Machine Learning for Medical Imaging

Adam Yala

MIT Jameel Clinic and CSAIL

Machine Learning for Medical Imaging

These presentation uses slides from Regina Barzilay, Stefanie Jegelka, Tommi Jaakkola, Yann Le Cun, Marc'Aurelio Ranzato, Alyosha Efros, Jonathan Shewchuk, and Ruizhi “Ray” Liao

Agenda

Data: What is medical imaging?

Method Foundations: How do we build models on imaging data?

Applications: How can we catch cancer earlier?

Interpretation: How can we audit our models?

Agenda

Data: What is medical imaging?

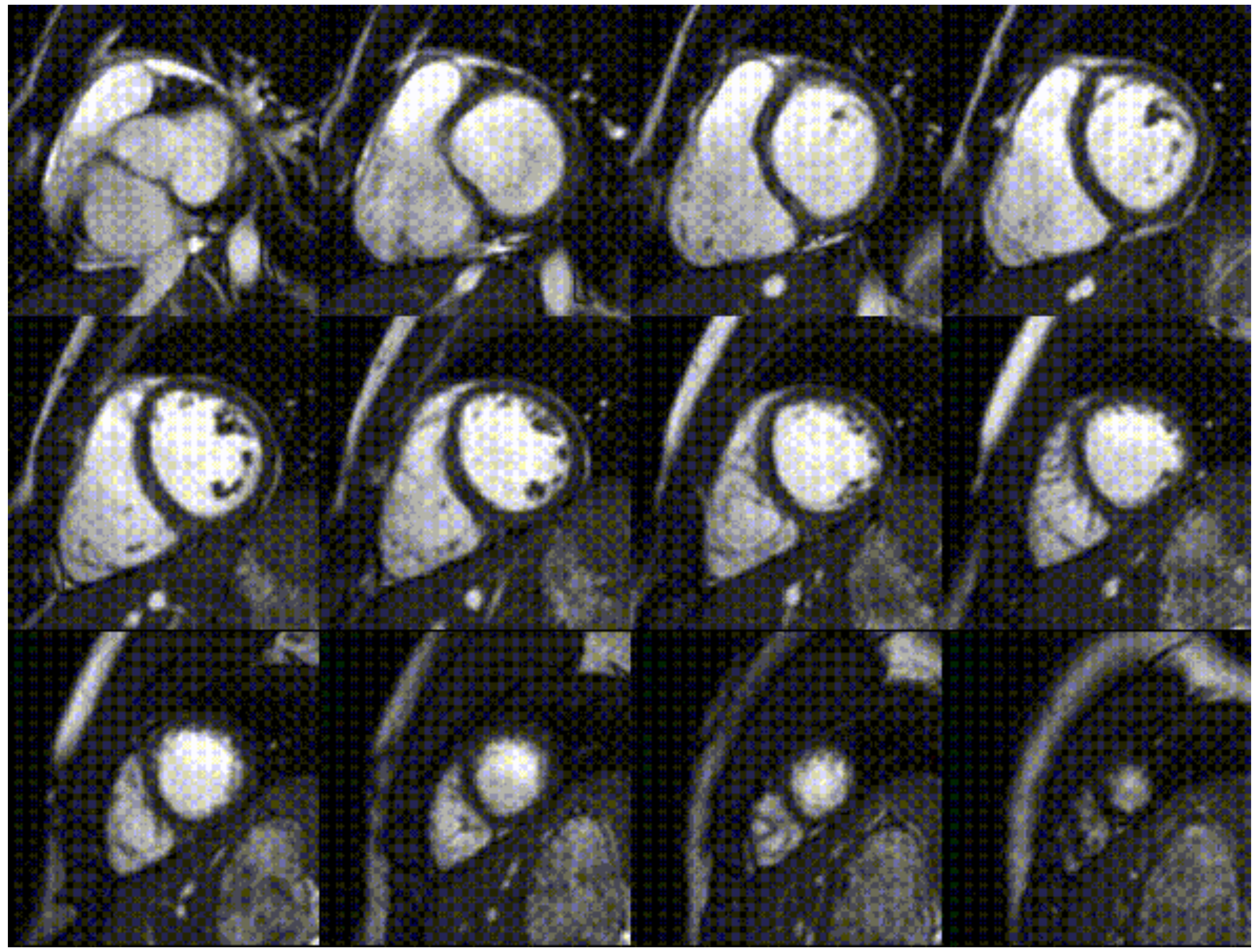
Method Foundations: How do we build models on imaging data?

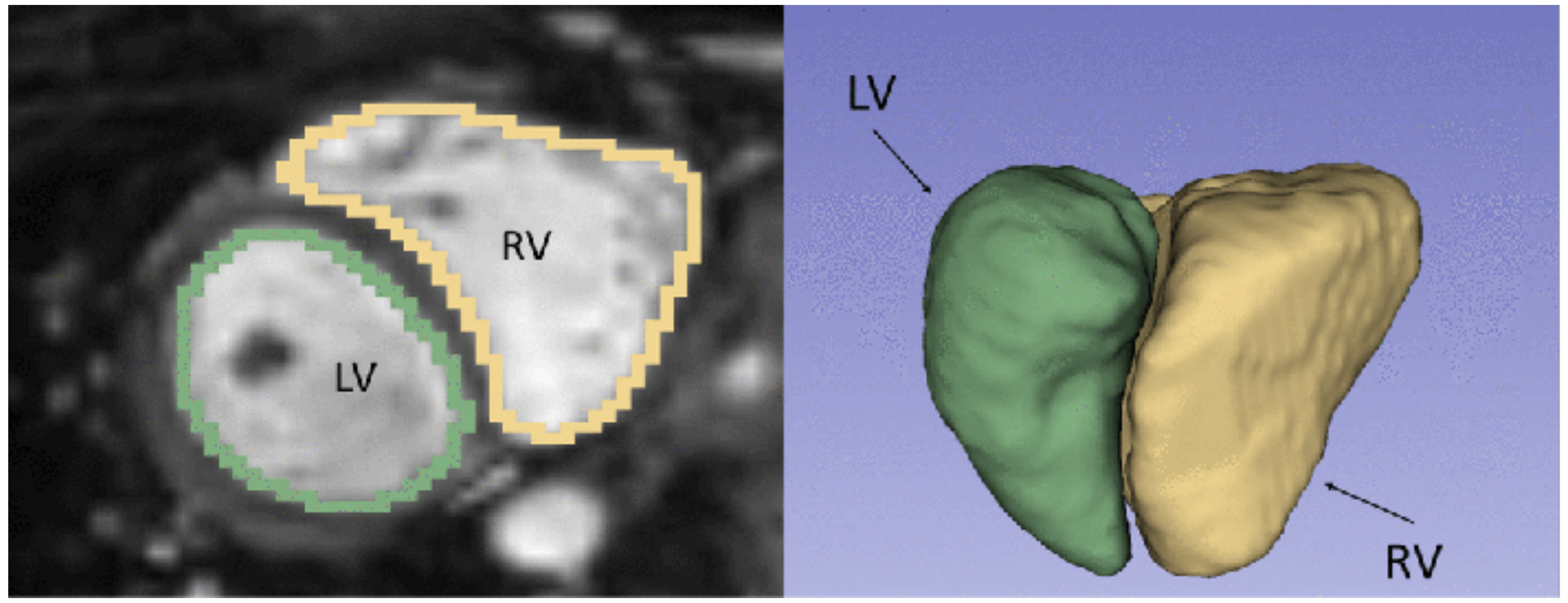
Applications: How can we catch cancer earlier?

Interpretation: How can we audit our models?

What is medical imaging?

- Medical imaging is the technique and process of imaging the interior of a body for clinical analysis and medical intervention.
- Critical for diagnosis, treatment, monitoring disease and more





Source: https://en.wikipedia.org/wiki/Medical_imaging

Source: Ruizhi “Ray” Liao

Fundamentals of medical imaging

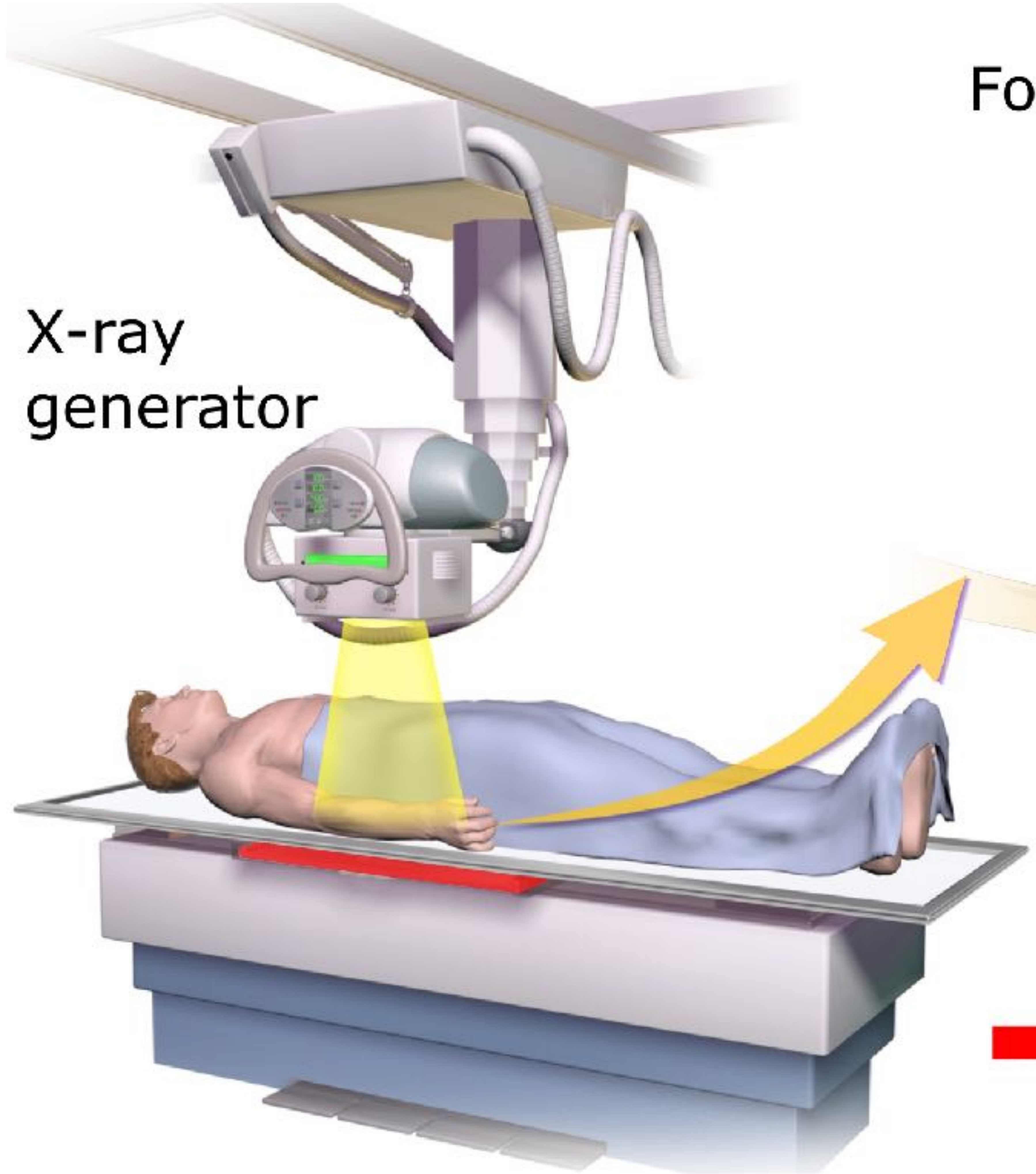
- A generator releases energy.
  - **Create “invisible light”**
- Body tissues interact with the energy.
- A receptor reconstructs an image based on the remaining/reflected energy.

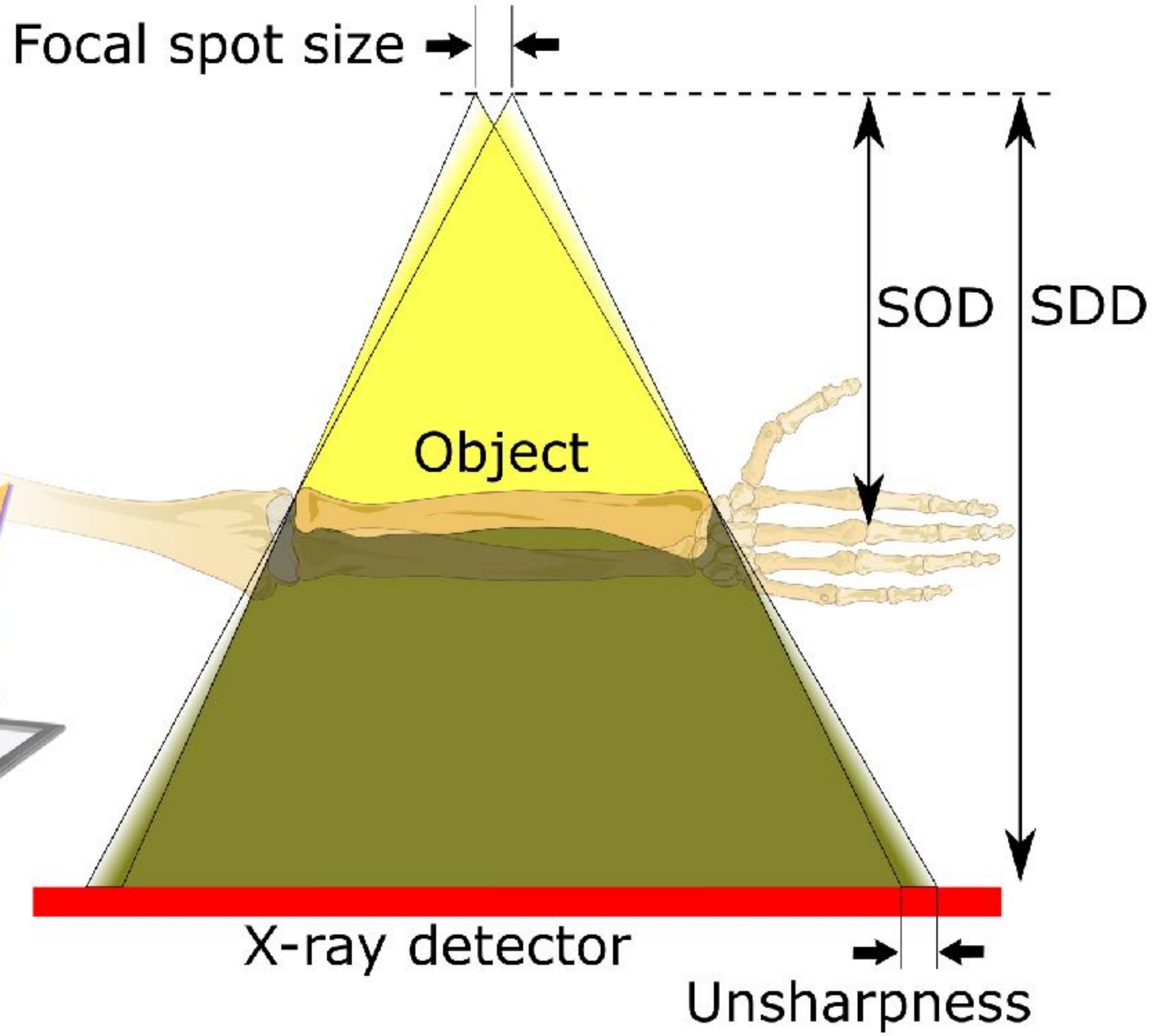
Examples:

- Projectional radiography (X-rays), CT: electromagnetic radiation
- MRI: perturbation of magnetic field
- Ultrasound: sound waves

Source: Ruizhi “Ray” Liao

Projectional radiography / x-ray





SOD: source-object distance

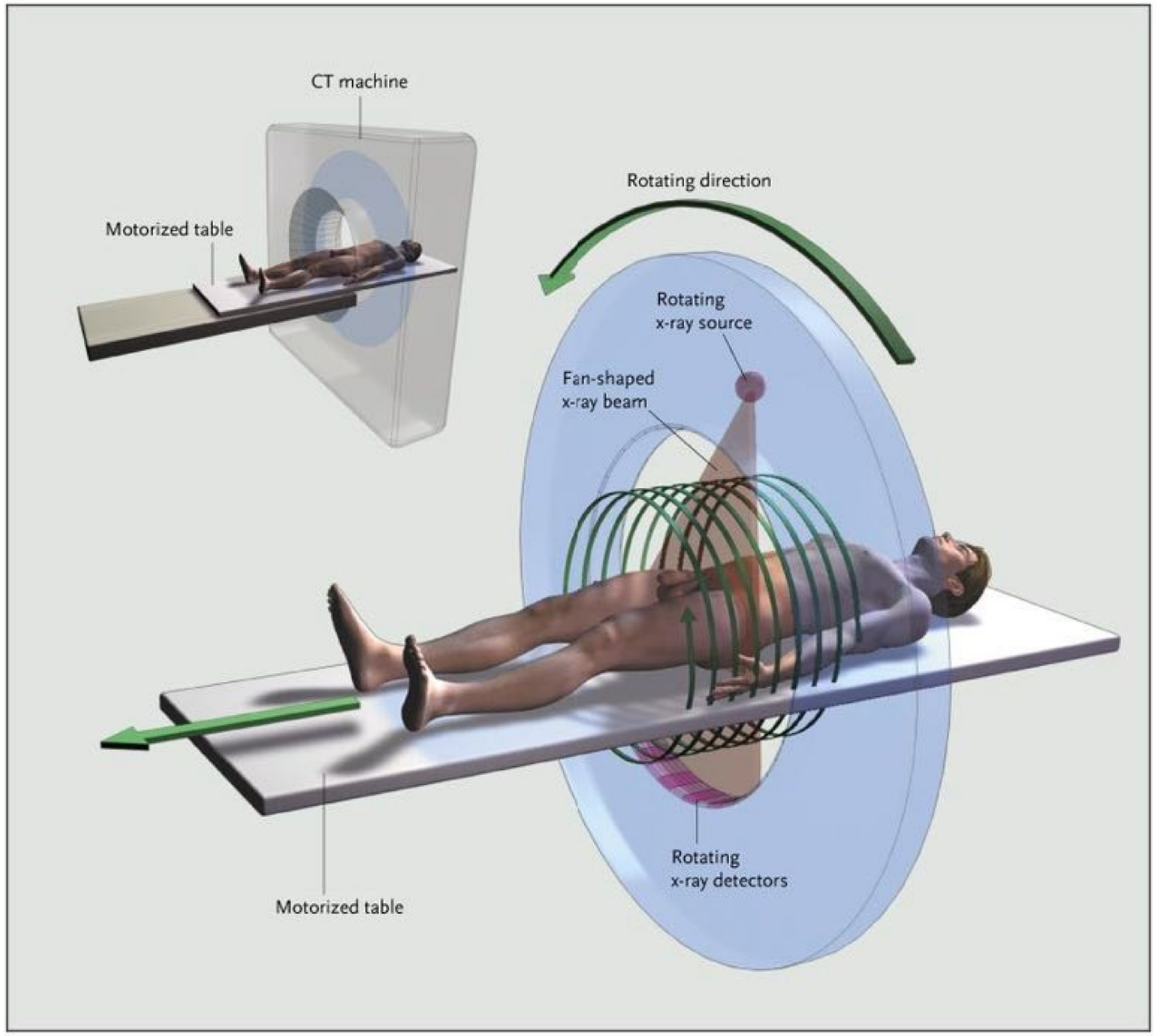
SDD: source-detector distance

Source: https://en.wikipedia.org/wiki/Projectional_radiography

Source: Ruizhi “Ray” Liao

Computed tomography (CT) scan

- Multiple x-ray measurements taken from different angles to produce tomographic images of a body.
- Advantages
  - Eliminating the superimposition of 2D x-ray images (due to its 3D nature)
  - High resolution (due to it high radiation energy)
- Adverse effect
  - One scan can have 100 to 1,000 times higher dose than 2D x-rays

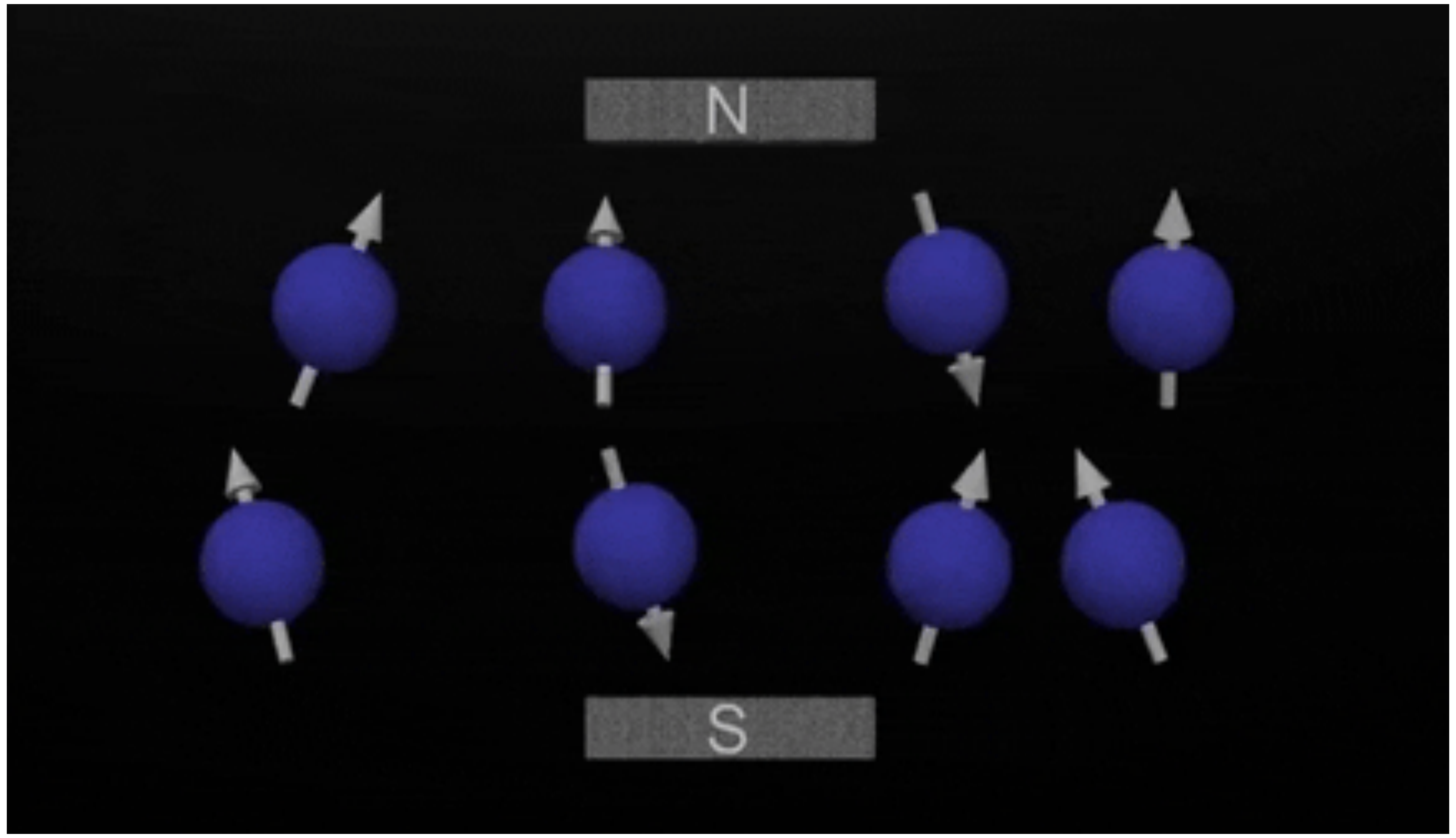


Source: https://www.nejm.org/doi/full/10.1056/nejmra072149

Source: Ruizhi “Ray” Liao

Magnetic resonance imaging (MRI)

- Nuclei with spin have a magnetic moments (spin magnetic moments). By itself, there is no energetic difference for any particular orientation of the nuclear magnet.

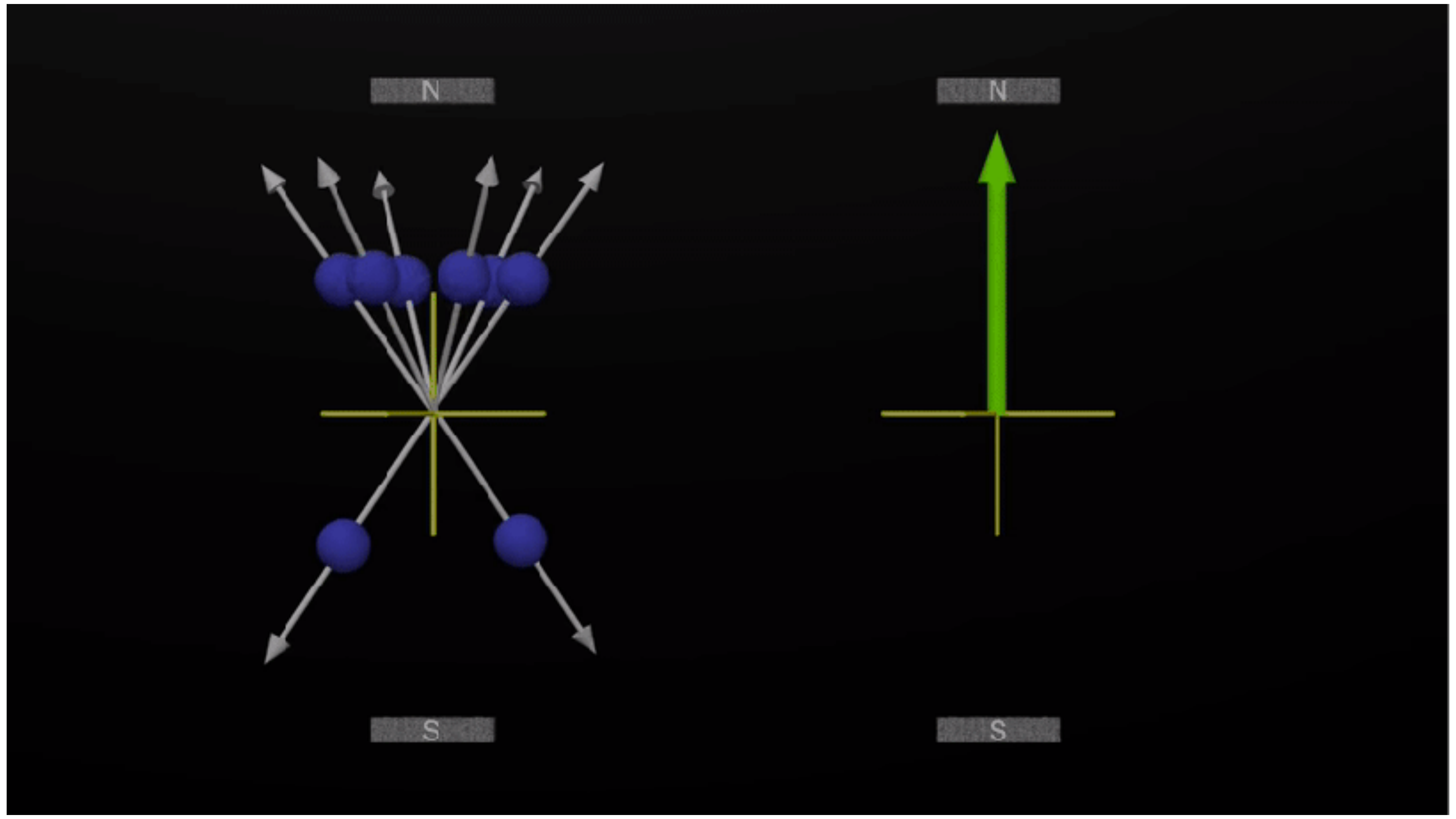


Source: https://en.wikipedia.org/wiki/Nuclear_magnetic_resonance

Source: Ruizhi "Ray" Liao

Magnetic resonance imaging (MRI)

- A strong constant magnetic field is perturbed by a weak oscillating magnetic field.
- In response, the spin orientations of nuclei perturbe from equilibrium.
- Nuclei return to their thermal equilibrium states of the spins. The receptor detects and characterizes the relaxation time to reconstruct MRI images.

