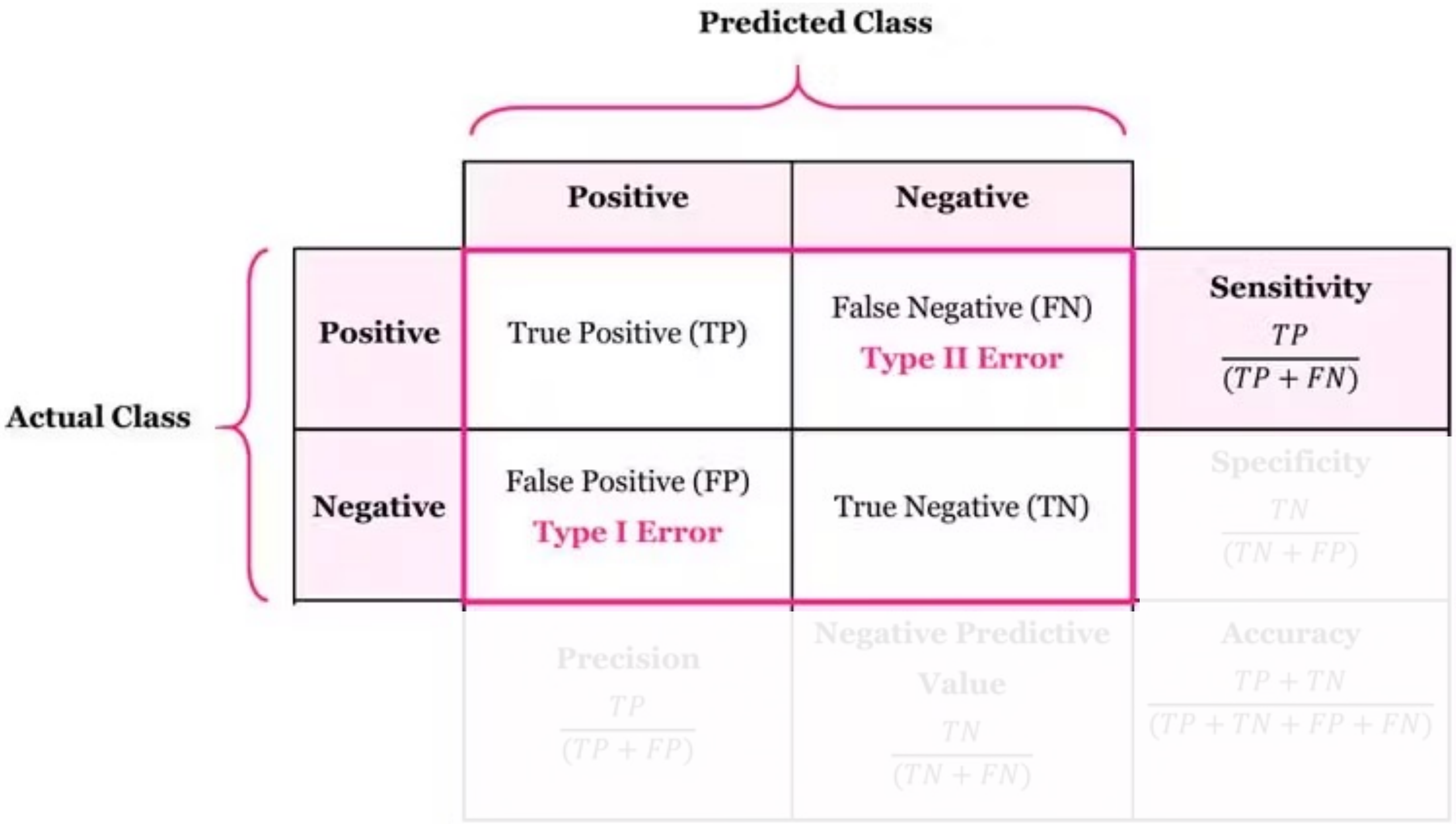
		Predicted Class	
		Positive	**Negative**
Actual Class	**Positive**	True Positive (TP)	False Negative (FN) **Type II Error**
	Negative	False Positive (FP) **Type I Error**	True Negative (TN)

Classification with unbalanced classes



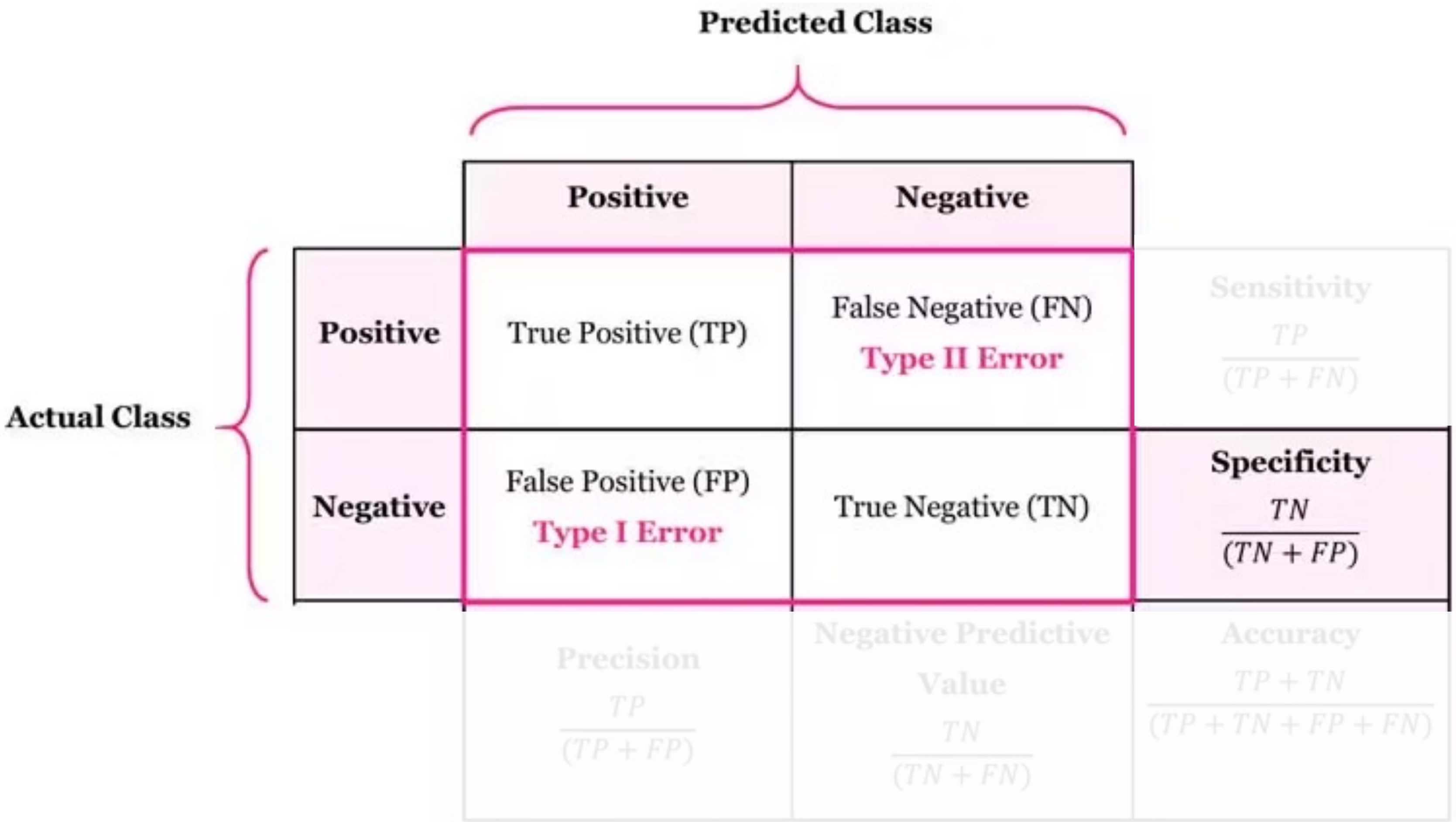
Accuracy: what proportion of predictions is correct