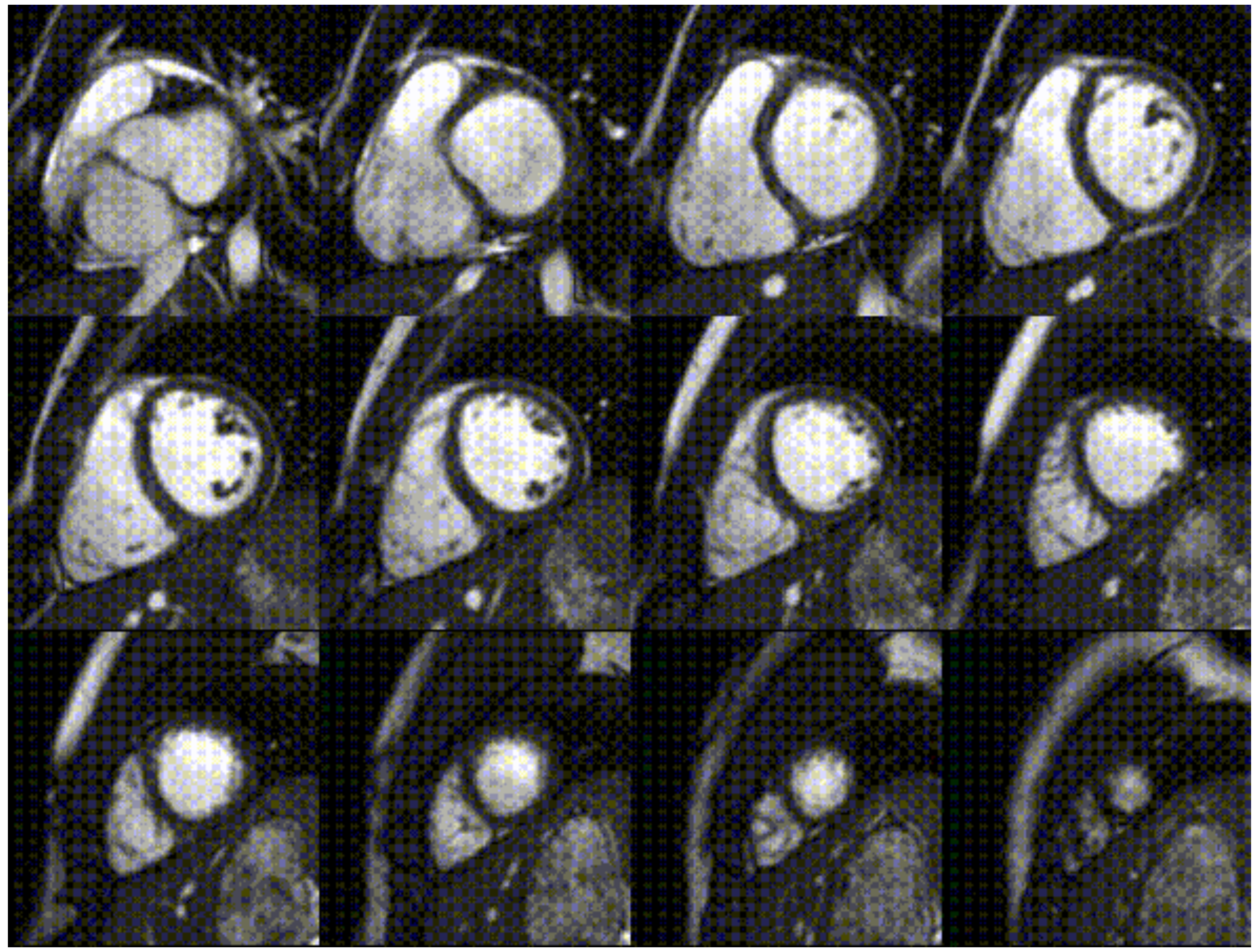


Source: https://en.wikipedia.org/wiki/Nuclear_magnetic_resonance

Source: Ruizhi “Ray” Liao

MRI captures chemical characteristics of body tissues

Tissue	T1 (ms)	T2 (ms)
Fat (adipose)	250	80
Liver	500	45
Kidney	650	60
White matter	800	90
Grey matter	900	100
Cerebrospinal fluid	2,400	280

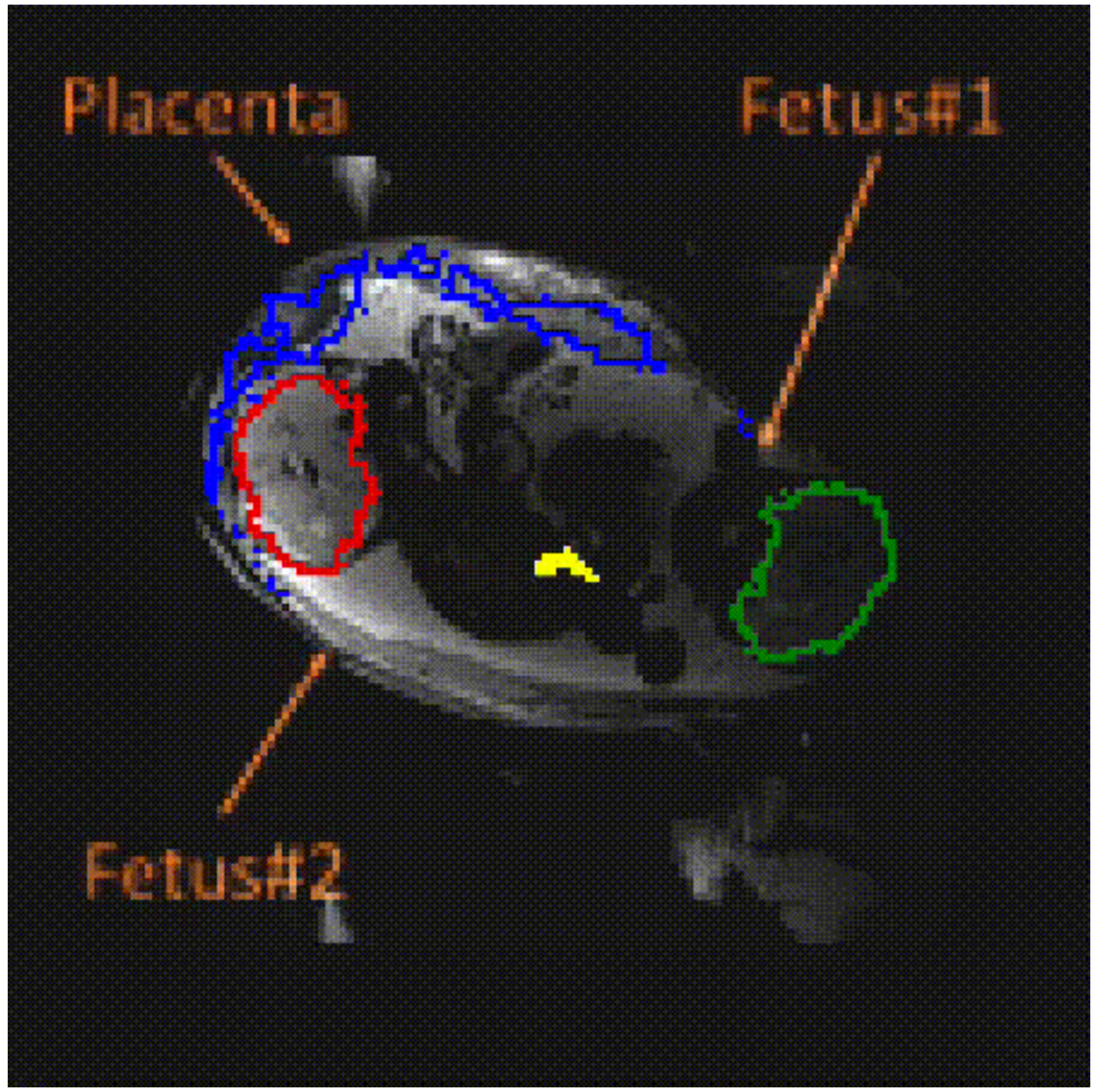


Source: http://www.ajnr.org/content/27/3/475/F1

Source: Ruizhi “Ray” Liao

Why do we need to understand medical image modalities?

- Because the intensity values of different modalities may capture different characteristics of body tissues.



MRI

Ultrasound

Liao et al, Luo et al.

Source: Ruizhi "Ray" Liao

Why model medical images?

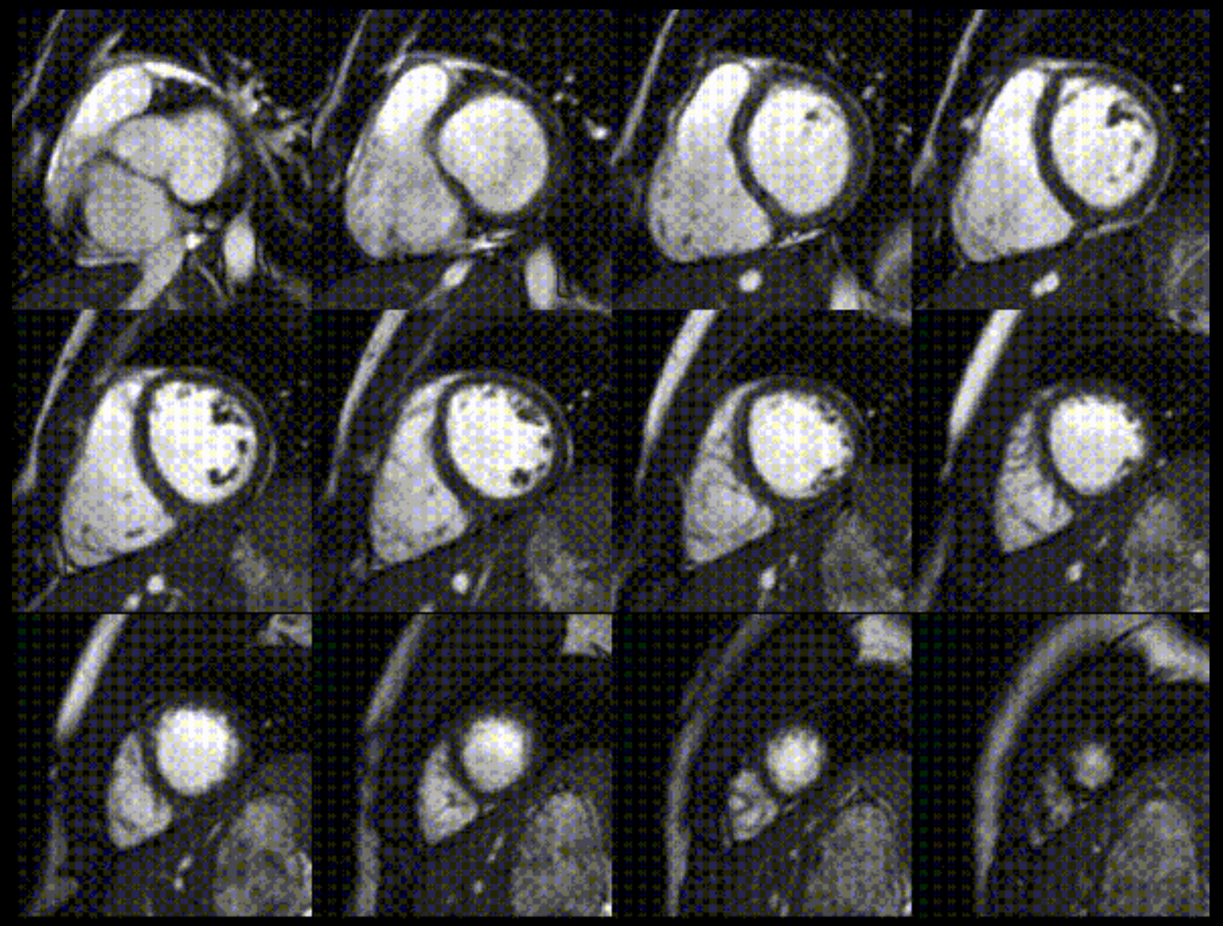
Images offer rich phenotype of tissue

Opportunities:

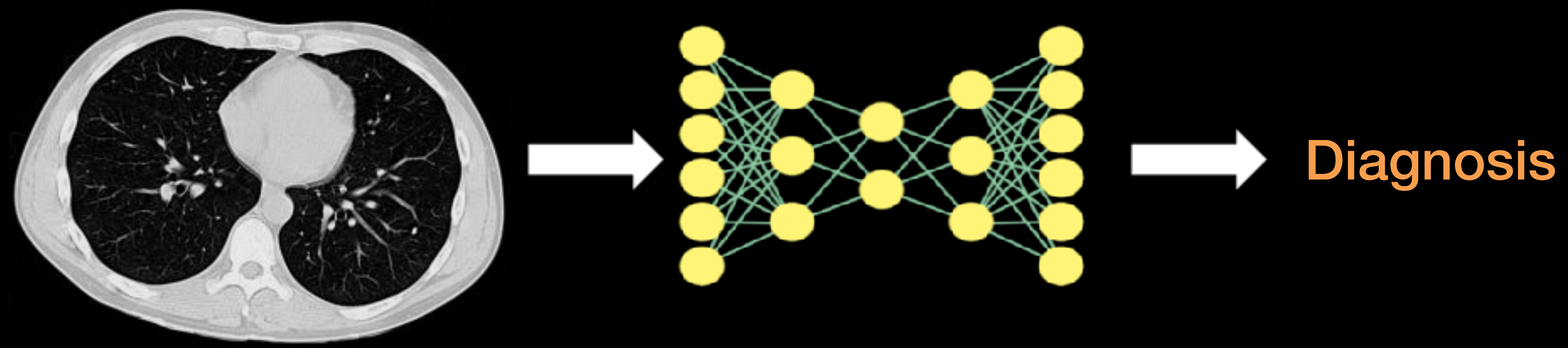
Improve diagnosis and procedures

Understanding of treatment response

Enable new treatments

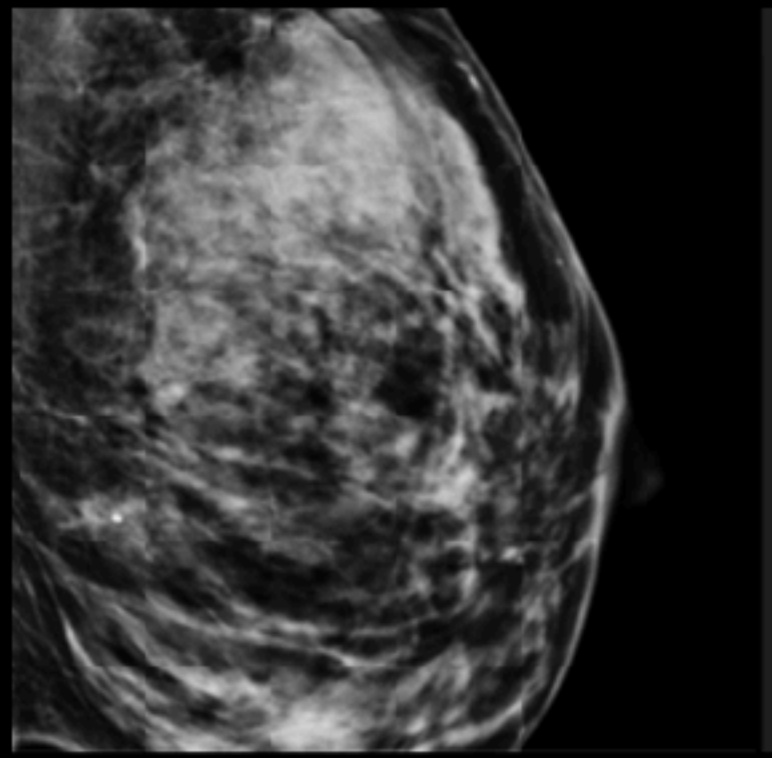


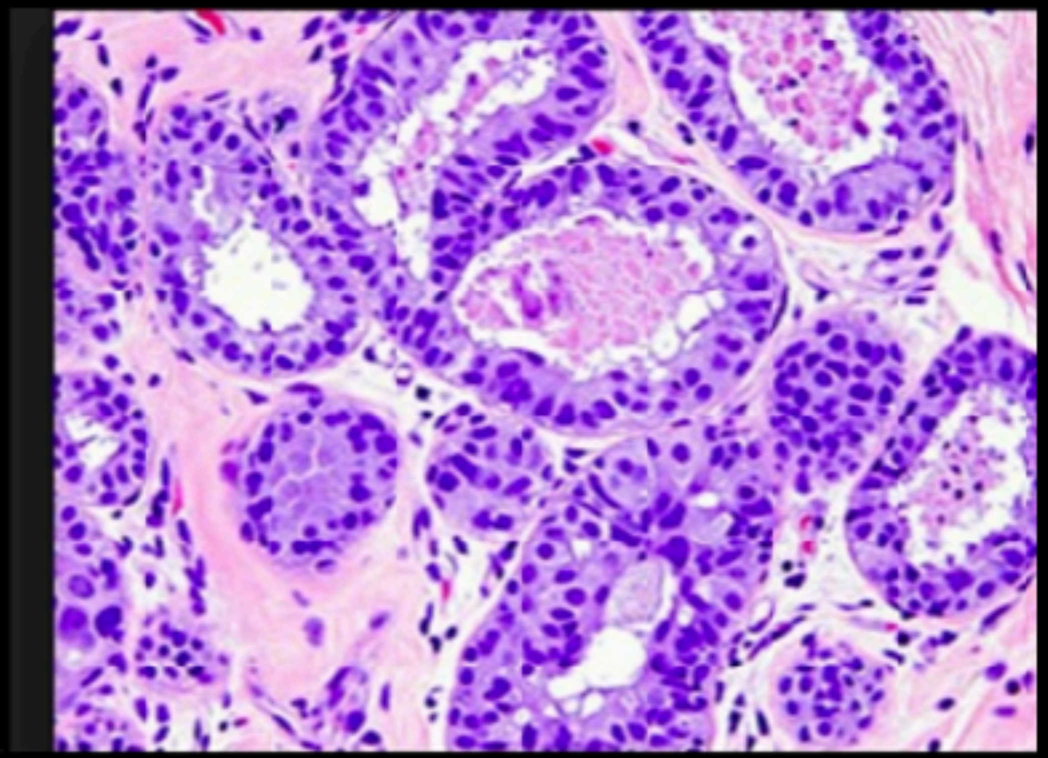
A wealth of opportunities: Lower cost

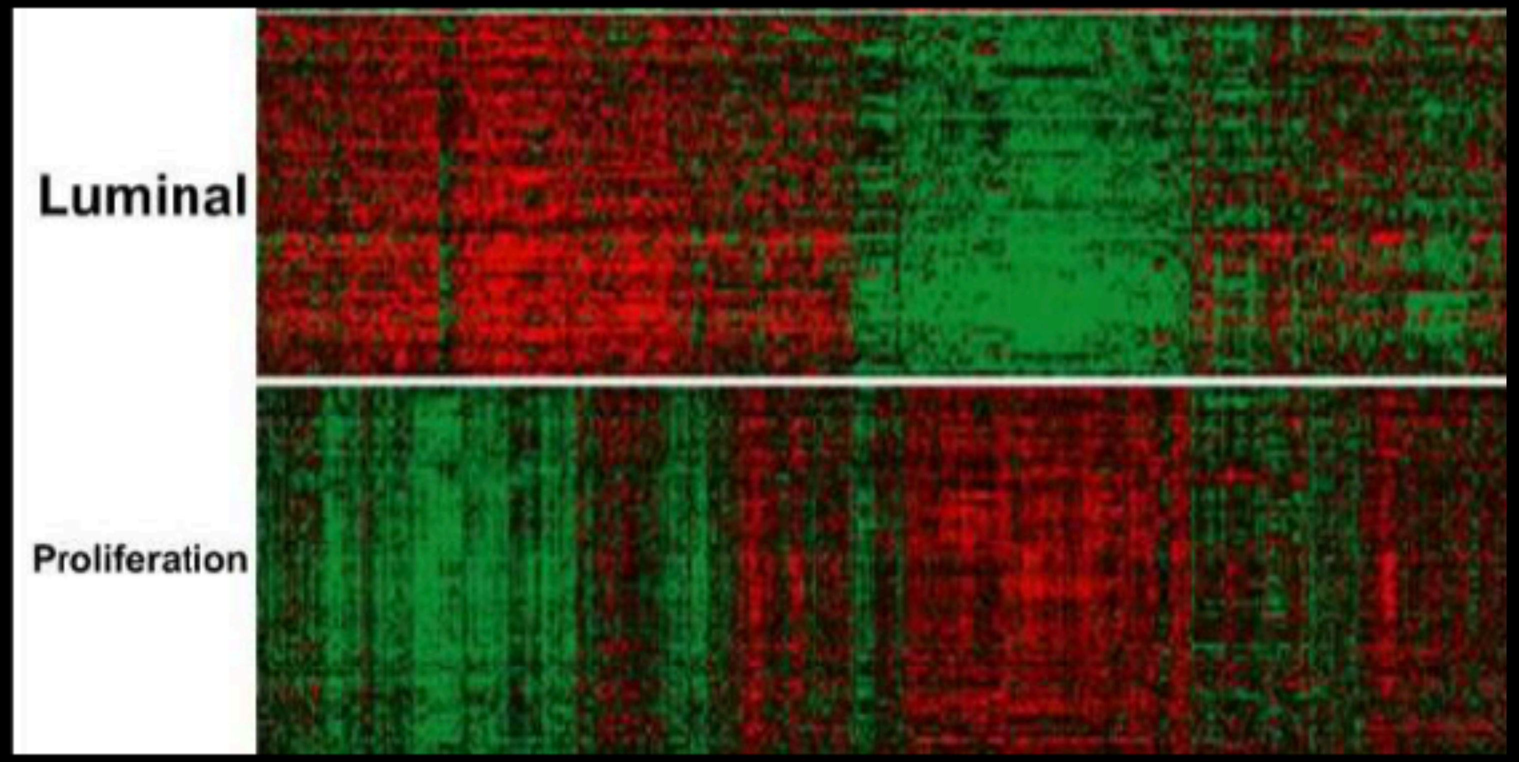


Improve global access to care

A wealth of opportunities: Better outcomes

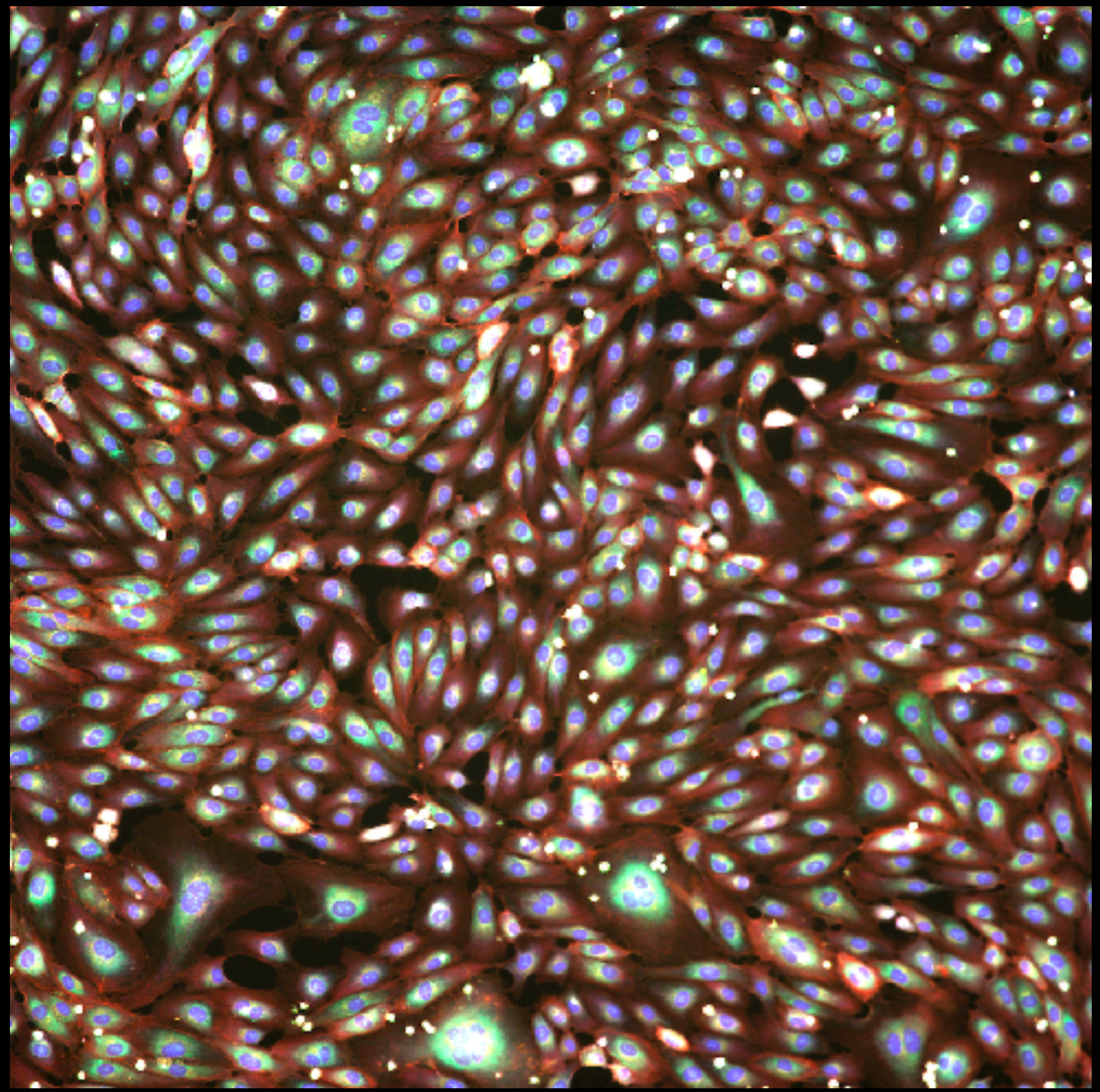






Predict Recurrences, Sensitivity to Treatment, Population at Risk

A wealth of opportunities: New discoveries



Review Article | Published: 22 December 2020

Image-based profiling for drug discovery: due for a machine-learning upgrade?

Srinivas Niranj Chandrasekaran, Hugo Ceulemans, Justin D. Boyd & Anne E. Carpenter

Identify drugs that perturb cells to "healthy" states

Agenda

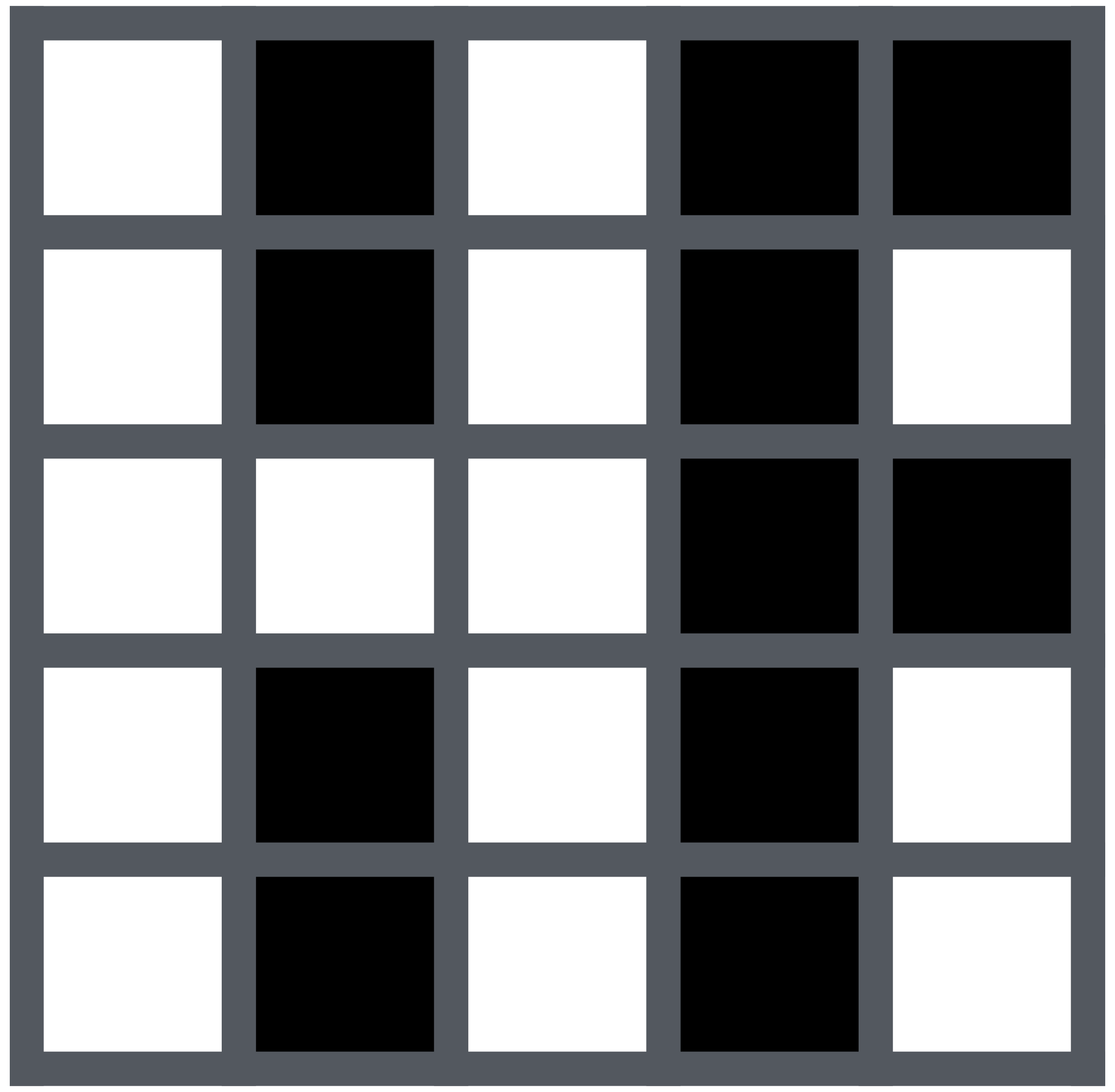
Data: What is medical imaging?

Method Foundations: How do we build models on imaging data?

Applications: How can we catch cancer earlier?

Interpretation: How can we audit our models?

Images are matrices (tensors)



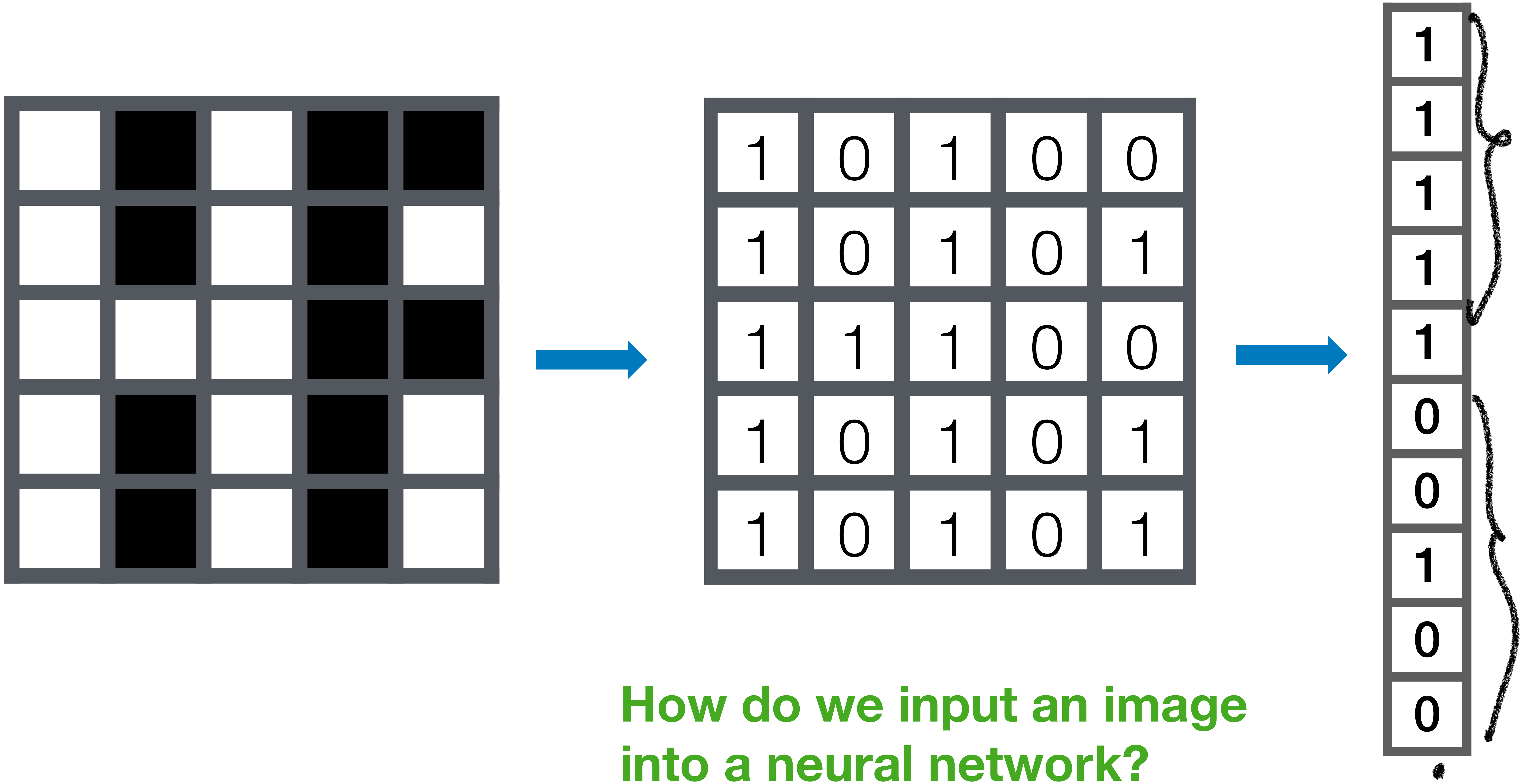
1	0	1	0	0
1	0	1	0	1
1	1	1	0	0
1	0	1	0	1
1	0	1	0	1

How do we input an image into a neural network?

...

Image credit: Tamara Broderick

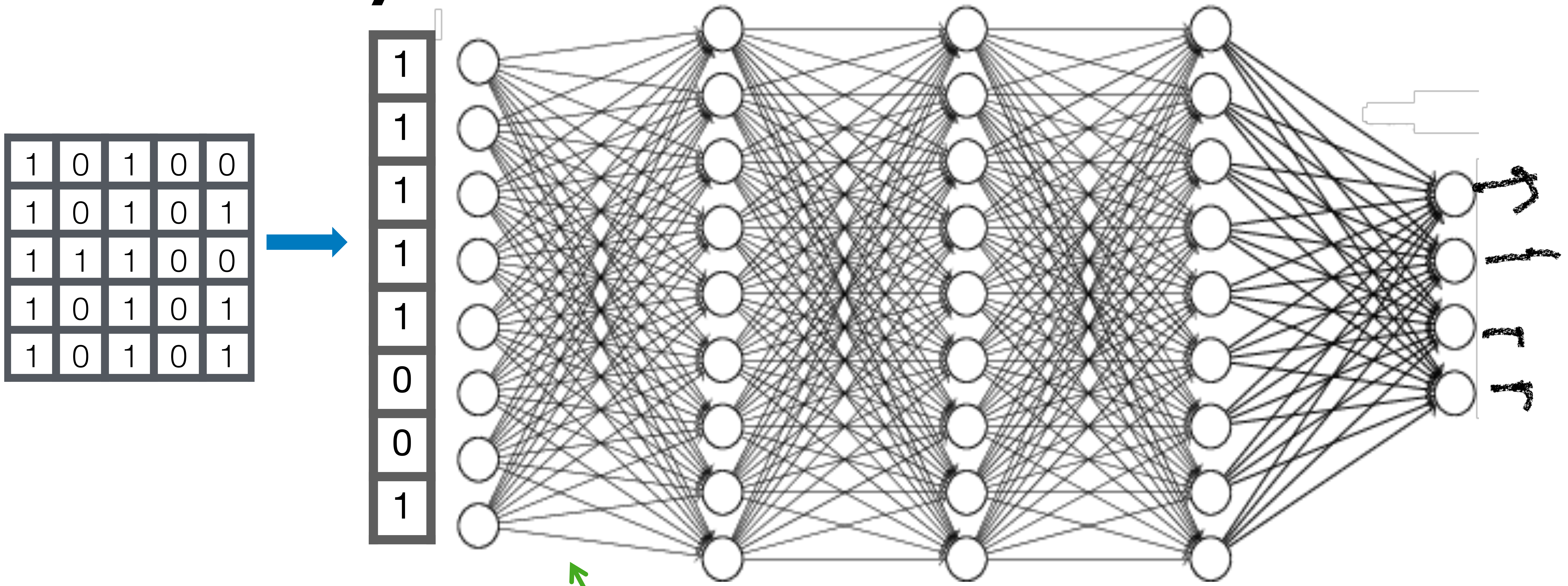
Images are matrices (tensors)



How do we input an image into a neural network?

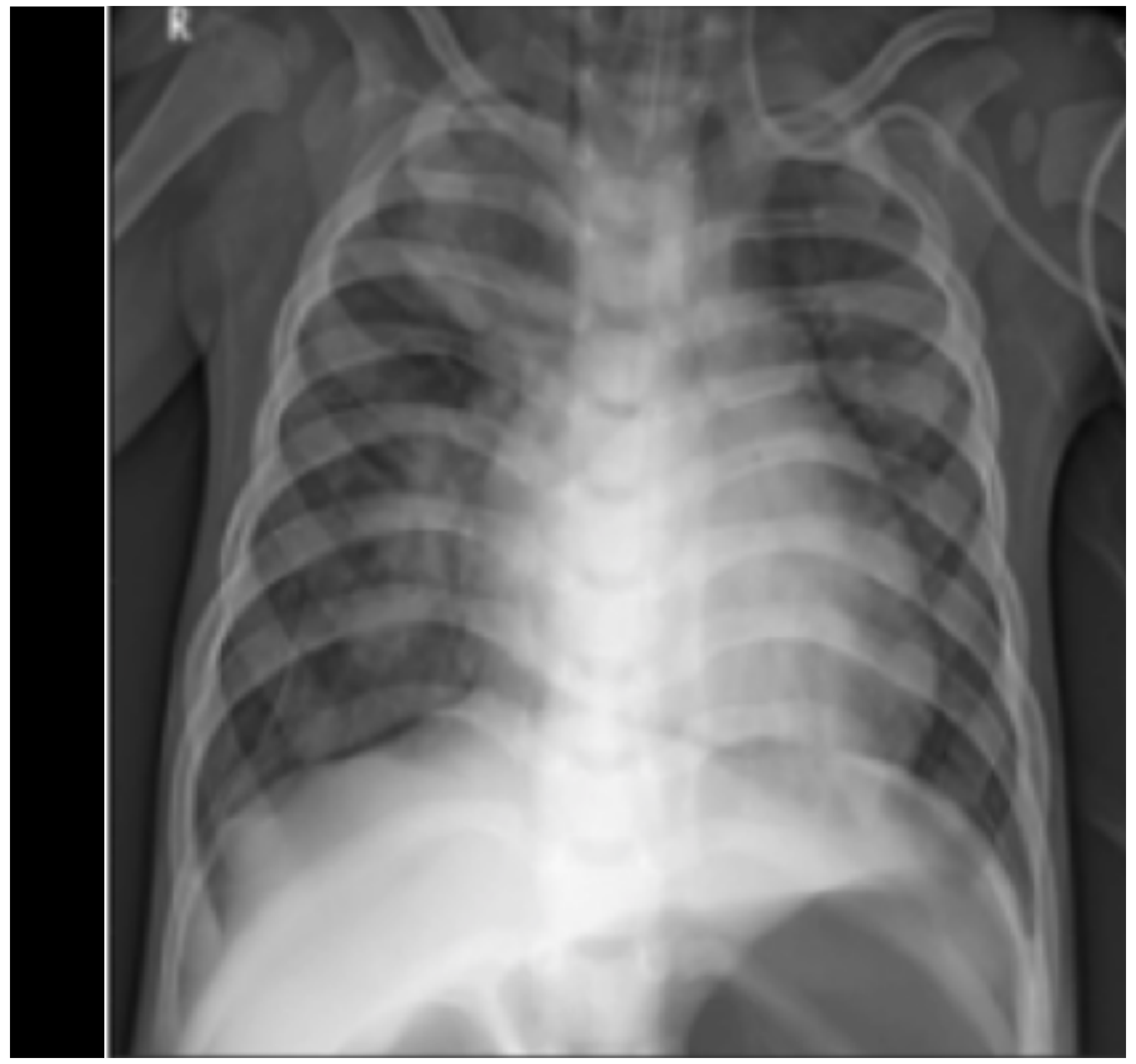
Idea: make it a vector and use FNN!

Fully connected network?



Fully connected layer: every input is connected to every output

Fully connected network?

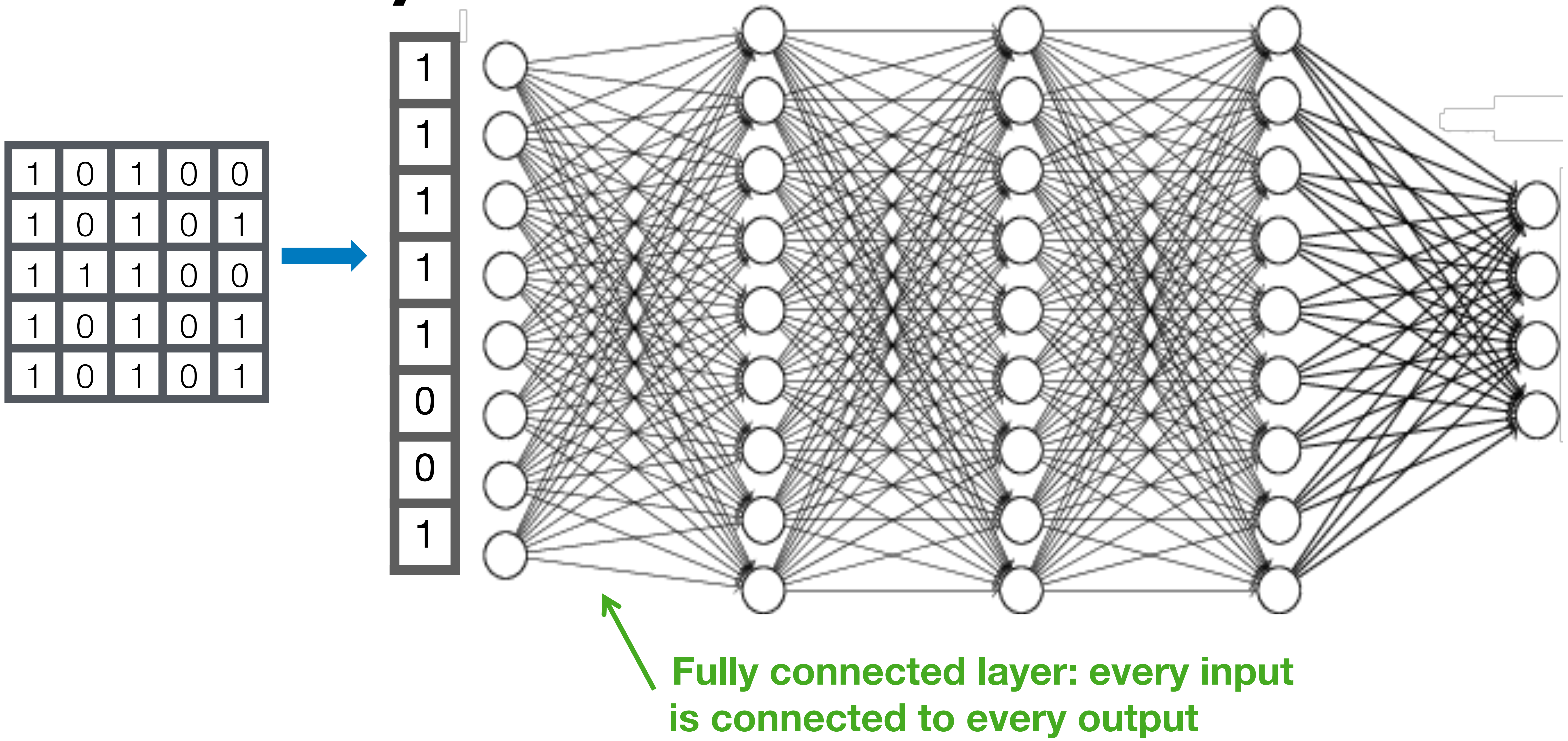


$x^0 = [0,1,1\ ,\ \ldots\ ,\ 0]$

$x^1 = [0,0,1,\ \ldots\ ,\ 0]$

Small padding, very different feature vectors!

Fully connected network?



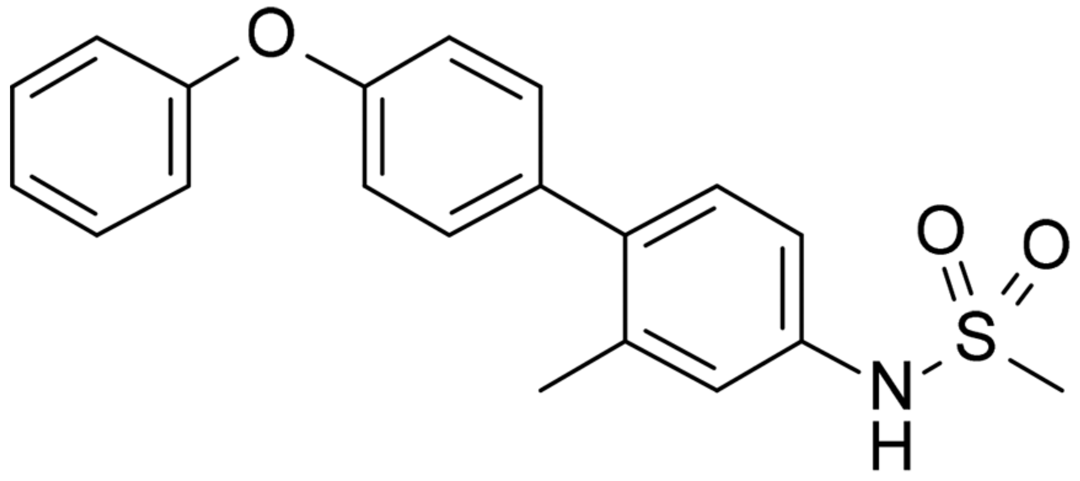
Example:
200*200 image
40K hidden units
2B parameters

Advanced Neural networks

- We need specialized architectures/approaches to effectively handle complex objects
  - RNNs, CNNs, GNNs, transformers, etc.

text / sequences (RNNs, transformers)

images / video (CNNs)



graphs (GNNs)

Locality and translation invariance









Desiderata

- Capture spatial dependencies: pixel position and neighborhood have semantic meaning
- Handle Translations: Elements of interest can appear anywhere in the image
- Robustly scale for large images

CNN key ideas

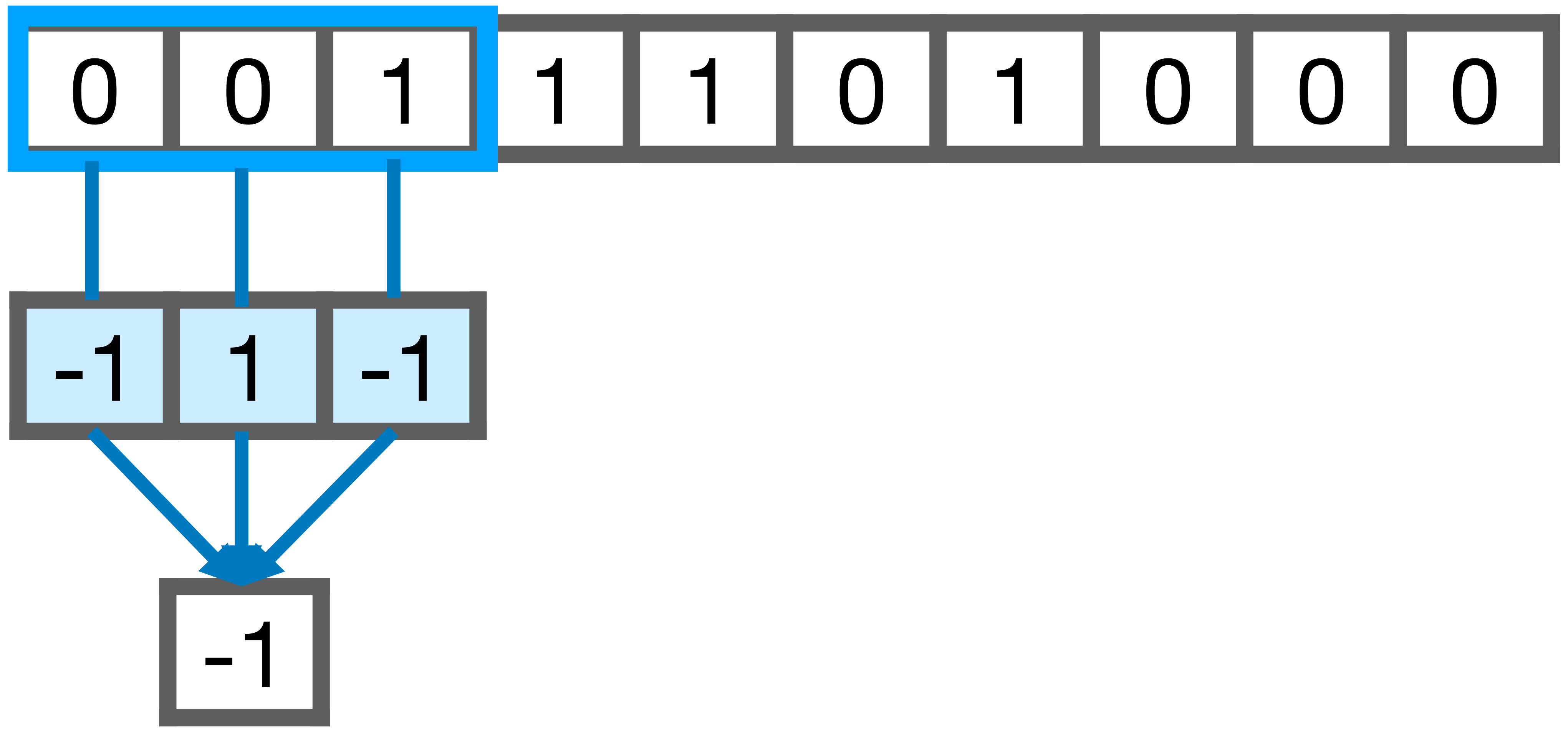
- **Capture spatial dependencies: convolutions**
  - spatial locality

- Handle Translations: pooling
  - abstract away locality

- Robustly scale for large images: weight sharing
  - apply the same detector to all the patches