Classification with unbalanced classes



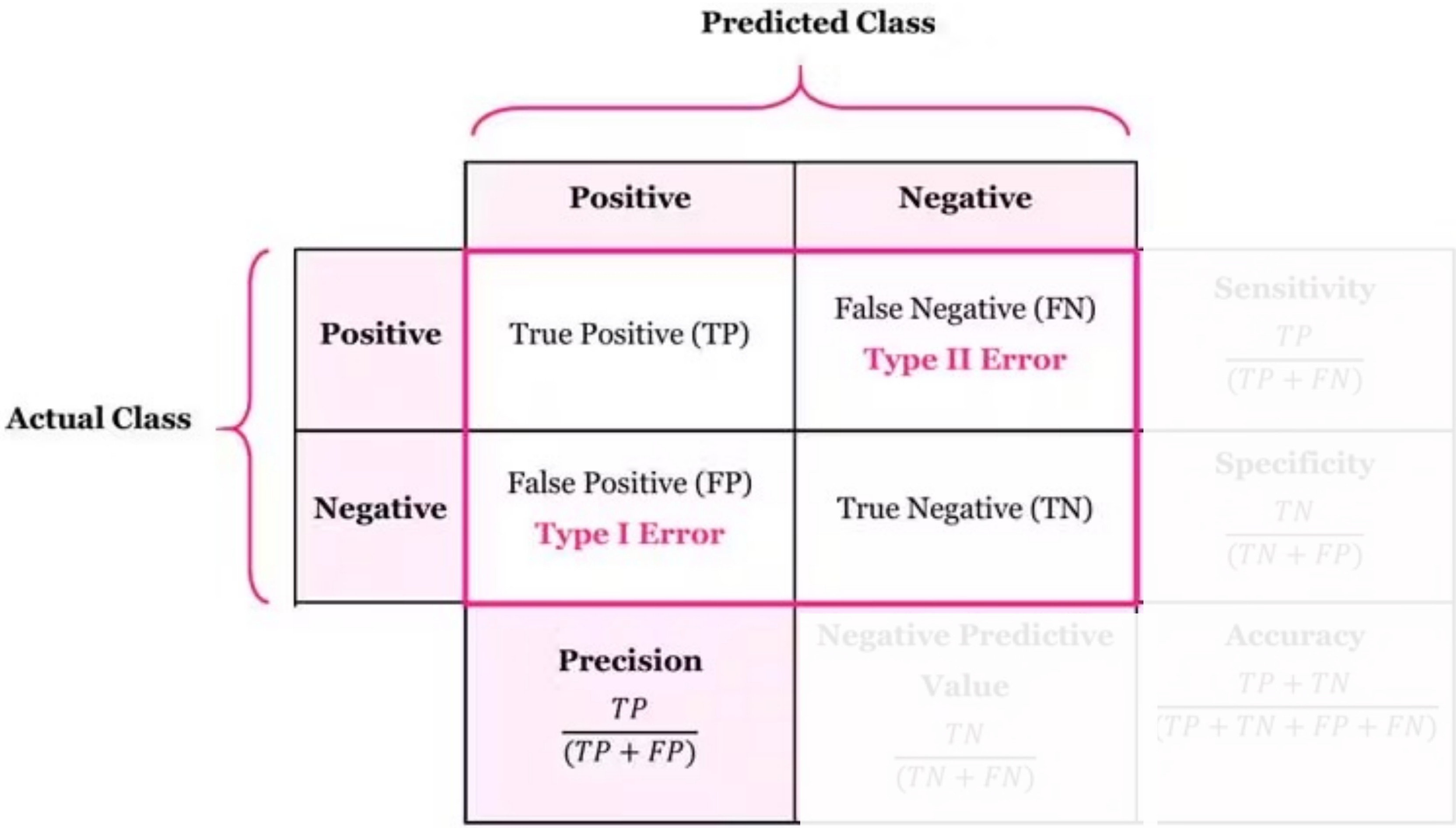
Sensitivity (**recall**): what proportion of sick people are diagnosed with the condition?