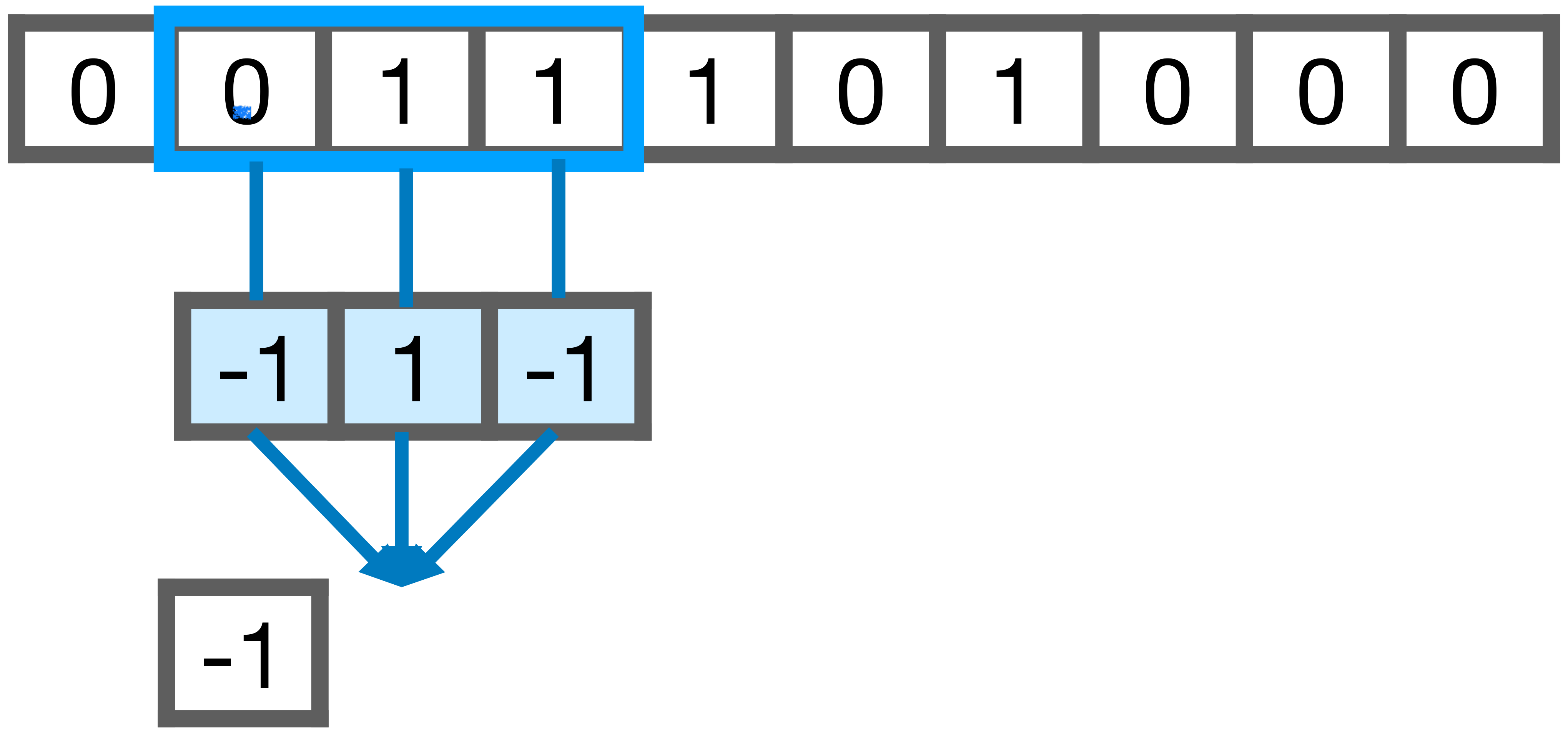
Convolutional Layer: 1D example

- 1D image
- Filter
- After convolution



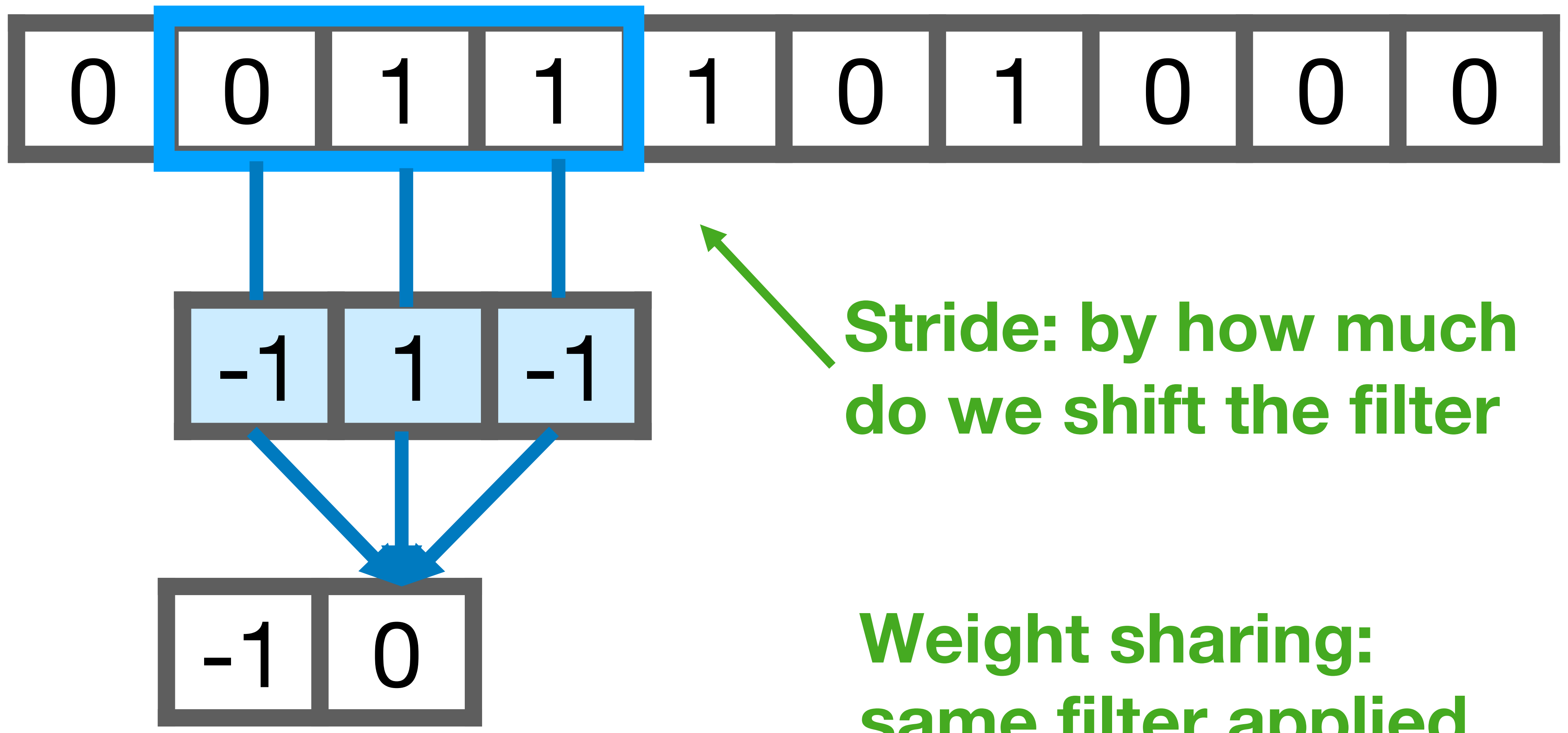
Convolutional Layer: 1D example

- 1D image
- Filter
- After convolution



Convolutional Layer: 1D example

- 1D image
- Filter
- After convolution

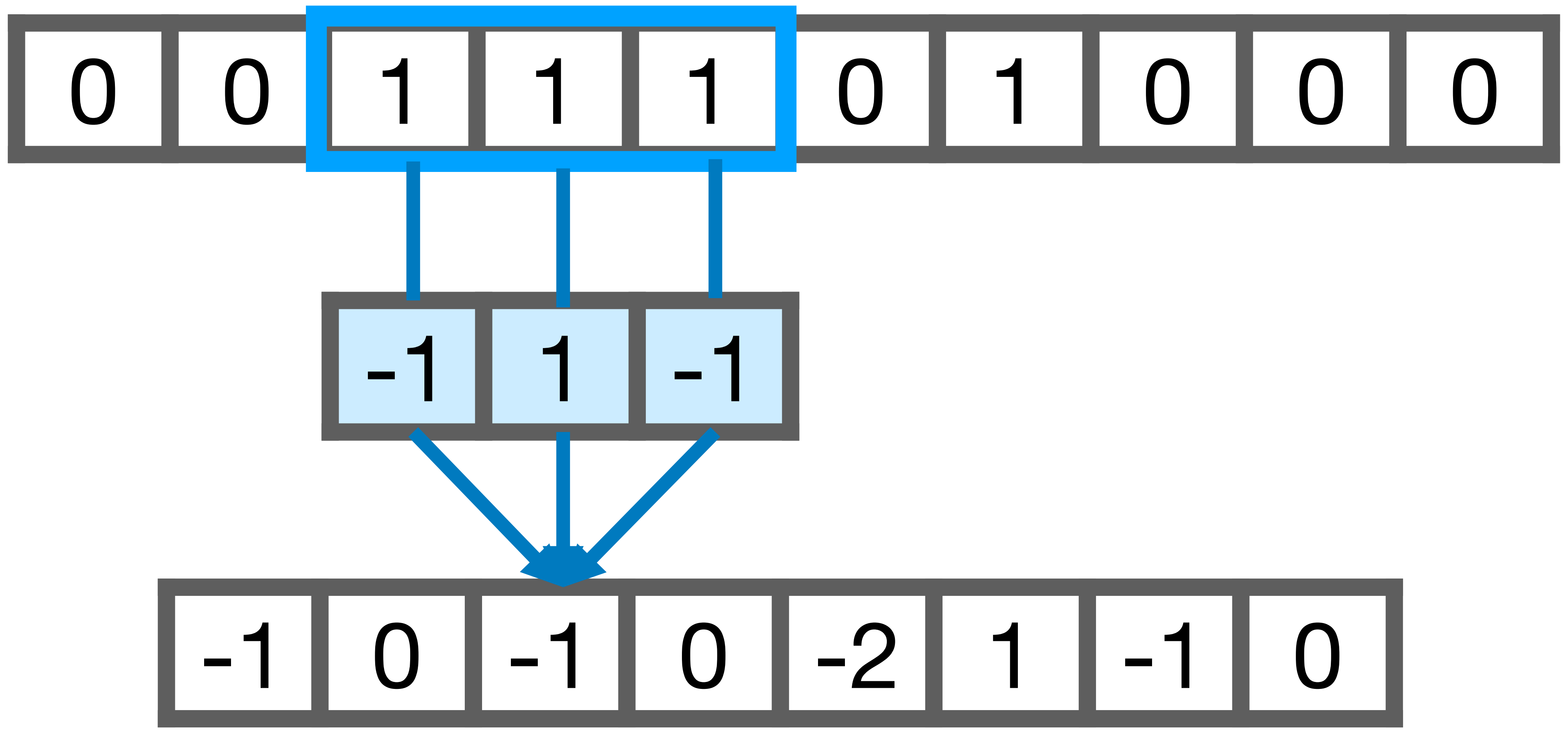


Stride: by how much do we shift the filter

Weight sharing: same filter applied to both (all) patches

Convolutional Layer: 1D example

- 1D image
- Filter
- After convolution



Big advantage: due to weight sharing, needs much fewer weights than a fully connected network

Convolutional Layer: 1D example

- 1D image

0	0	1	1	1	0	1	0	0	0

- Filter

-1	1	-1

- After convolution

-1	0	-1	0	-2	1	-1	0

- After ReLu

Padding

- 1D image

0	0	1	1	1	0	1	0	0	0

- Filter

-1	1	-1

- After convolution

-1	0	-1	0	-2	1	-1	0

- After ReLu

0	0	0	0	0	1	0	0

Output is smaller! (why?)

Remedy: pad input with zeros

Padding

- 1D image

0	0	0	1	1	1	0	1	0	0	0	0

- Filter

-1	1	-1

- After convolution

0	-1	0	-1	0	-2	1	-1	0	0

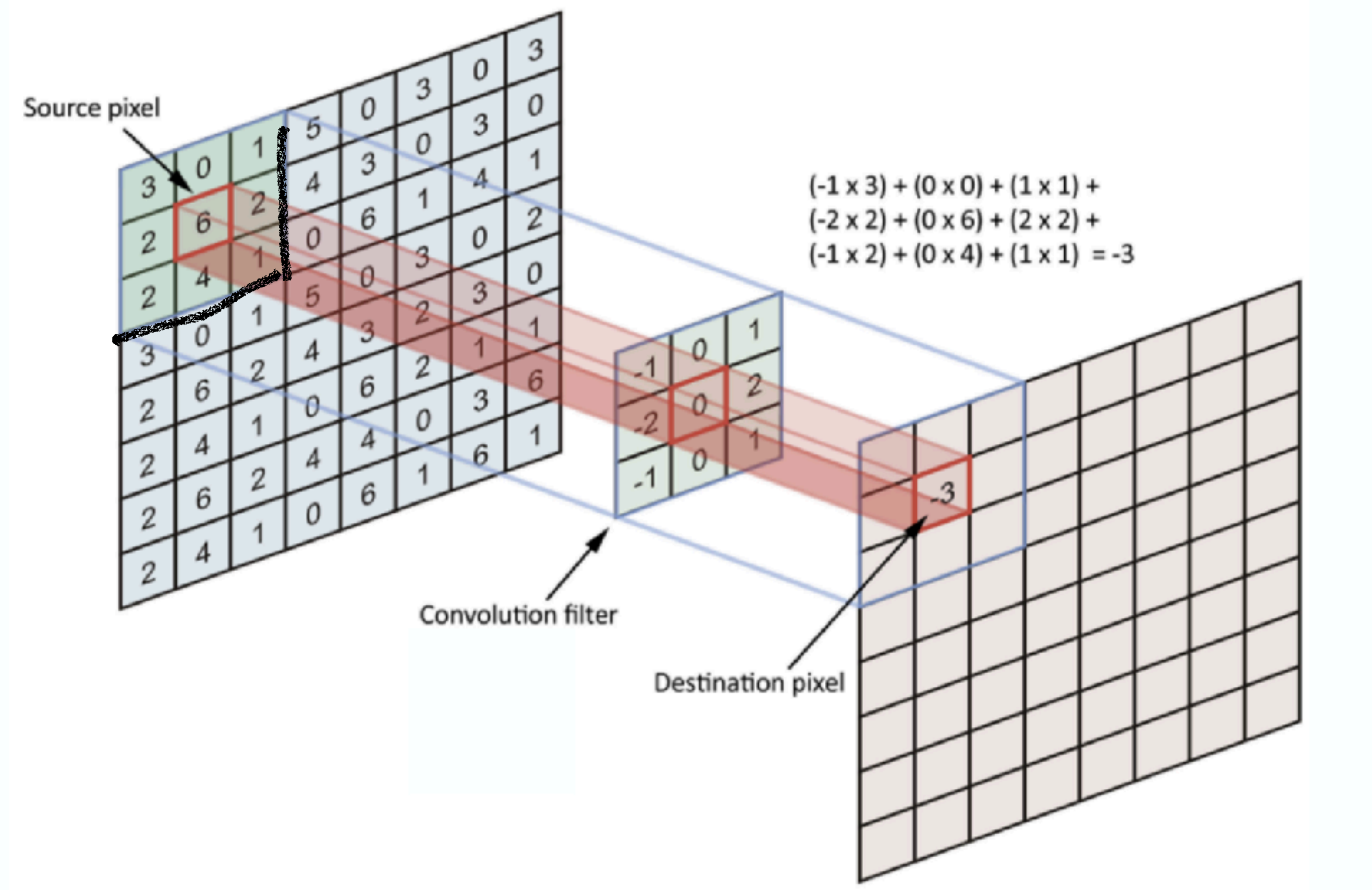
- After ReLu

0	0	0	0	0	0	1	0	0	0

Output is smaller! (why?)

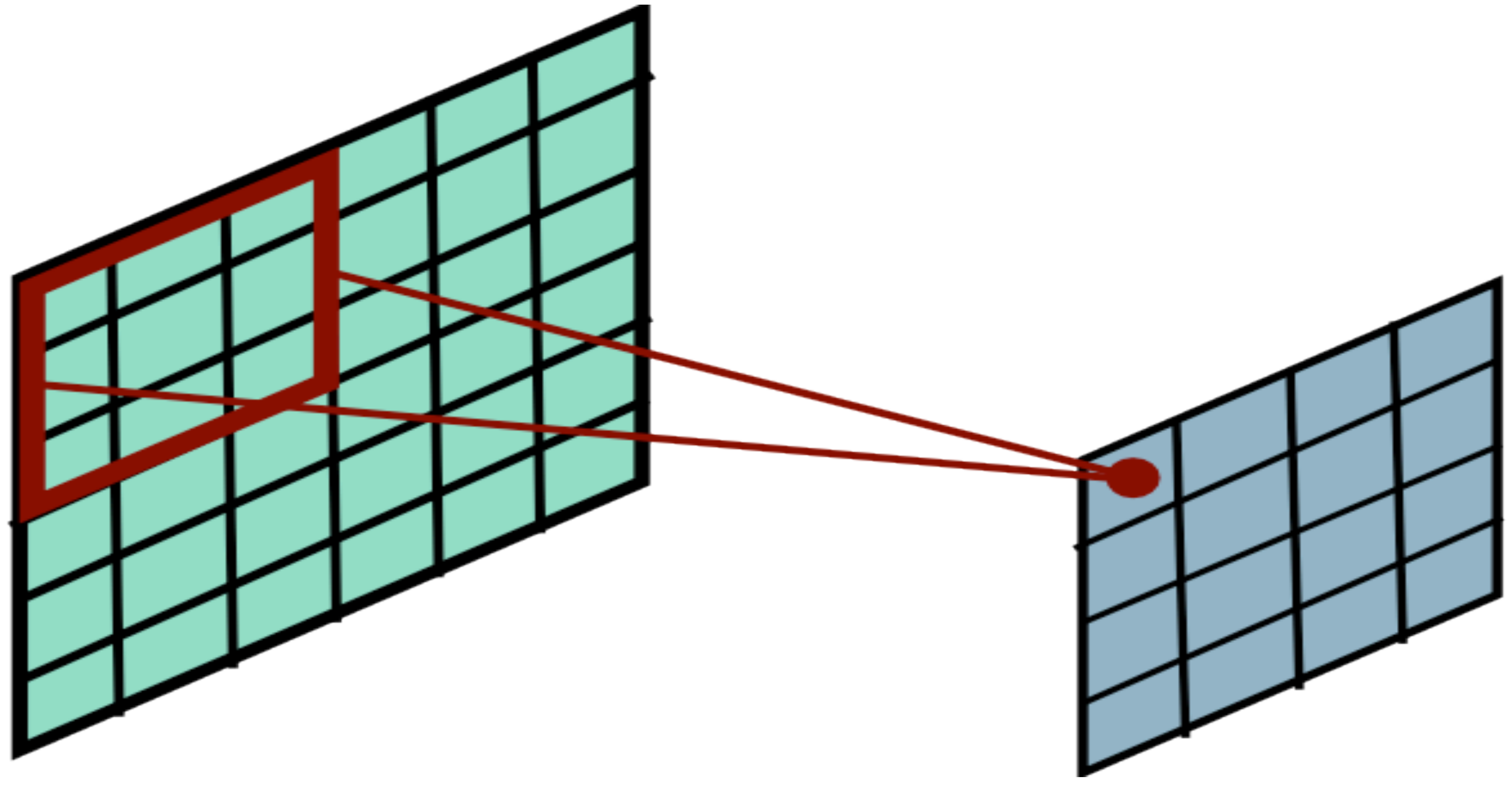
Remedy: pad input with zeros

2D Convolutions



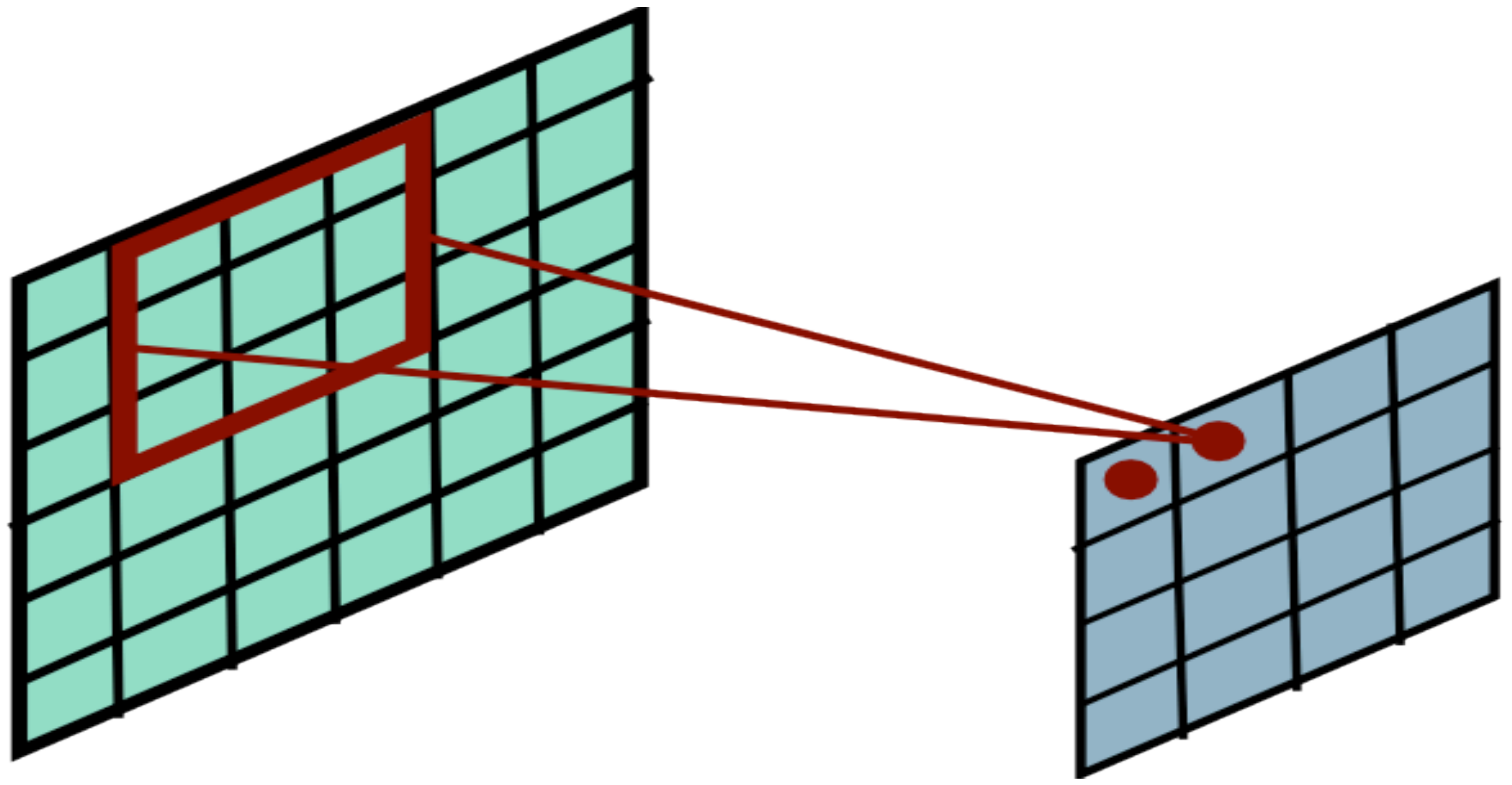
The convolution operation.

Convolutional Layer

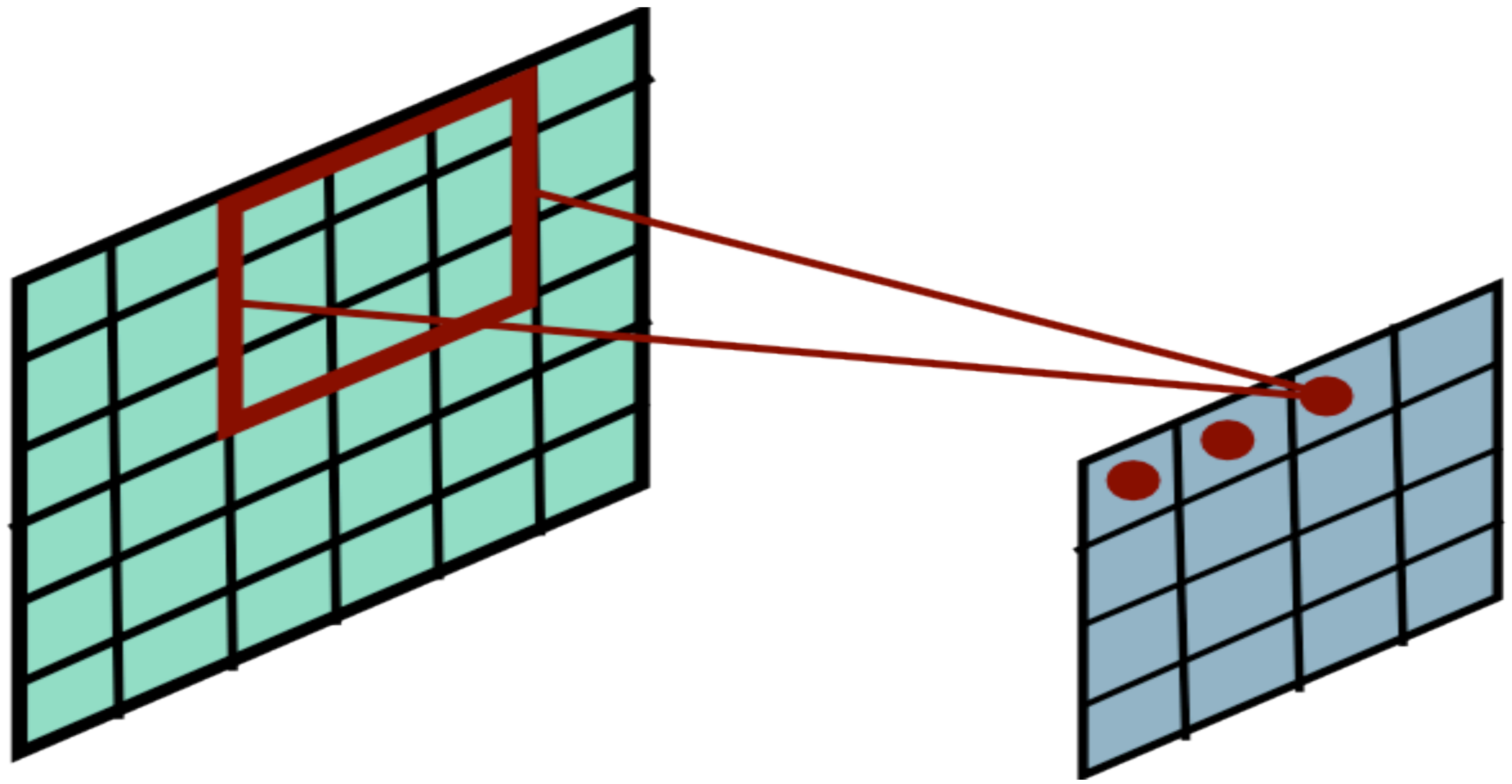


activation map.

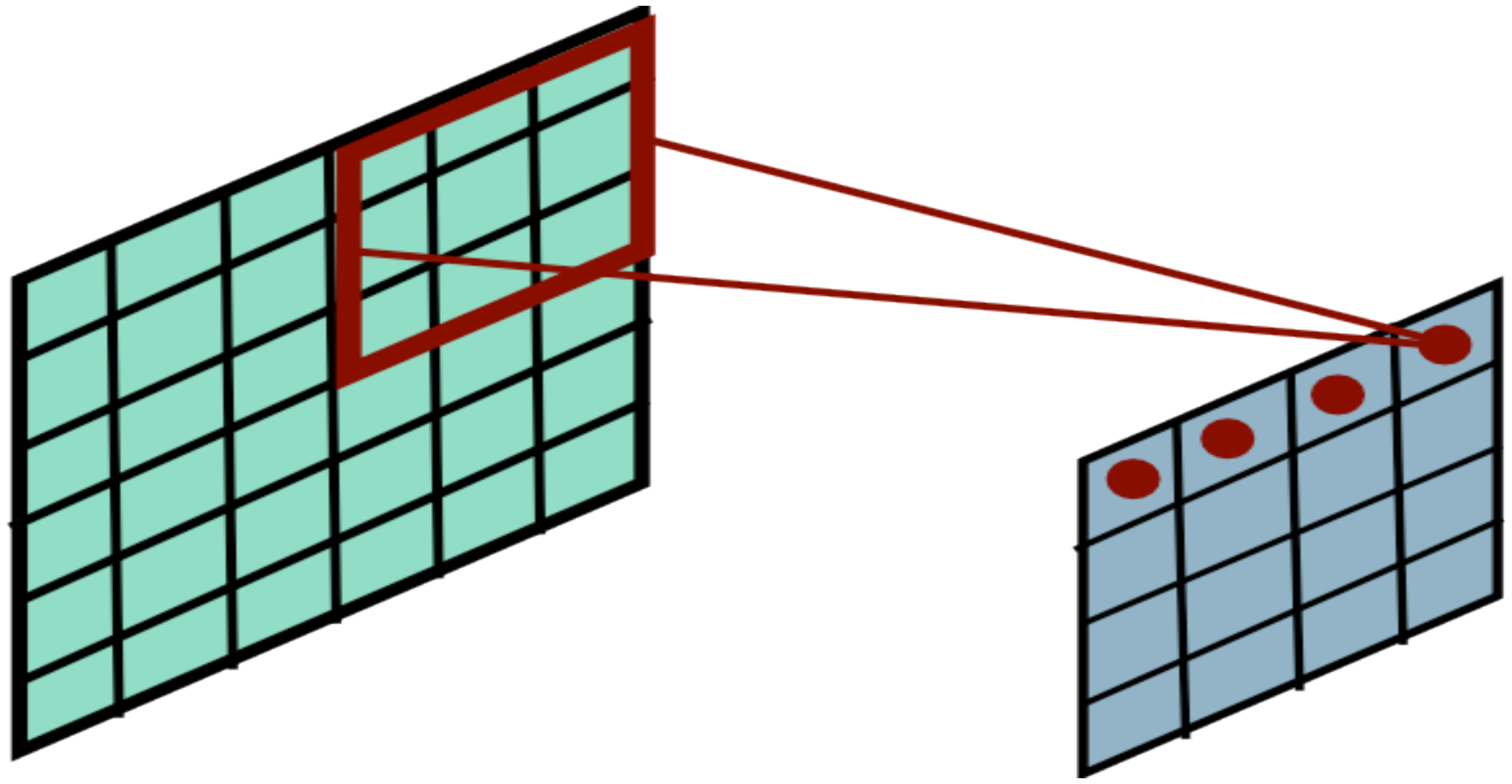
Convolutional Layer



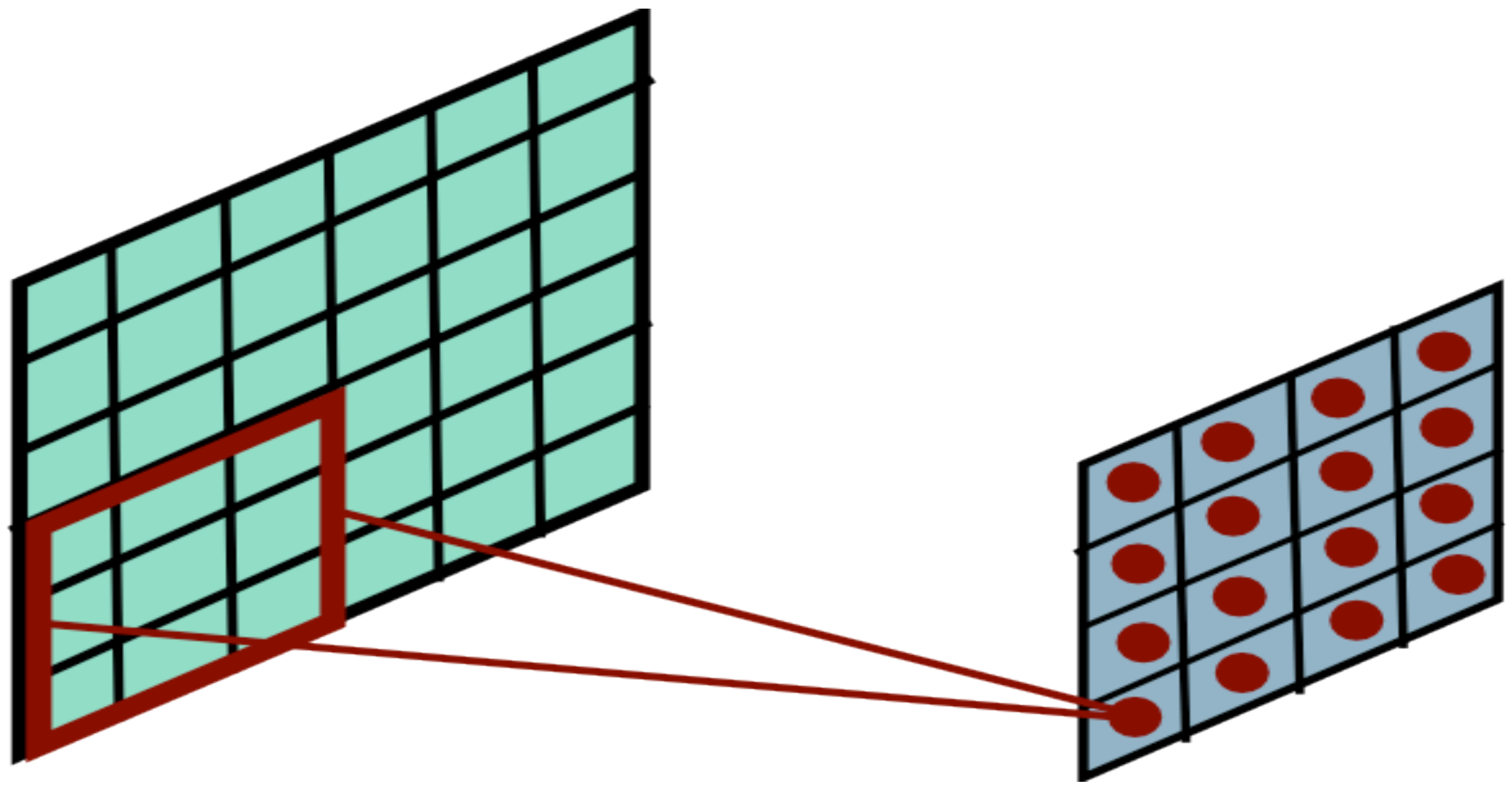
Convolutional Layer



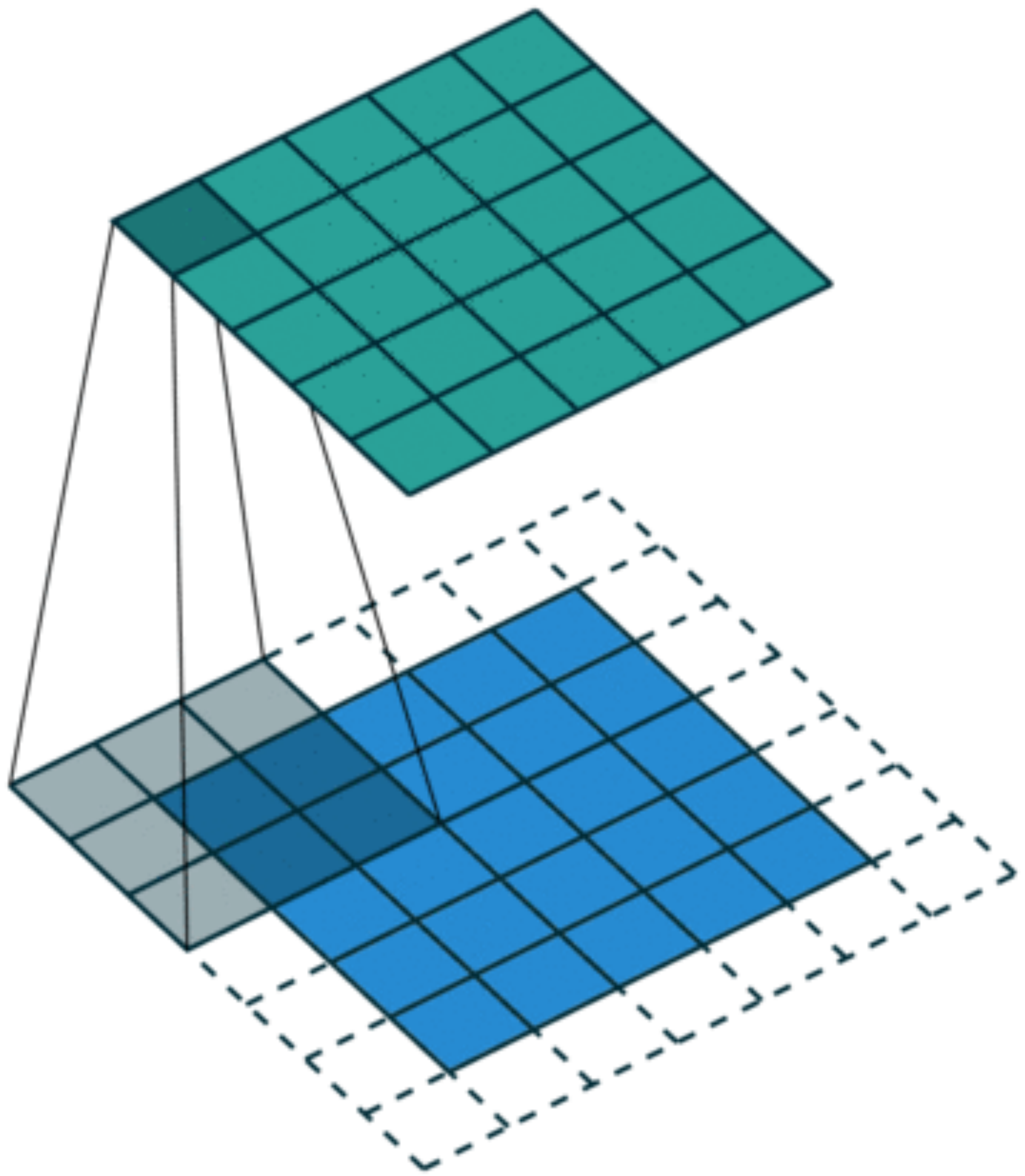
Convolutional Layer



Convolutional Layer

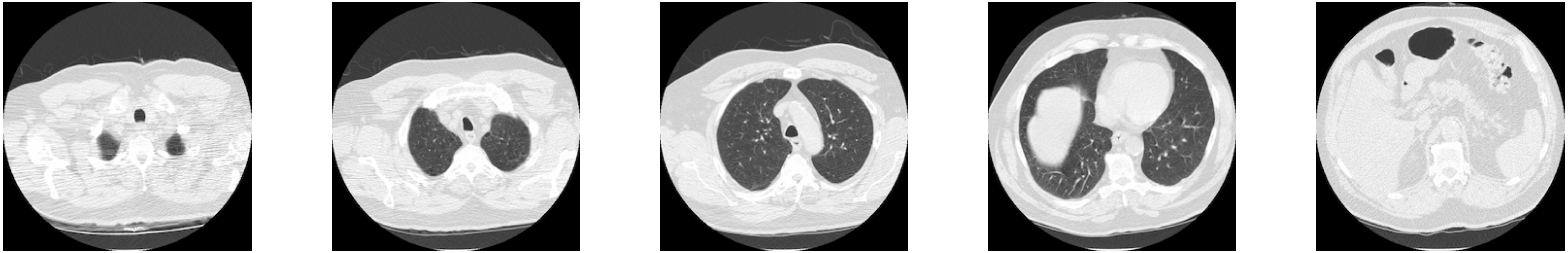


Convolution with Padding

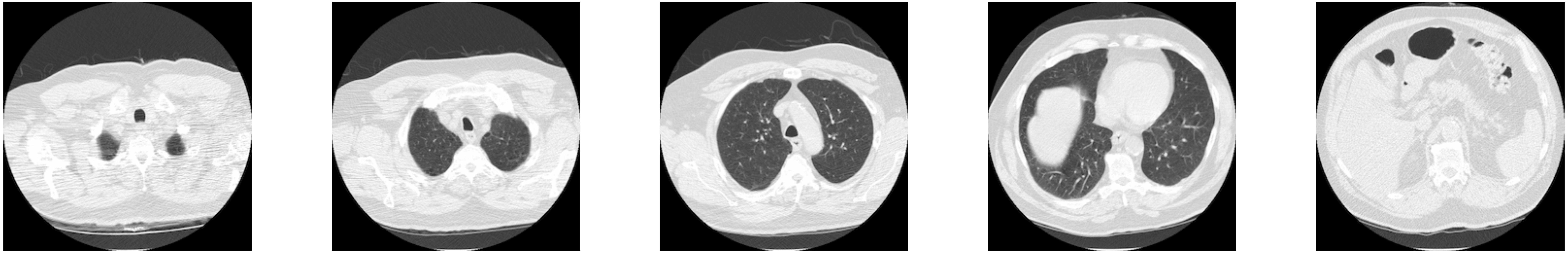


source: Theano, deeplearning.net

Question: How would you apply this idea to a CT-scan?



Question: How would you apply this idea to a CT-scan?



3D Convolutions

Examples of Convolutions

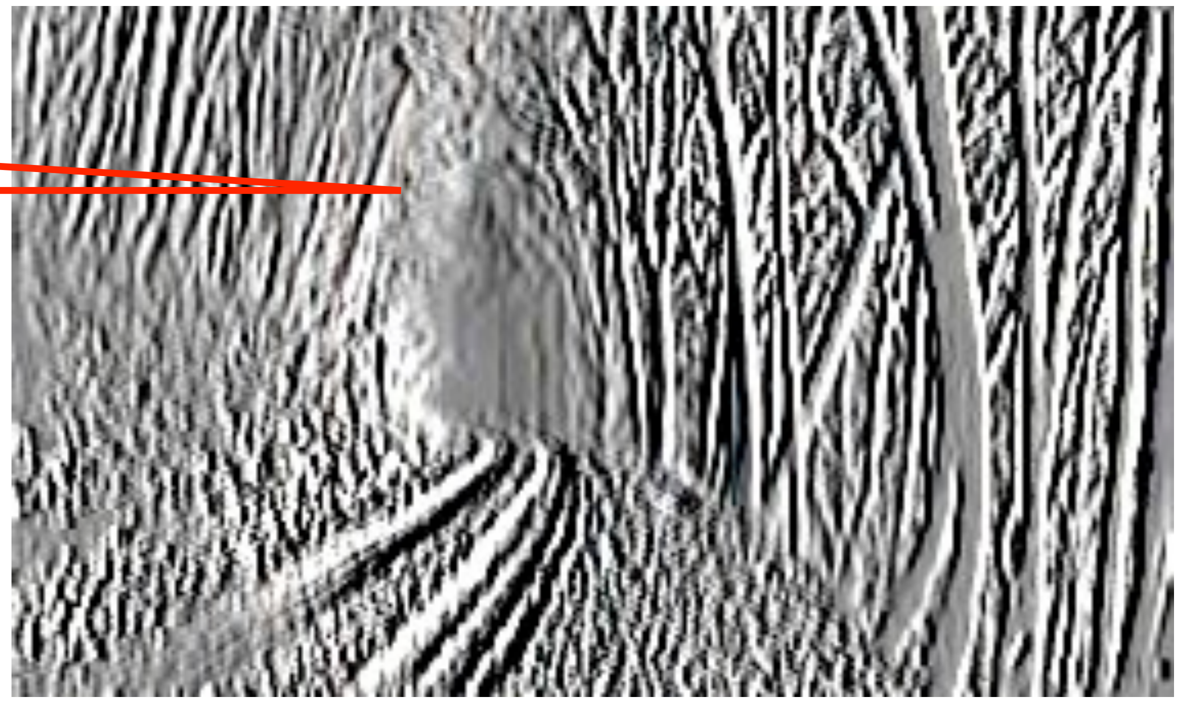
filter
mask

$$
\begin{bmatrix} 10 & 10 & 10 & 0 & 0 & 0 \\ 10 & 10 & 10 & 0 & 0 & 0 \\ 10 & 10 & 10 & 0 & 0 & 0 \\ 10 & 10 & 10 & 0 & 0 & 0 \\ 10 & 10 & 10 & 0 & 0 & 0 \\ 10 & 10 & 10 & 0 & 0 & 0 \end{bmatrix} * \begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix} = \begin{bmatrix} 0 & 30 & 30 & 0 \\ 0 & 30 & 30 & 0 \\ 0 & 30 & 30 & 0 \\ 0 & 30 & 30 & 0 \end{bmatrix}
$$

input

Examples of Convolutions

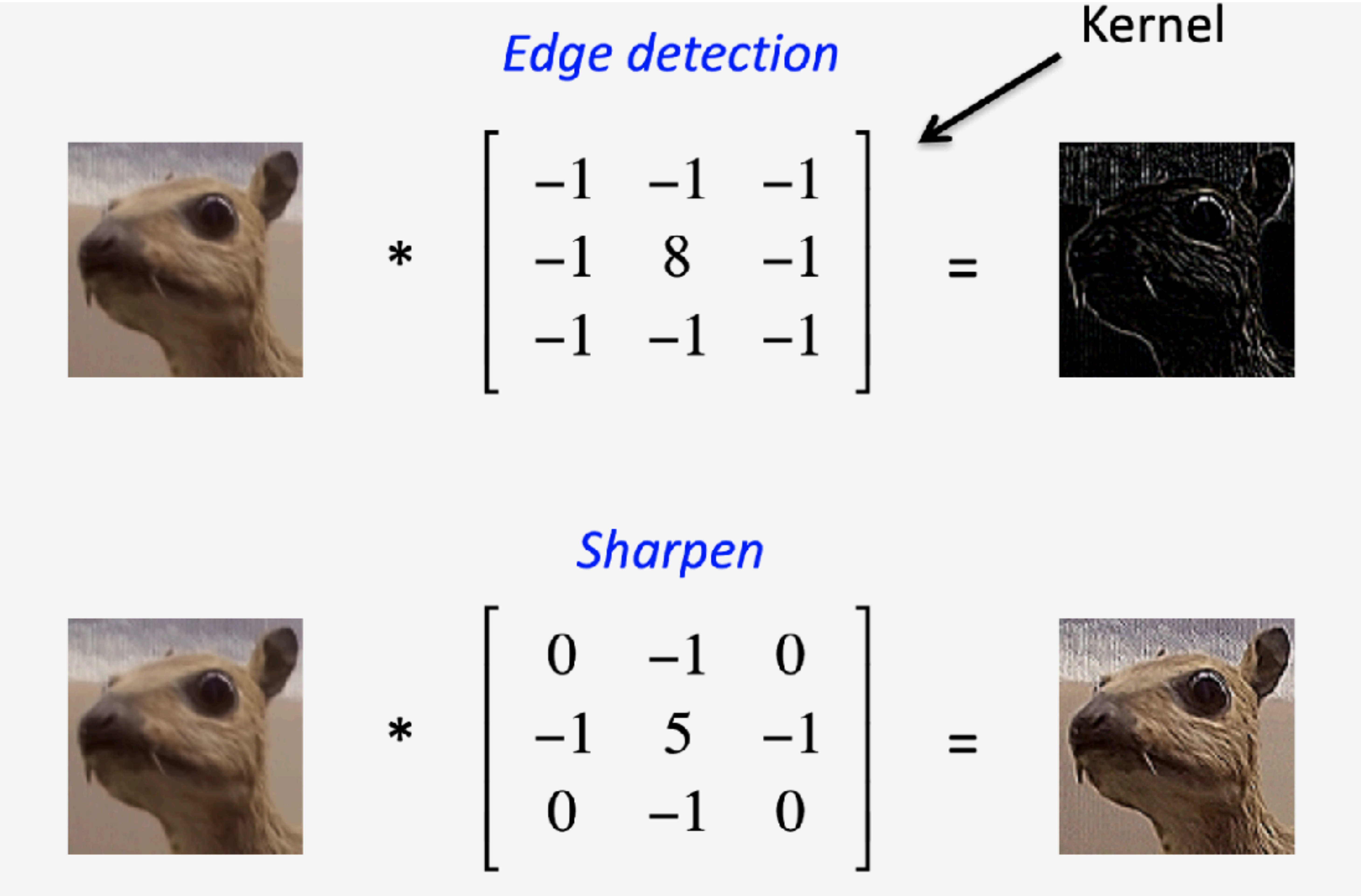
$$* \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix} =$$



mask

(vector, NOT a matrix!)

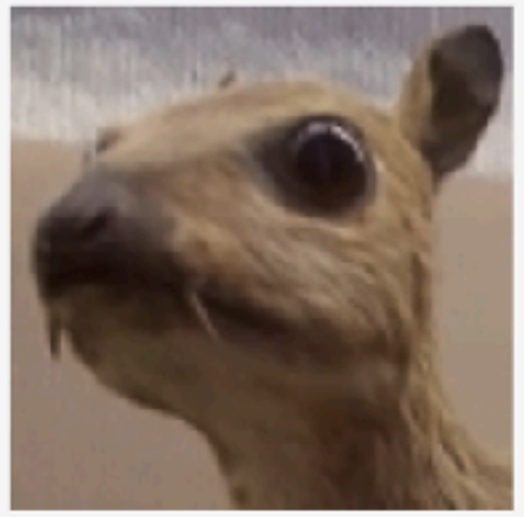
Examples of Convolutions



Question: How can we make convolutions more expressive?

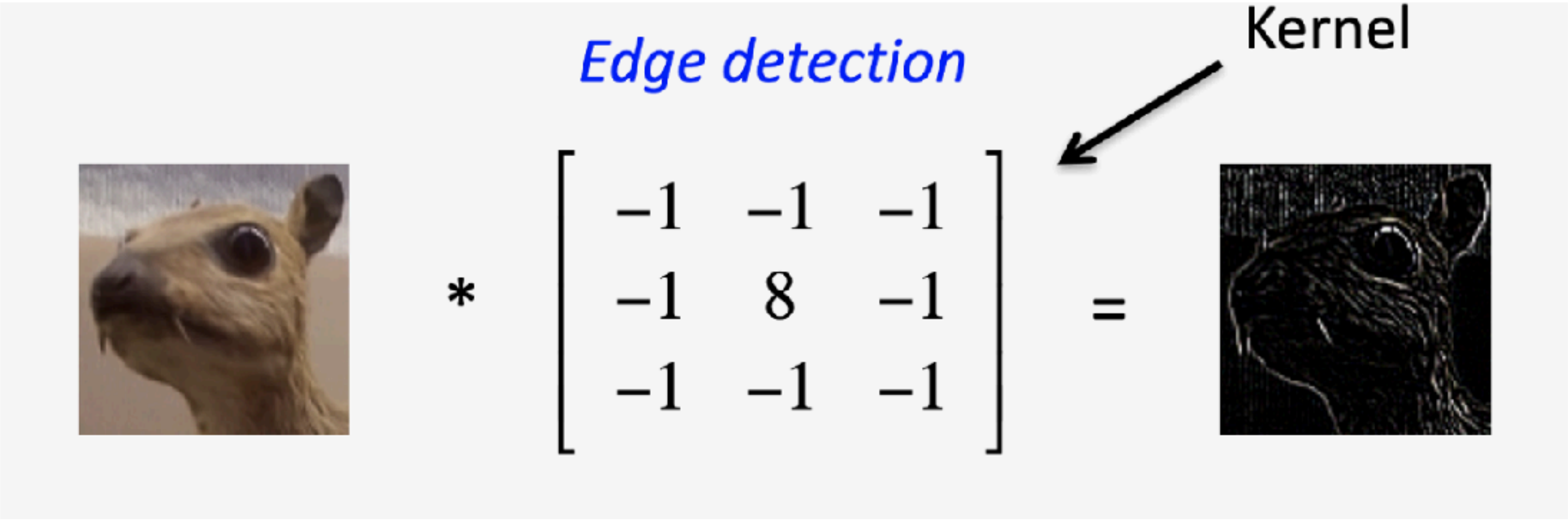
Edge detection

Kernel



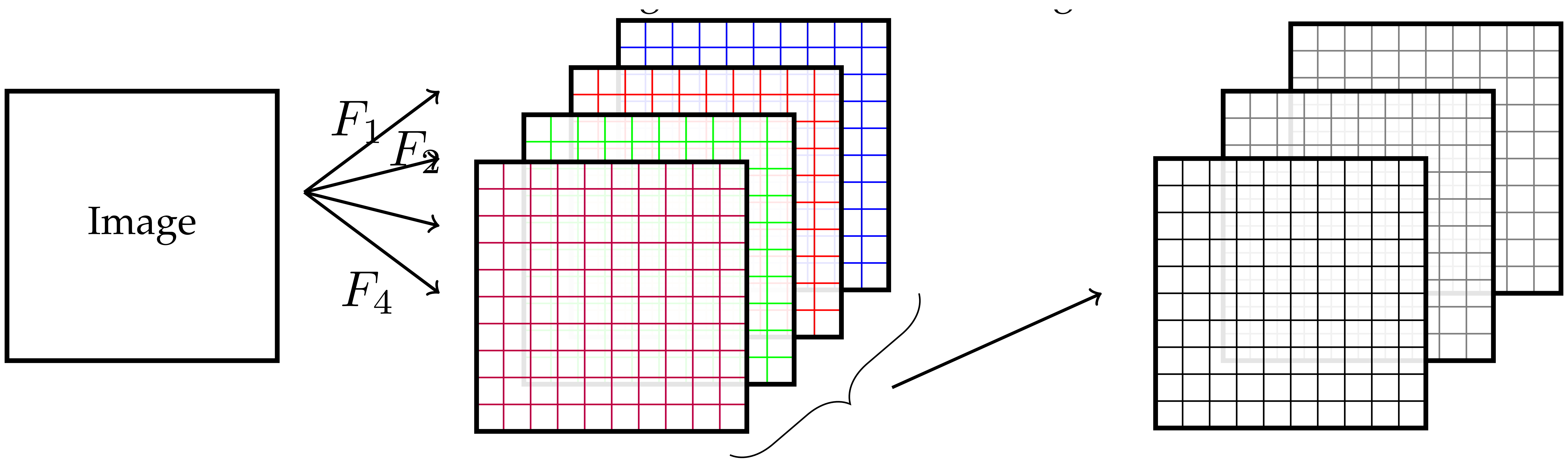
$$* \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix} =$$

Question: How can we make convolutions more expressive?



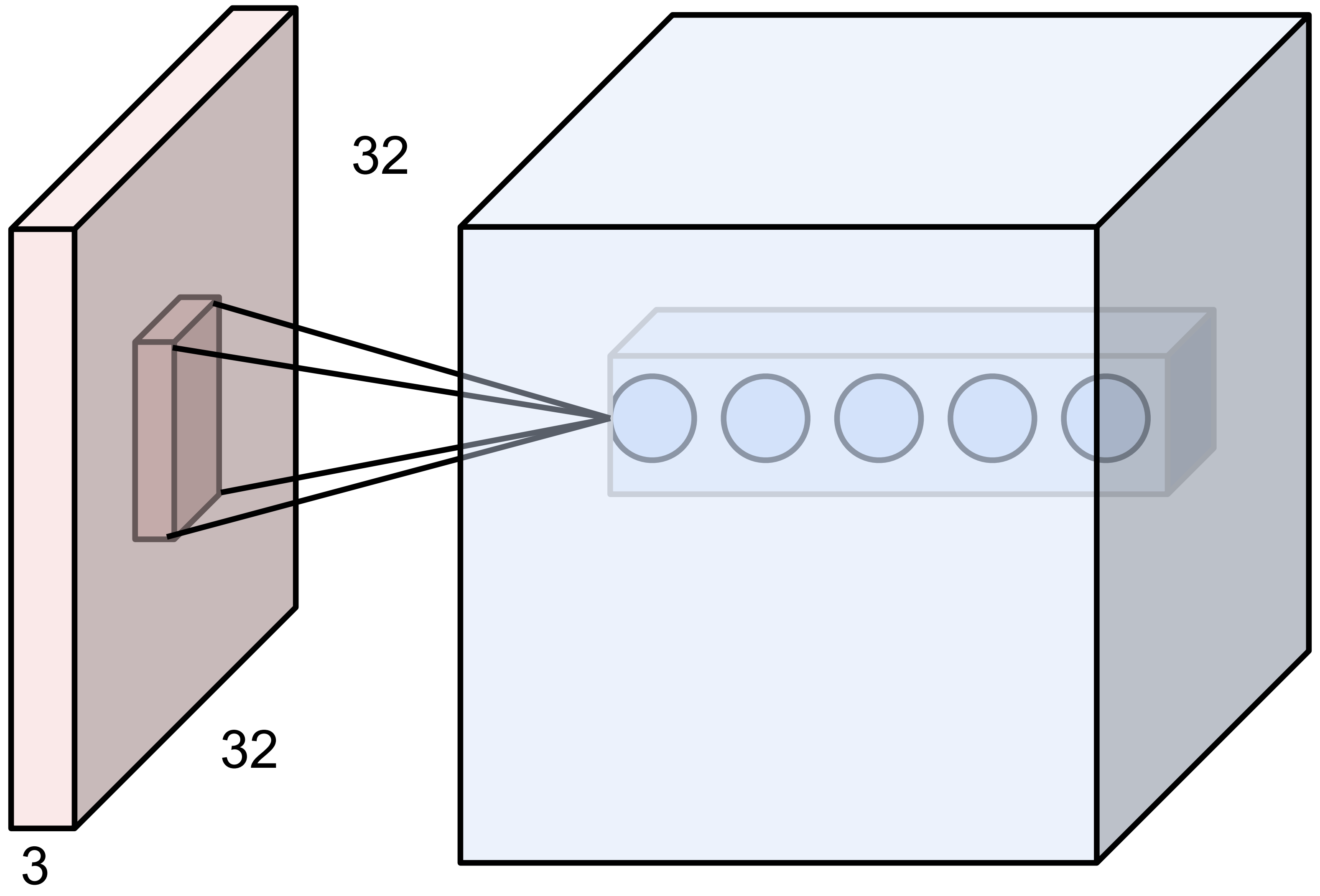
Width: Many kernels in parallel
Depth: Composing kernels