Classification with unbalanced classes



Specificity: what proportion of healthy people are diagnosed as not having the condition?

Classification with unbalanced classes



Precision: what proportion of positive diagnoses are correct?

Classification with unbalanced classes

Actual Class \ Predicted Class	Positive	Negative	
Positive	True Positive (TP)	False Negative (FN) **Type II Error**	**Sensitivity** $\frac{TP}{(TP + FN)}$
Negative	False Positive (FP) **Type I Error**	True Negative (TN)	**Specificity** $\frac{TN}{(TN + FP)}$
	Precision $\frac{TP}{(TP + FP)}$	**Negative Predictive Value** $\frac{TN}{(TN + FN)}$	**Accuracy** $\frac{TP + TN}{(TP + TN + FP + FN)}$

Accuracy: what proportion of predictions is correct
Sensitivity (recall): what proportion of sick people are diagnosed with the condition?
Specificity: what proportion of healthy people are diagnosed as not having the condition?
Precision: what proportion of positive diagnoses are correct?

Classification with unbalanced classes

		Predicted Class		
		Positive	**Negative**	
Actual Class	**Positive**	True Positive (TP)	False Negative (FN) **Type II Error**	**Sensitivity** $\frac{TP}{(TP+FN)}$
	Negative	False Positive (FP) **Type I Error**	True Negative (TN)	**Specificity** $\frac{TN}{(TN+FP)}$
		Precision $\frac{TP}{(TP+FP)}$	**Negative Predictive Value** $\frac{TN}{(TN+FN)}$	**Accuracy** $\frac{TP+TN}{(TP+TN+FP+FN)}$

Accuracy: what proportion of predictions is correct
Sensitivity (recall): what proportion of sick people are diagnosed with the condition?
Specificity: what proportion of healthy people are diagnosed as not having the condition?
Precision: what proportion of positive diagnoses are correct?

Classification with unbalanced classes

- 100 patients, 90 healthy, 10 sick

- Algorithm that is always negative:
  - 90% accuracy, 100% specificity, 0% recall

- Algorithm that is always positive:
  - 10% accuracy, 0% specificity, 100% recall

Machine learning for diagnosis: regulation

- 责任主体仍是研发者、生产者以及使用者
- 国家药监局器审中心《人工智能医疗器械注册审查指导原则》2022年3月：
  - **人工智能医疗器械是指基于“医疗器械数据”，采用人工智能技术实现其预期用途（即医疗用途）的医疗器械**

从智能产品类角度，可以细分为：

* 智能辅助诊断产品（如消化系统、心脑血管系统、神经系统、骨科、眼科、皮肤科、肿瘤等领域）；

* 智能辅助治疗产品（如内窥镜手术、神经外科手术、骨科手术、穿刺手术、口腔种植手术等领域）；

* 智能监护与生命支持产品（如研发监测心电、脑电、血糖、血氧、呼吸、睡眠等生理参数的智能监护产品或生命支持产品；

* 突破智能重症监护（ICU）、智能急救、智能新生儿监护等）；

* 智能康复理疗产品（如认知言语视听障碍康复、运动障碍康复等重点领域，研发融合脑机接口、人-机-电融合、虚拟现实/增强现实等技术的智能医用康复产品；精神类疾病、神经退行性疾病等领域，研发融合人工智能技术的理疗产品）；

* 智能中医诊疗产品（研发融合人工智能技术的脉诊仪、目诊仪、舌诊仪、四相仪等中医诊疗产品）。

Machine learning for diagnosis: regulation

产品注册重点关注以下要求

- **算法研究资料**：明确软件安全性级别（轻微、中等、严重），明确过拟合与欠拟合、假阴性与假阳性、数据污染与数据偏倚等风险的控制措施
- **用户培训方案**：软件安全性级别为严重级别、预期由患者使用或在基层医疗机构使用的产品
- 产品技术要求
- **说明书**：明确使用限制和必要警示提示信息；明确数据采集设备与采集过程；算法训练集、训练指标与结果

Thank you~

Wan Fang
Southern University of Science and Technology