Multiple channels/filters



Channels: outputs of convolution

Filter bank: Collection of filters in a layer

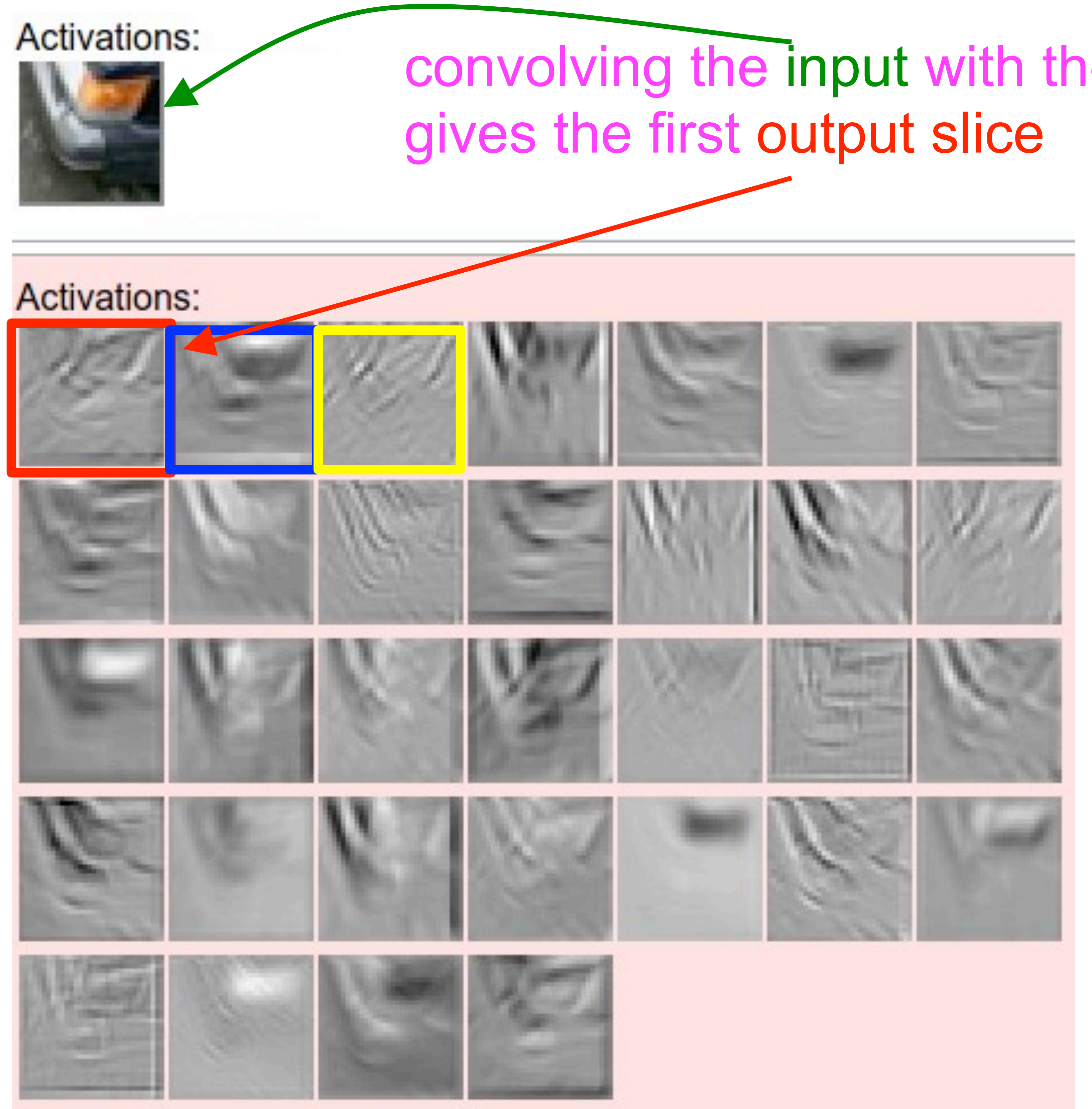


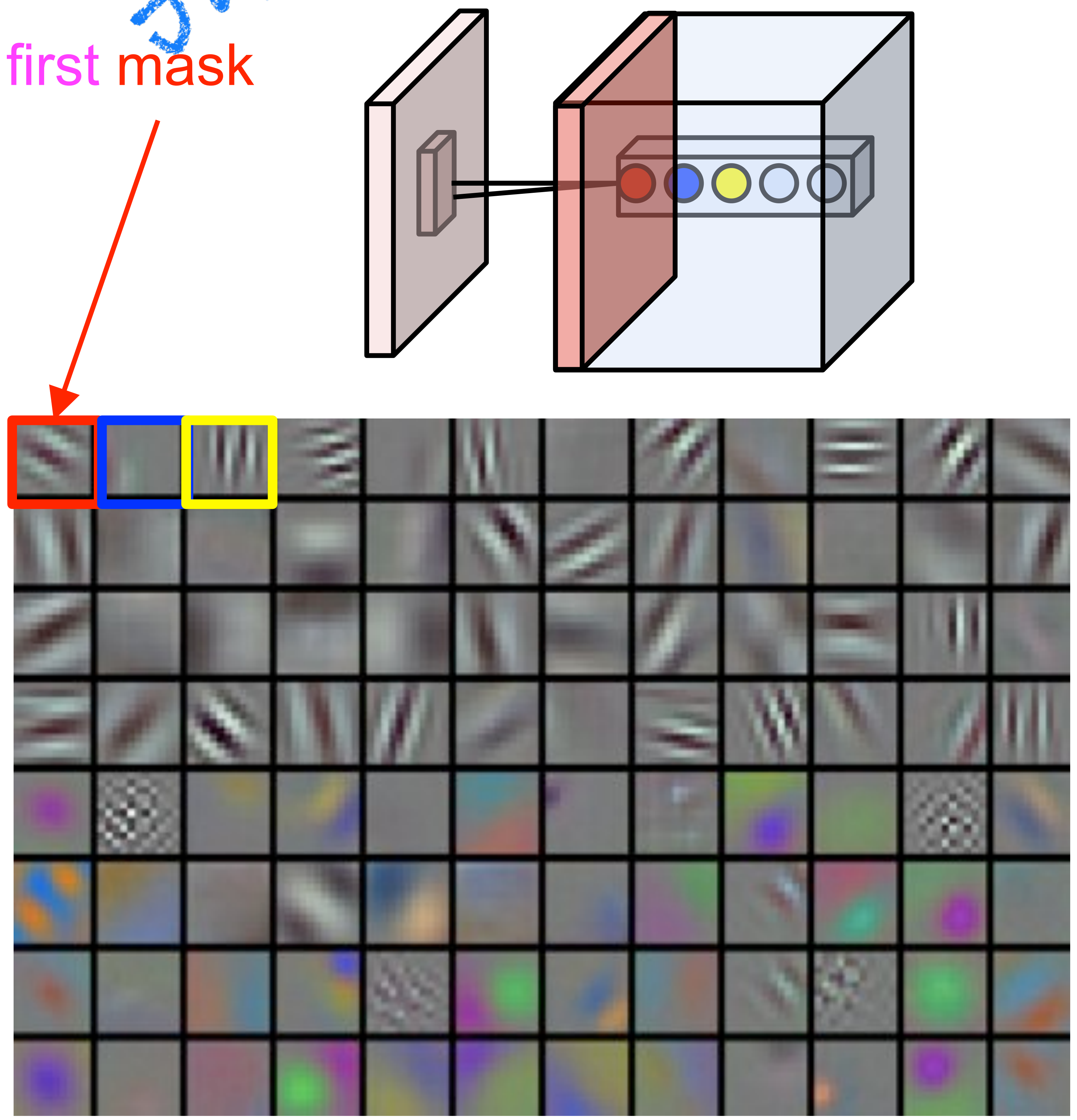
Hidden layer of “depth” 5: five neurons all looking at the same patch; five different masks.

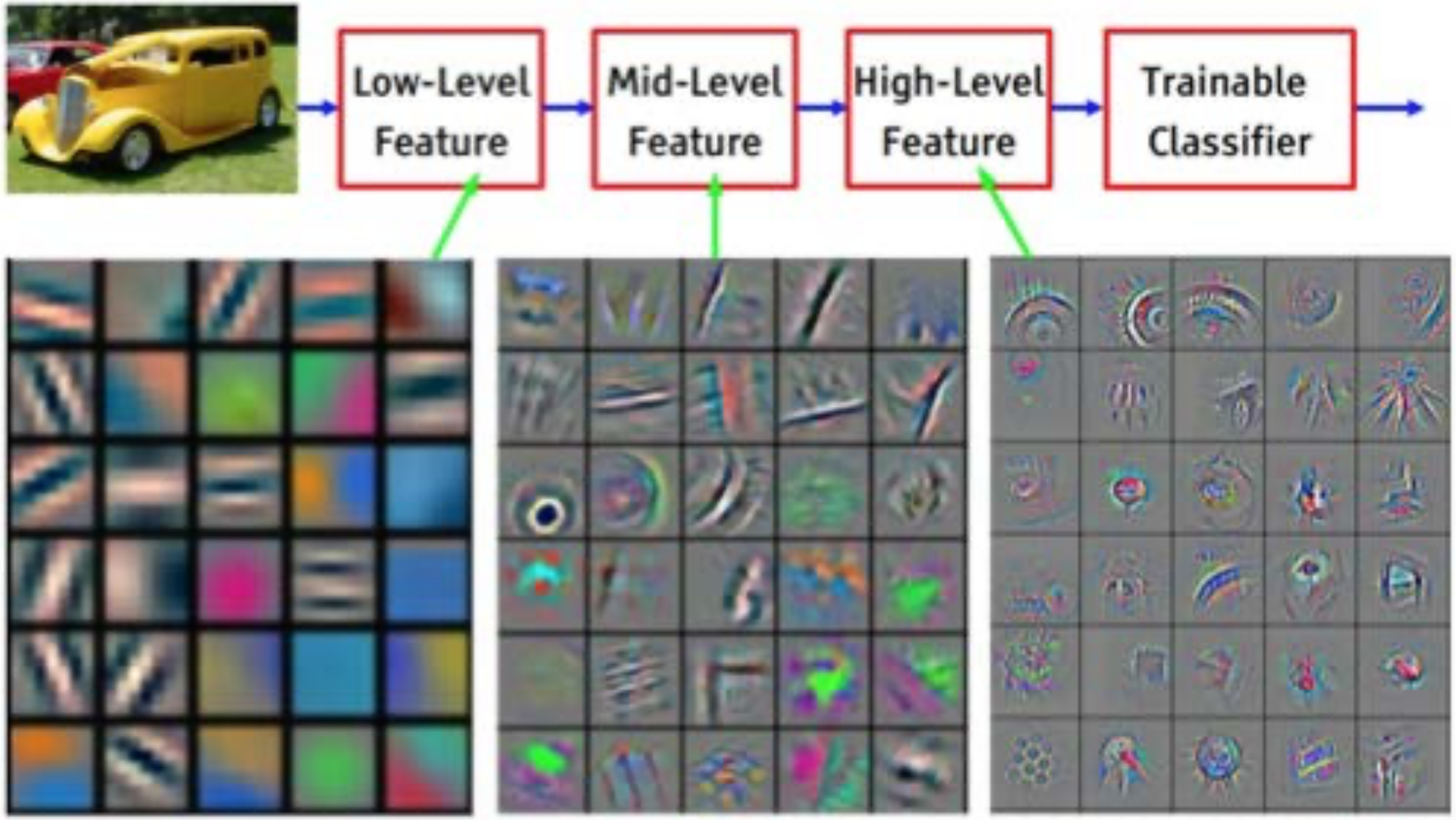
Apply the same 5 masks to each patch. Five neurons per patch.

conv. filter

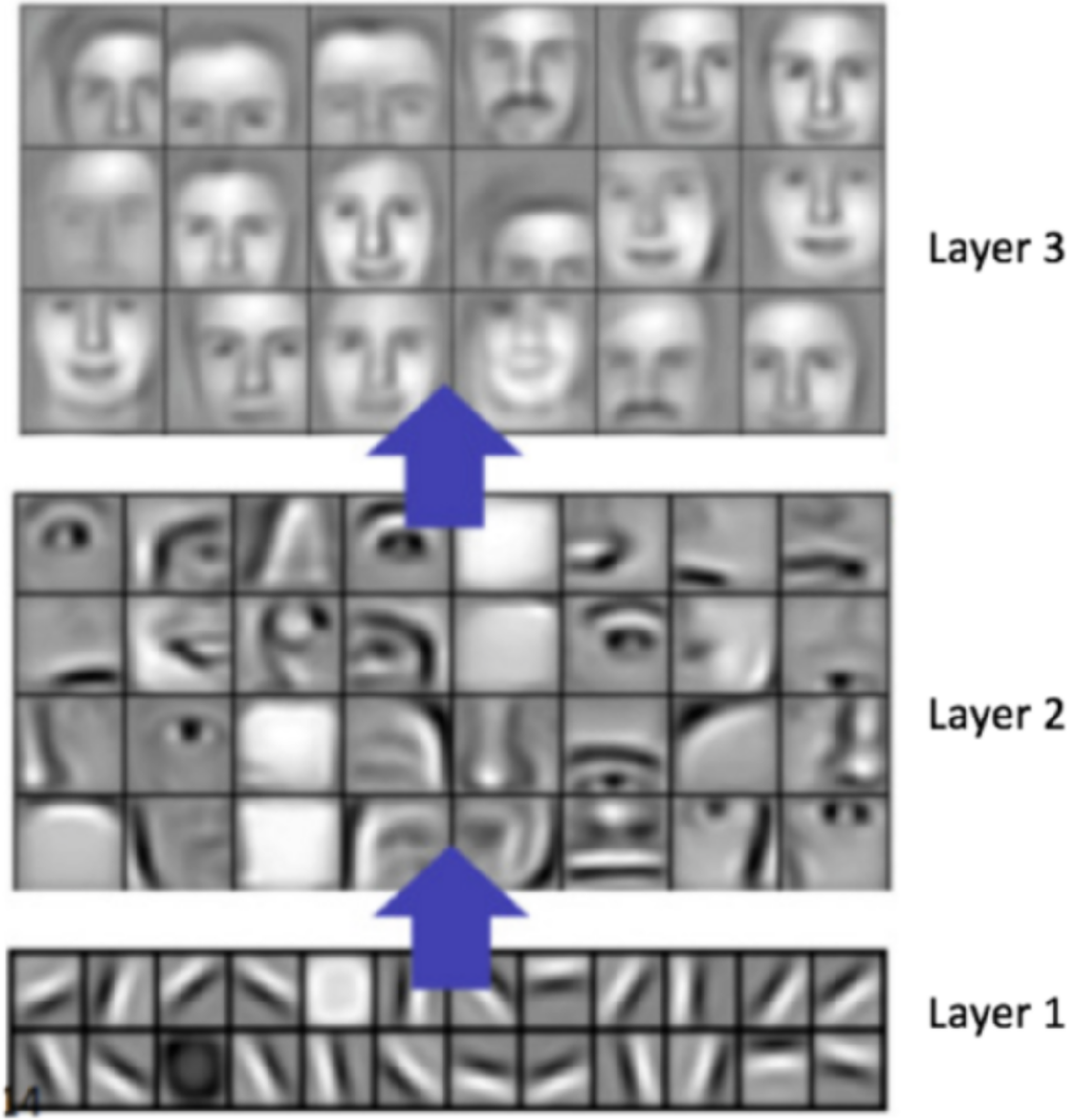
convolving the input with the first mask
gives the first output slice







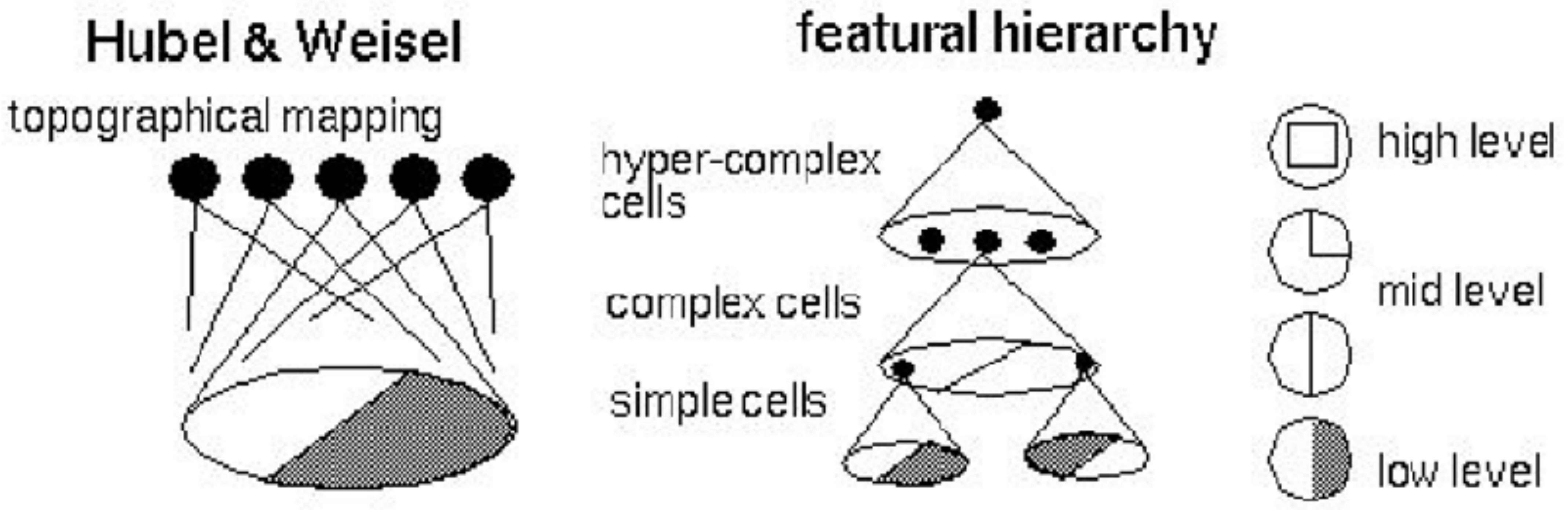
Feature visualization of convolutional net trained on ImageNet from [Zeiler & Fergus 2013]


Layer 3
Layer 2
Layer 1

What Humans Do?

- Hierarchical pattern recognition
  - edges -> simple parts-> parts -> objects -> scenes

- Hubel/Wiesel Architecture (1959, 1962, Nobel Prize,1981)



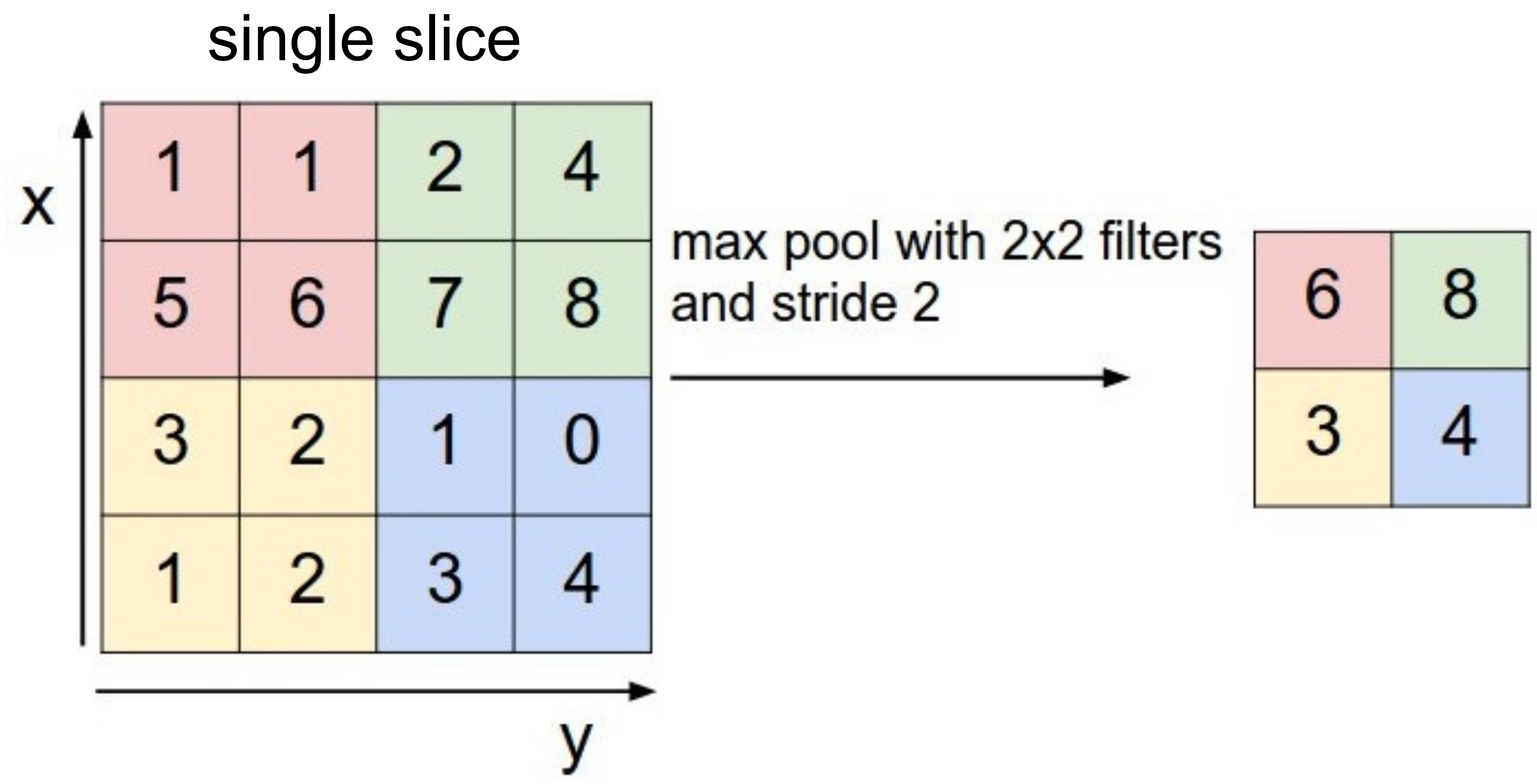
CNN key ideas

- Capture spatial dependencies: convolutions
  - spatial locality

- **Handle Translations: pooling**
  - abstract away locality

- Robustly scale for large images: weight sharing
  - apply the same detector to all the patches

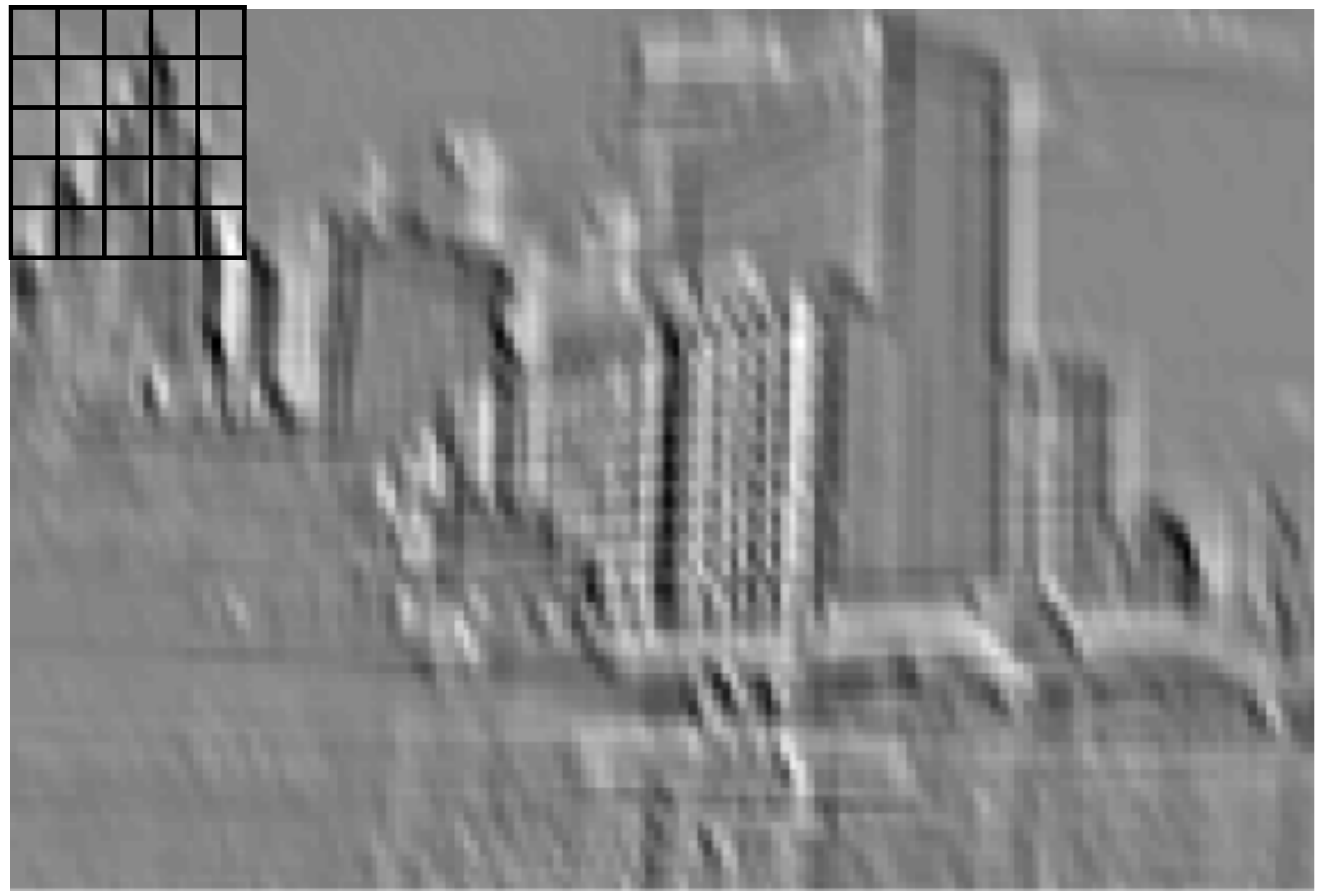
Max Pooling



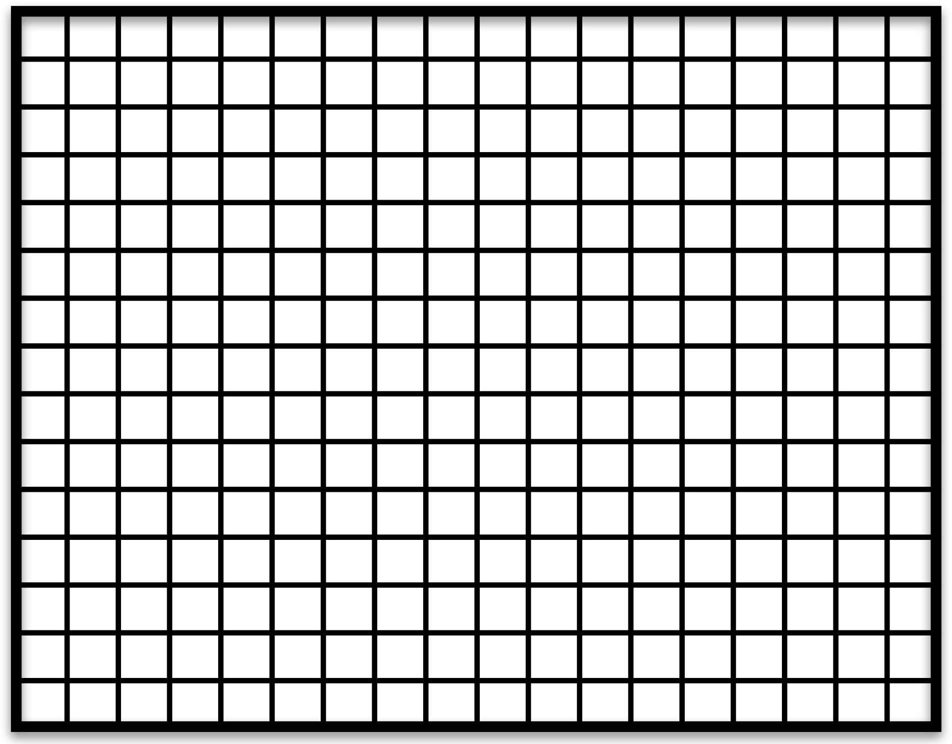
- Similar to filtering, but output the maximum entry instead of a weighted sum

Pooling

- We wish to know whether a feature was there but not exactly where it was



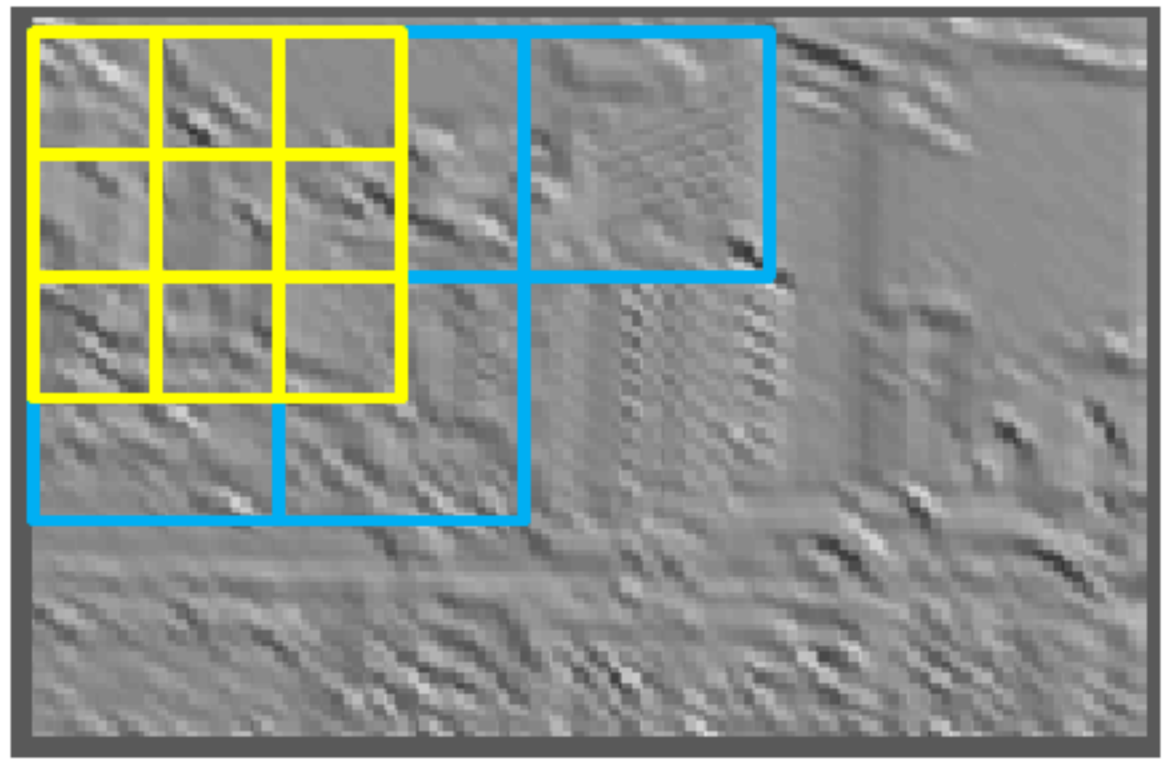
feature map



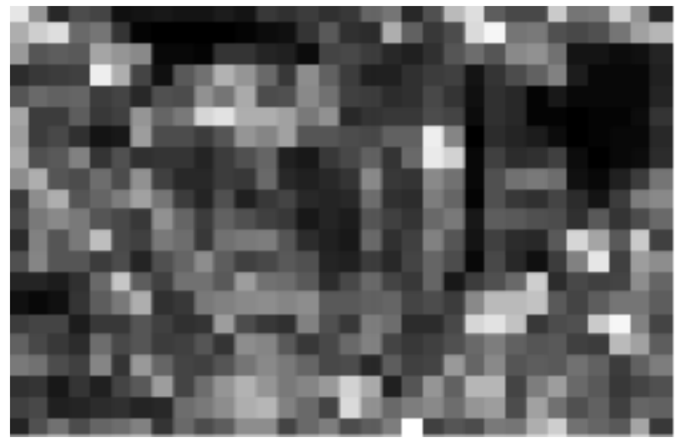
pooled map

Pooling (max)

- Pooling region and "stride" may vary
  - pooling induces translation invariance at the cost of spatial resolution
  - stride reduces the size of the resulting feature map

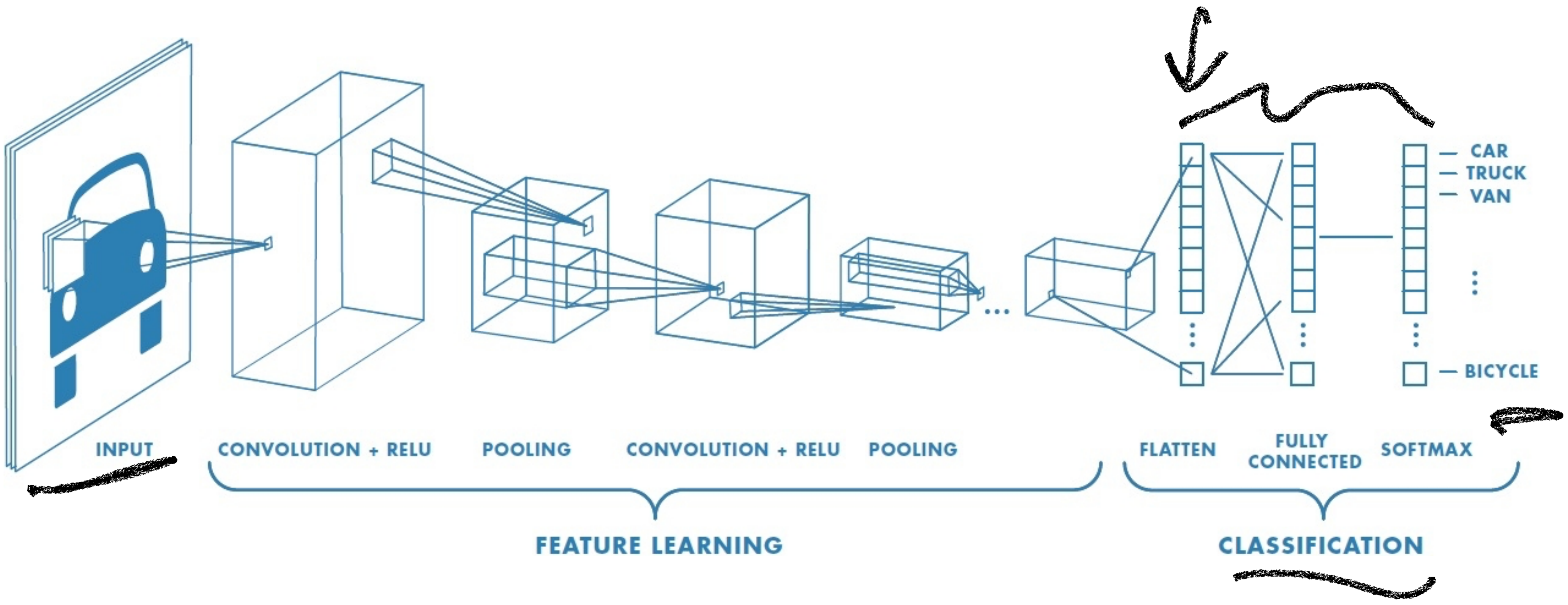


feature map



feature map
after max pooling

Example Architecture

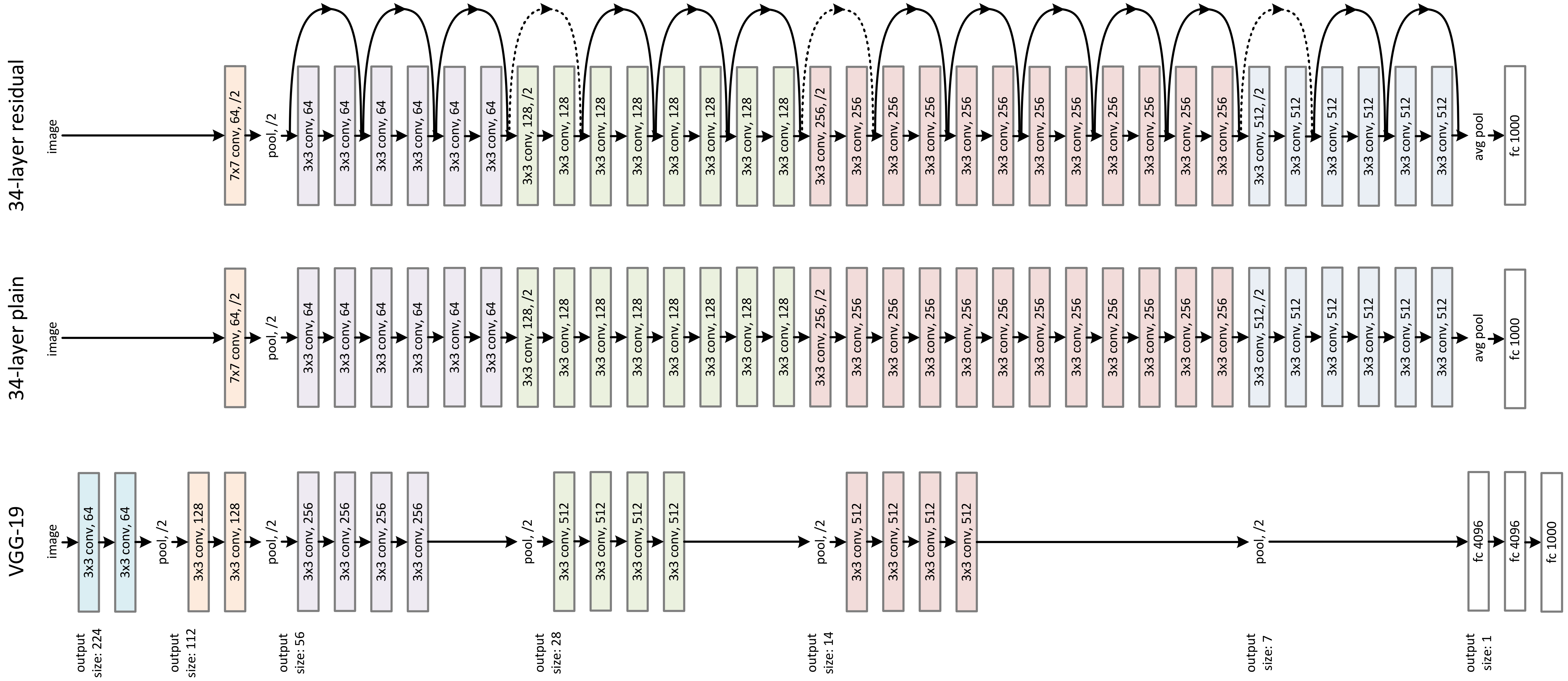


- Trainable via SGD and back-propagation

https://www.mathworks.com/solutions/deep-learning/convolutional-neural-network.html

Resnet (2015)

- A bit more modern…



[He et al. 2015]

IMAGENET **Large Scale Visual Recognition Challenge**

The Image Classification Challenge:
1,000 object classes
1,431,167 images

Output:		Output:	
Scale	✔	Scale	✗
T-shirt		T-shirt	
<u>Steel drum</u>		Giant panda	
Drumstick		Drumstick	
Mud turtle		Mud turtle	

Russakovsky et al. arXiv, 2014

Example: computer vision ’12-

- ImageNet classification (what’s it about)
  - over a million images, 1K image categories

2012 Teams	%error
Supervision (Toronto)	15.3
ISI (Tokyo)	26.1
VGG (Oxford)	26.9
XRCE/INRIA	27.0
UvA (Amsterdam)	29.6
INRIA/LEAR	33.4

2013 Teams	%error
Clarifai (NYU spinoff)	11.7
NUS (singapore)	12.9
Zeiler-Fergus (NYU)	13.5
A. Howard	13.5
OverFeat (NYU)	14.1
UvA (Amsterdam)	14.2
Adobe	15.2
VGG (Oxford)	15.2
VGG (Oxford)	23.0

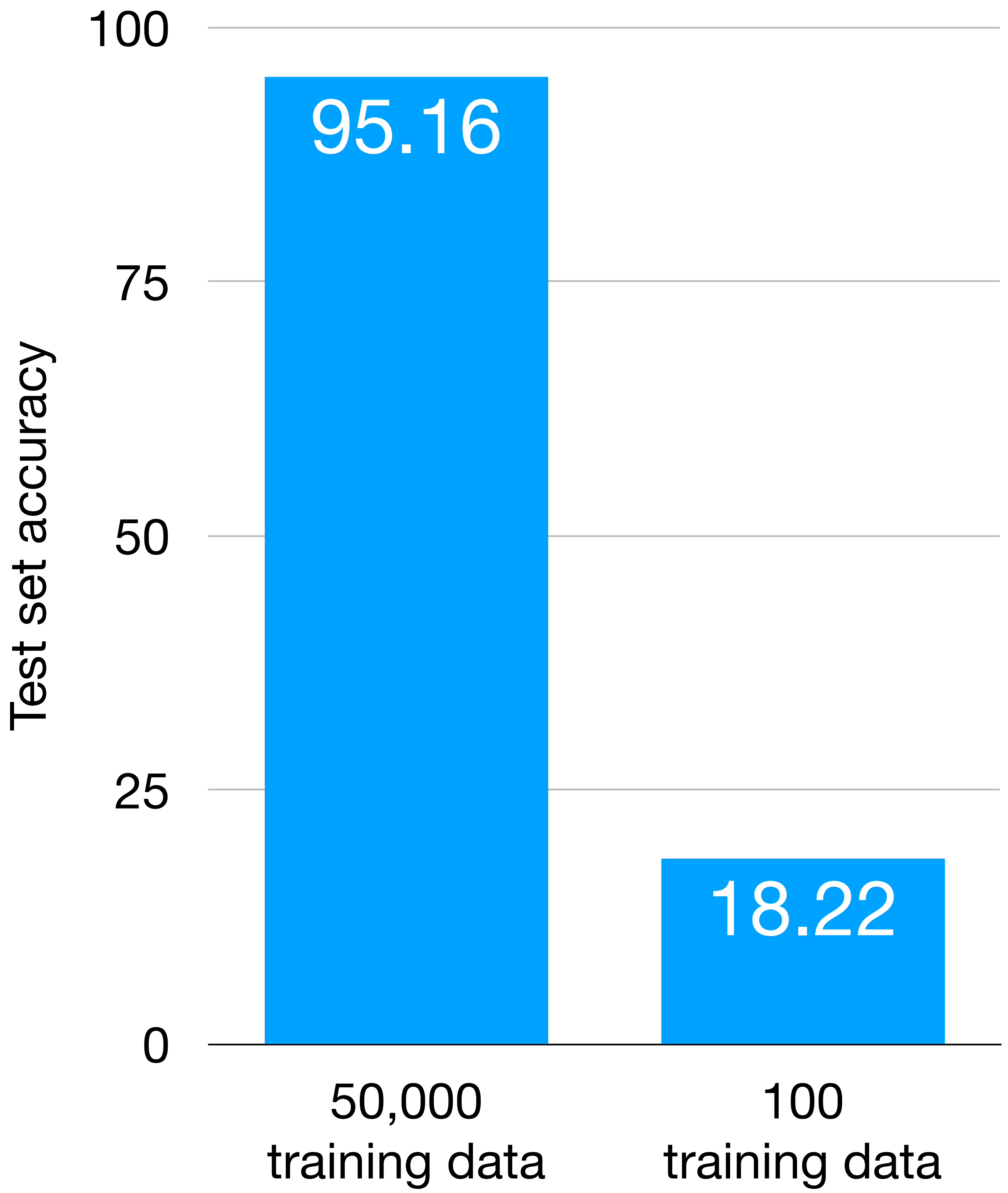
2014 Teams	%error
GoogLeNet	6.6
VGG (Oxford)	7.3
MSRA	8.0
A. Howard	8.1
DeeperVision	9.5
NUS-BST	9.7
TTIC-ECP	10.2
XYZ	11.2
UvA	12.1

[Figure by Yann LeCun]

Models are Data Hungry

CIFAR-10 dataset

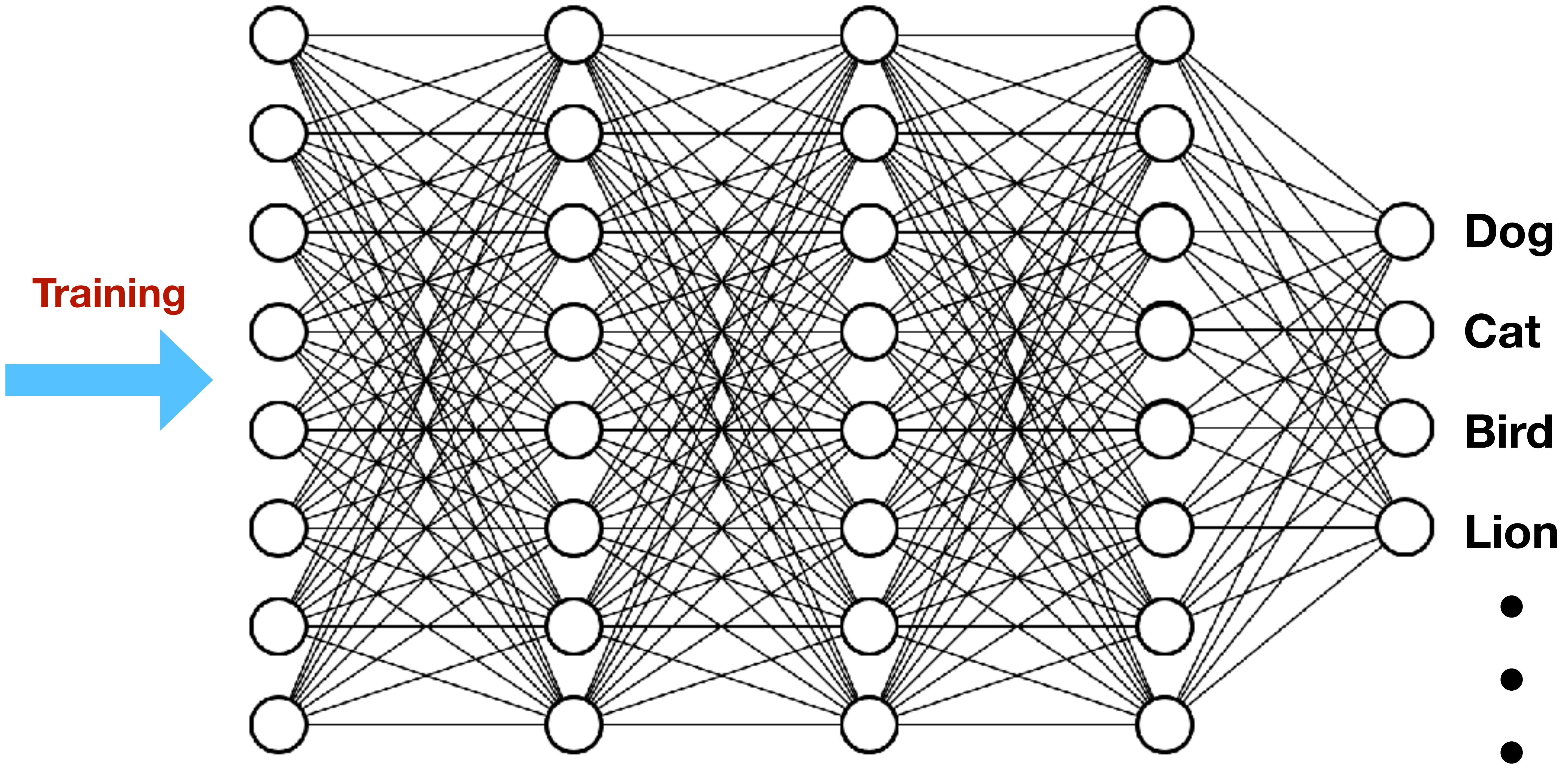
classes: 10
img per class: 5000



Initialization via Pre-Training

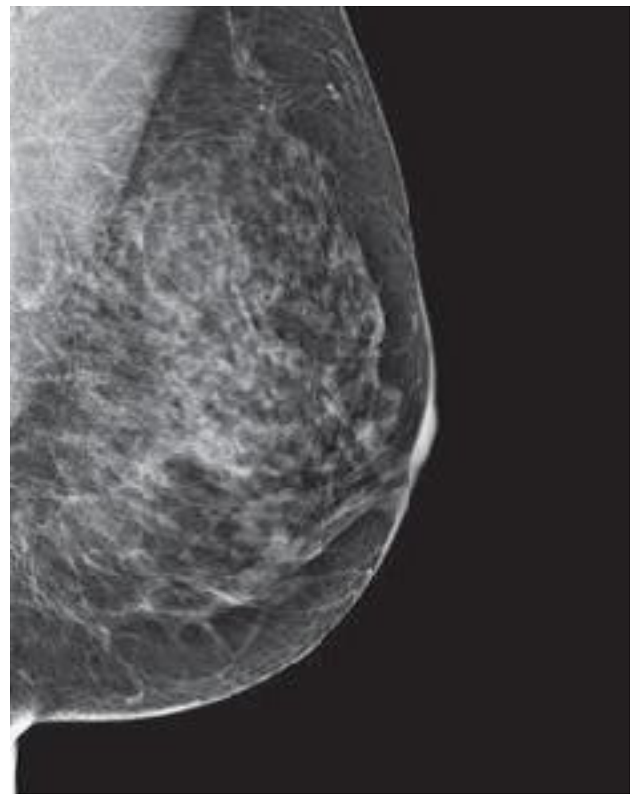
Transfer knowledge acquired from other dataset

ImageNet dataset:
1.2M labeled images

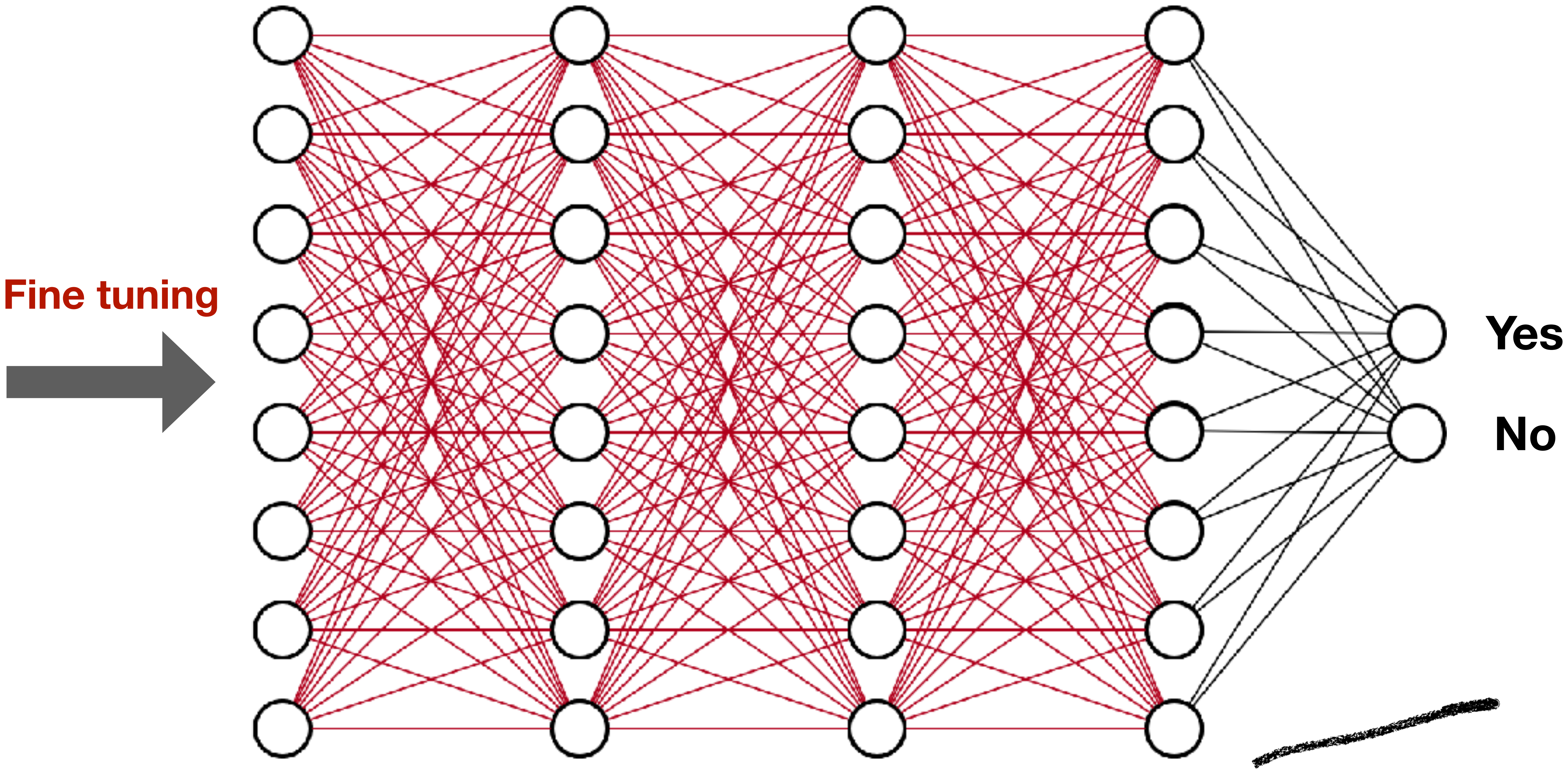


Initialization via Pre-Training

Weights in red are initialized from the pre-trained model.



Fine tuning



Weights in black are initialized from scratch and are updated by SGD during fine tuning stage.

Summary: CNNs

- **Convolution: “local detectors”**
  spatial locality

- **Weight sharing:**
  apply same detector to all image patches
  efficiency: much fewer parameters!, translation invariance

- **Pooling**
  abstract away locality

Agenda

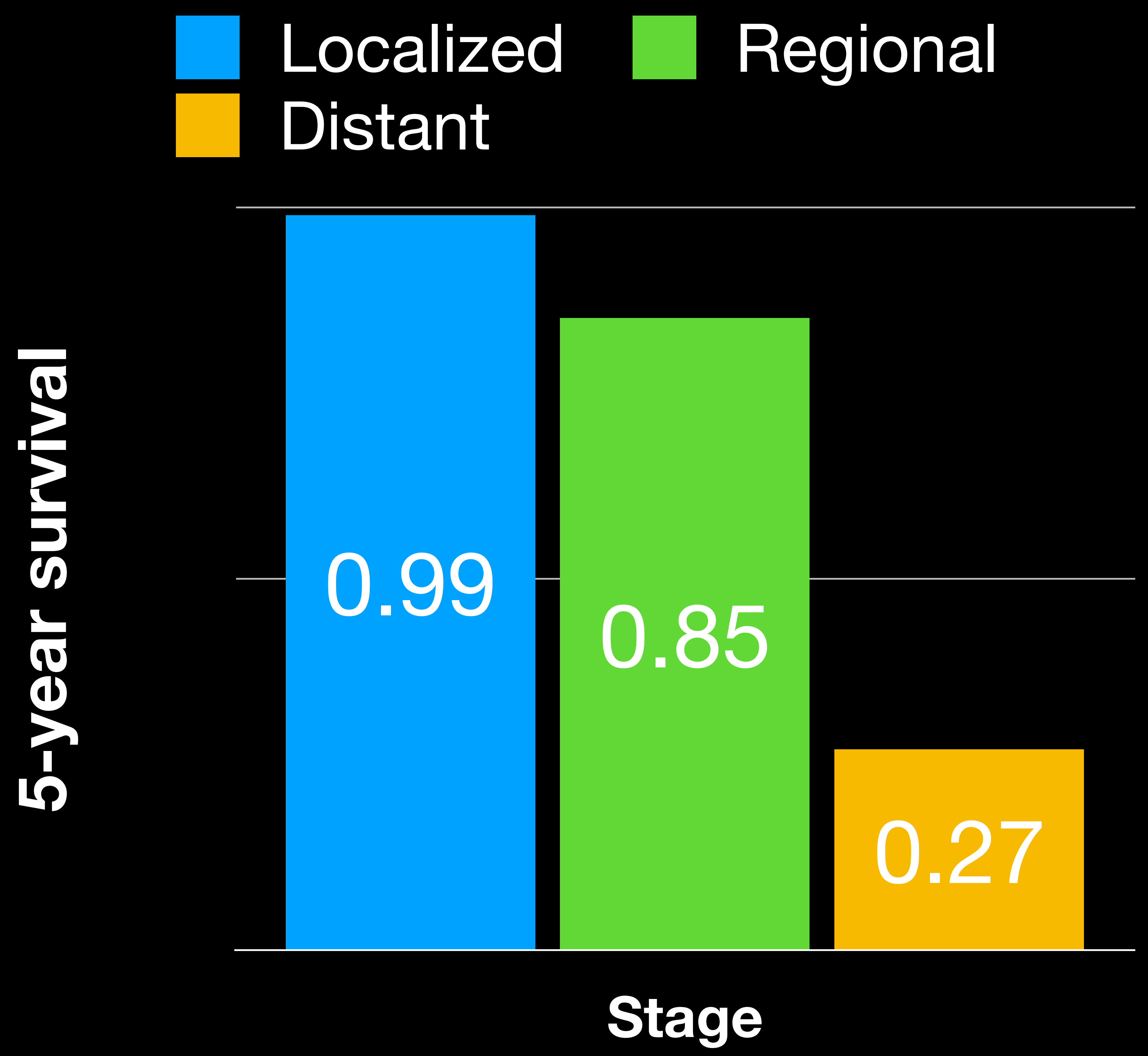
Data: What is medical imaging?

Method Foundations: How do we build models on imaging data?

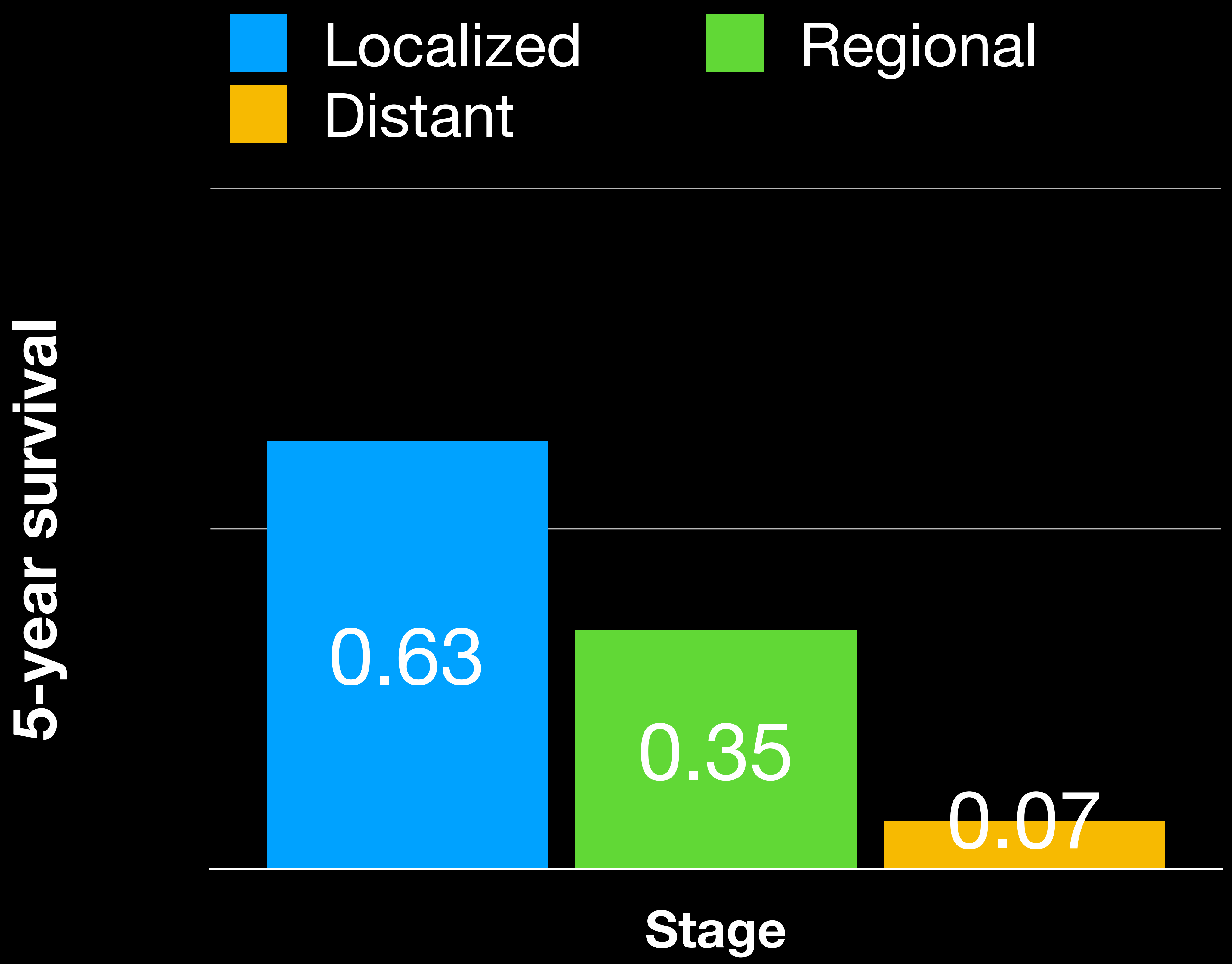
Applications: How can we catch cancer earlier?

Interpretation: How can we audit our models?

Early Detection is critical



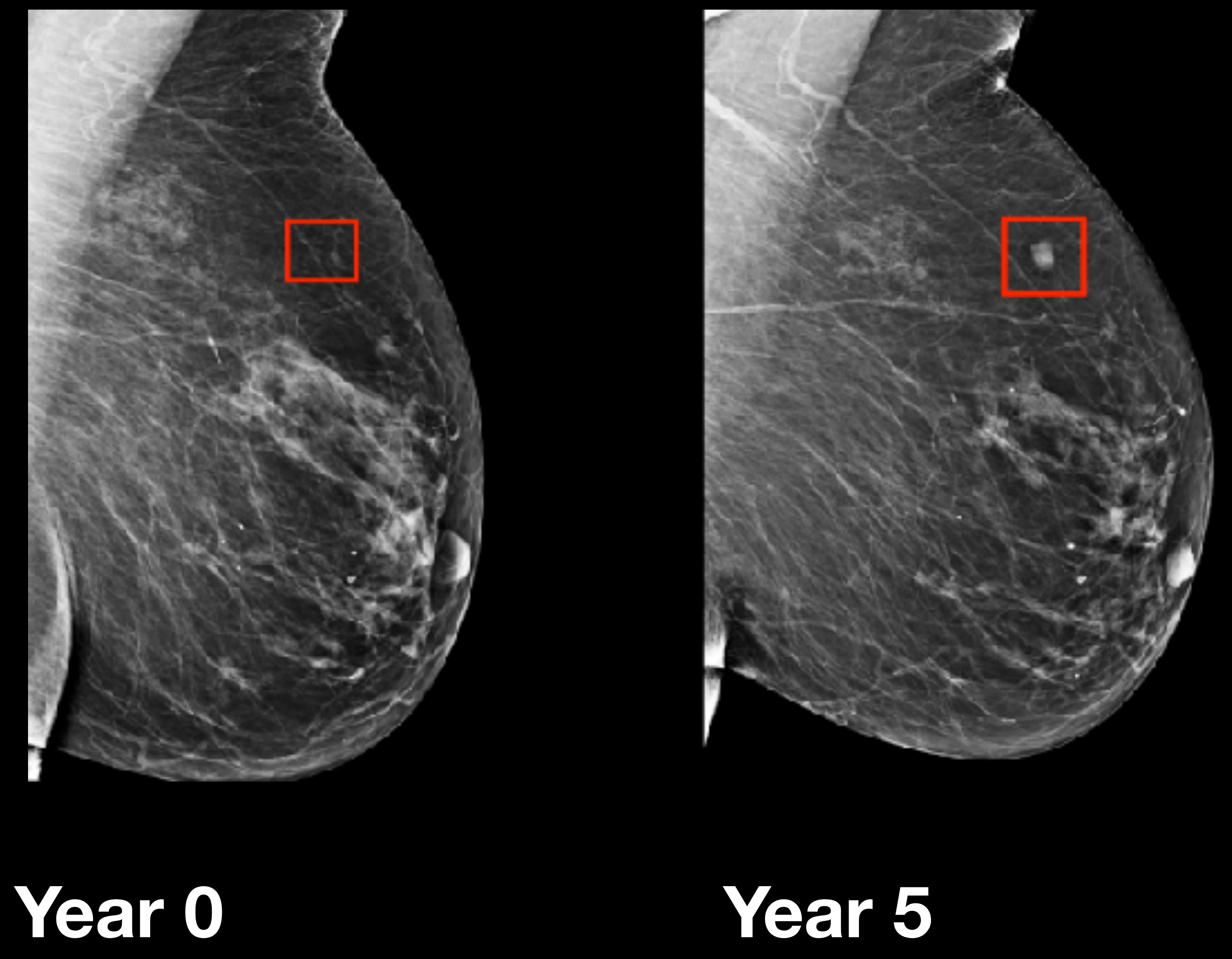
Breast Cancer



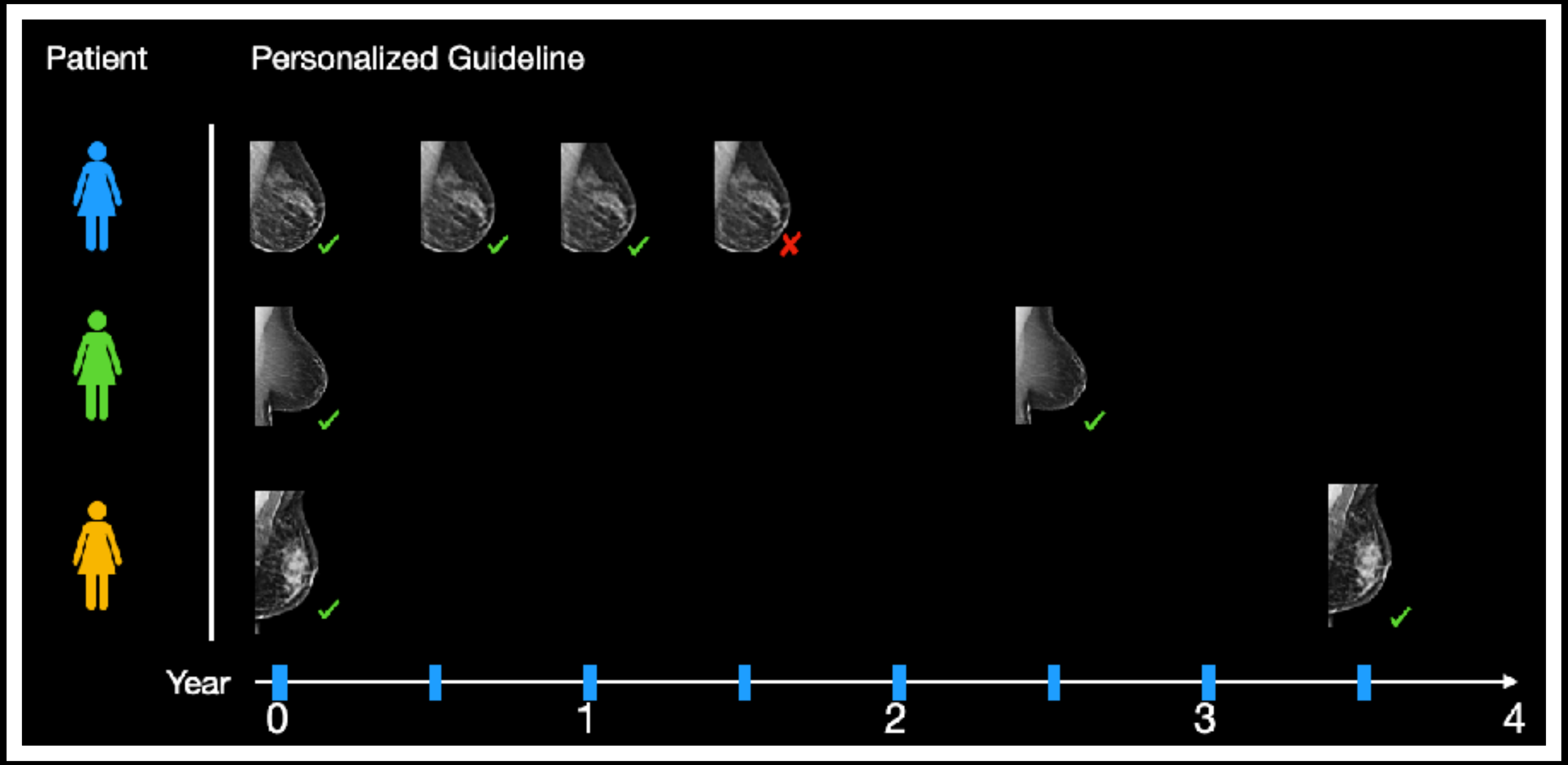
Lung Cancer

How to catch cancer earlier

Predict Cancer Risk



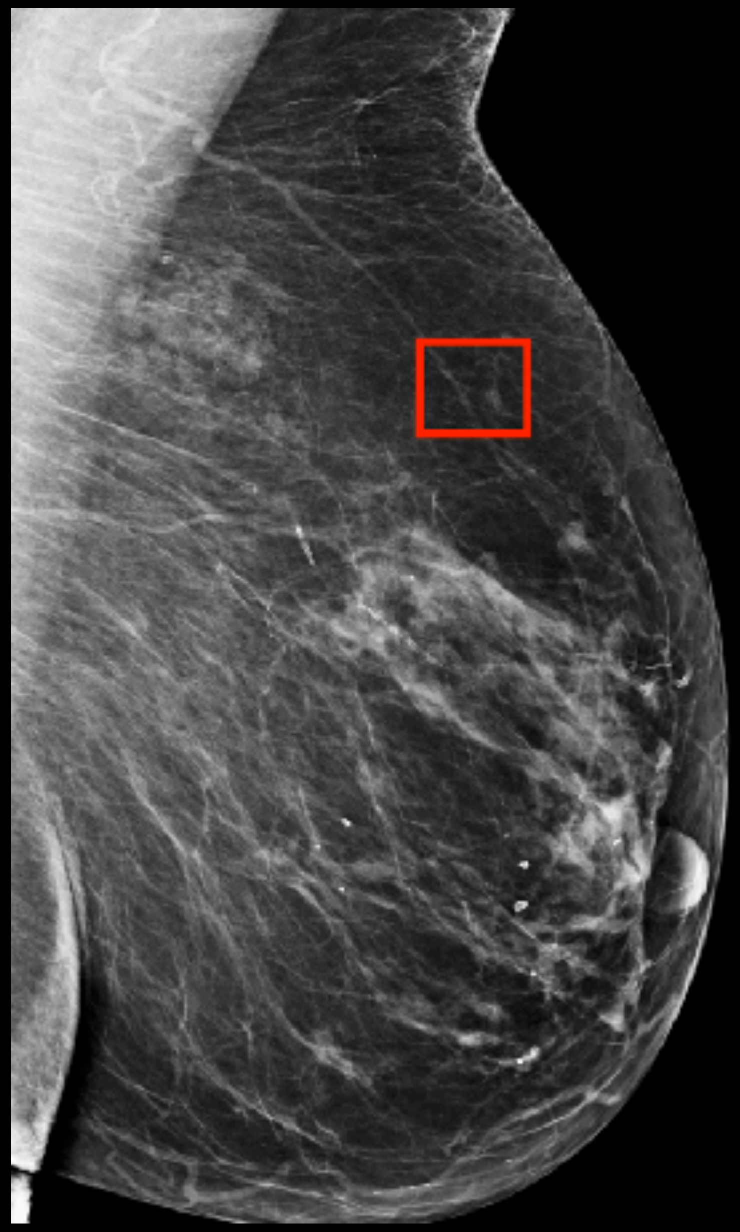
Create personalized screening policy



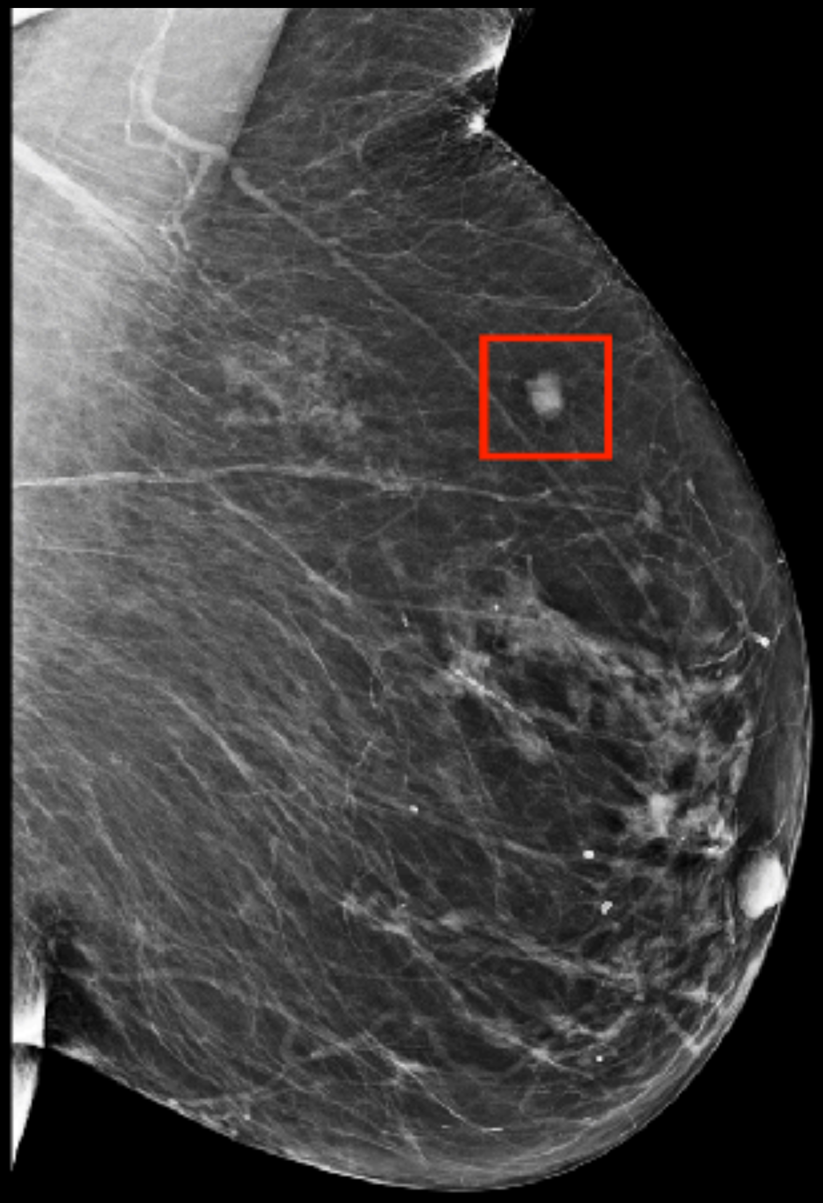
How to catch cancer earlier

Predict Cancer Risk

- Identify which population is at risk of developing cancer



Year 0



Year 5

Journal of Clinical Oncology®
An American Society of Clinical Oncology Journal

Multi-Institutional Validation of a Mammography-Based Breast Cancer Risk Model

Adam Yala, MEng[1,2]; Peter G. Mikhael, BS[1,2]; Fredrik Strand, MD, PhD[3,4]; Gigin Lin, MD, PhD[5]; Siddharth Satuluru, BS[6];

SCIENCE TRANSLATIONAL MEDICINE

Toward robust mammography-based models for breast cancer risk

Adam Yala[1,2*], Peter G. Mikhael[1,2], Fredrik Strand[3,4], Gigin Lin[5], Kevin Smith[6,7], Yung-Liang Leslie Lamb[8], Kevin Hughes[9], Constance Lehman[8†], Regina Barzilay[1,2†]

Radiology

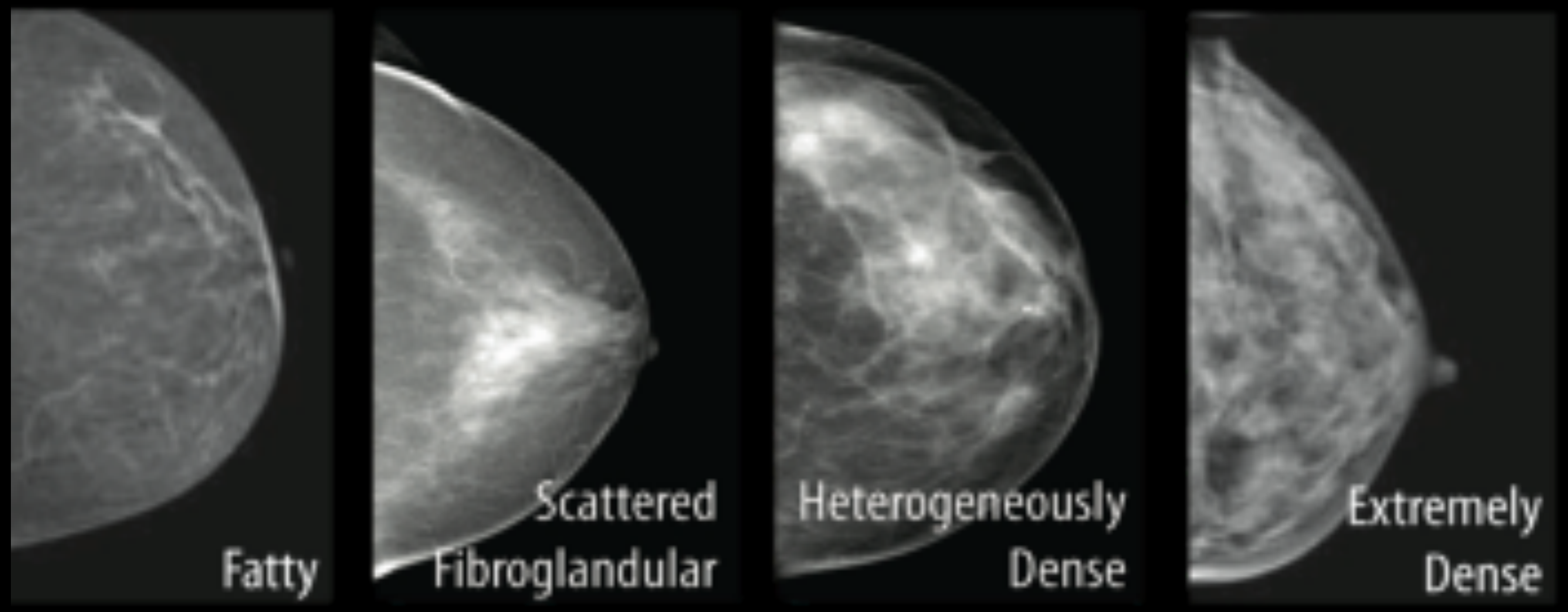
A Deep Learning Mammography-based Model for Improved Breast Cancer Risk Prediction

Adam Yala, MEng • Constance Lehman, MD, PhD • Tal Schuster, MS • Tally Portnoi, BS • Regina Barzilay, PhD

Traditional approach: use expert knowledge

Age
Family History
Prior Breast Procedure
Breast Density

→ Model → **Risk**



AUC: 0.63

AUC: 0.61 without Density

J Natl Cancer Inst. 2006 Sep 6;98(17):1204-14.

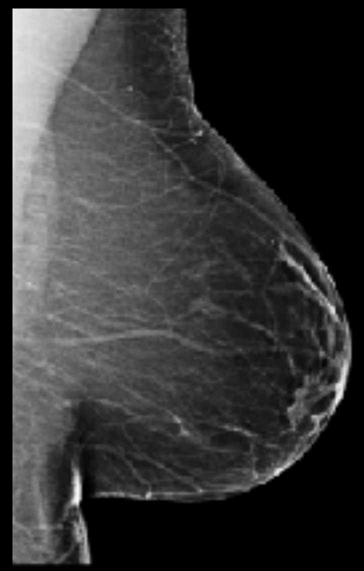
Prospective breast cancer risk prediction model for women undergoing screening mammography.

Barlow WE[1], White E, Ballard-Barbash R, Vacek PM, Titus-Ernstoff L, Carney PA, Tice JA, Buist DS, Geller BM, Rosenberg R, Yankaskas BC, Kerlikowske K.

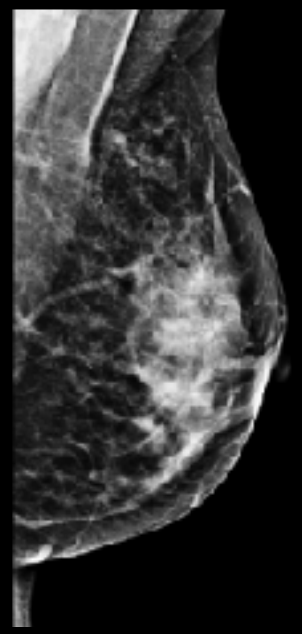
Learning to predict future cancer from imaging

Patient | **Future Outcome**

→ 3 year cancer



→ No cancer



→ 5 year cancer

2. ResNet

Standard Architecture
Resnet-18

2. ResNet

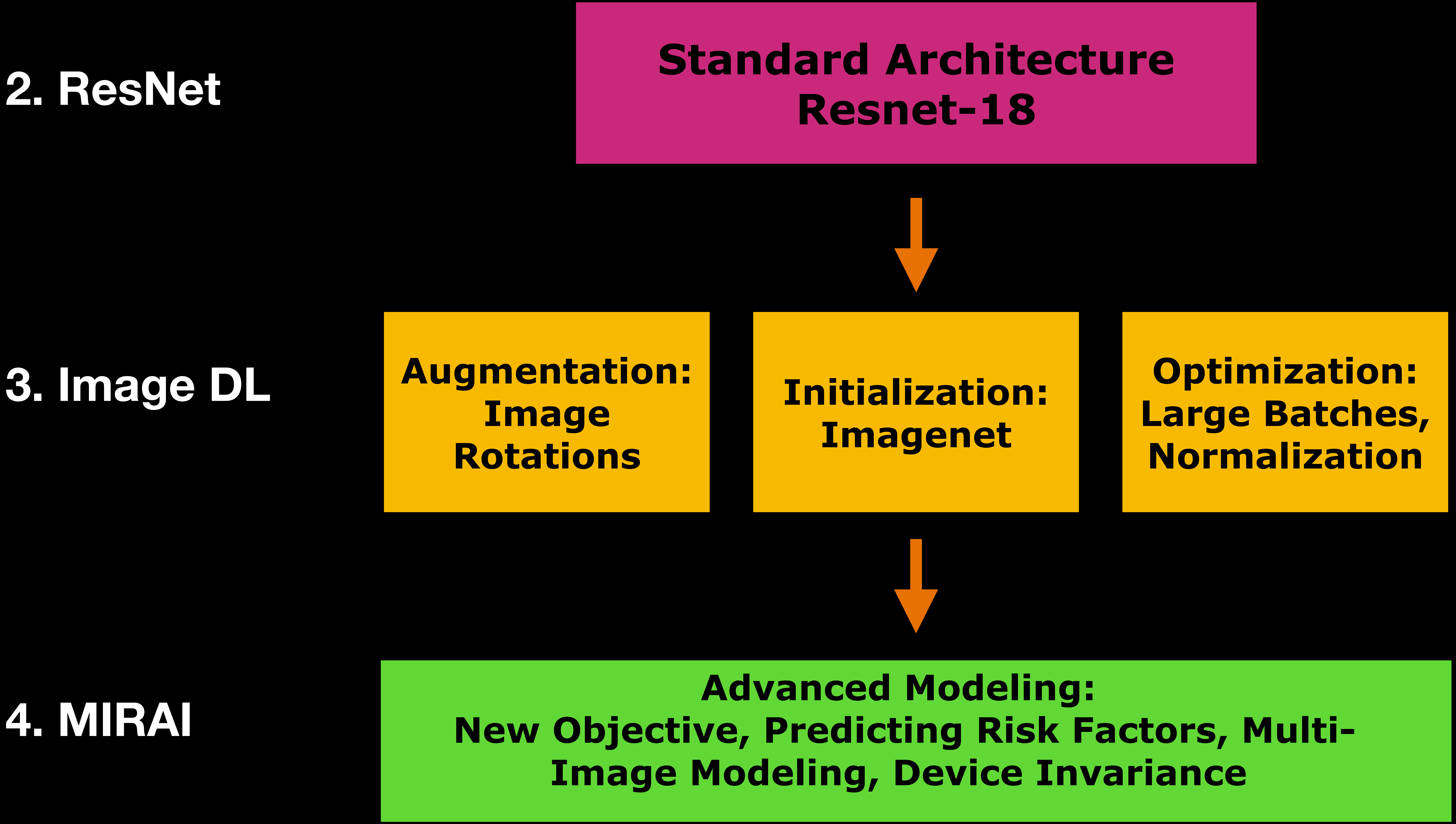
Standard Architecture
Resnet-18

3. Image DL

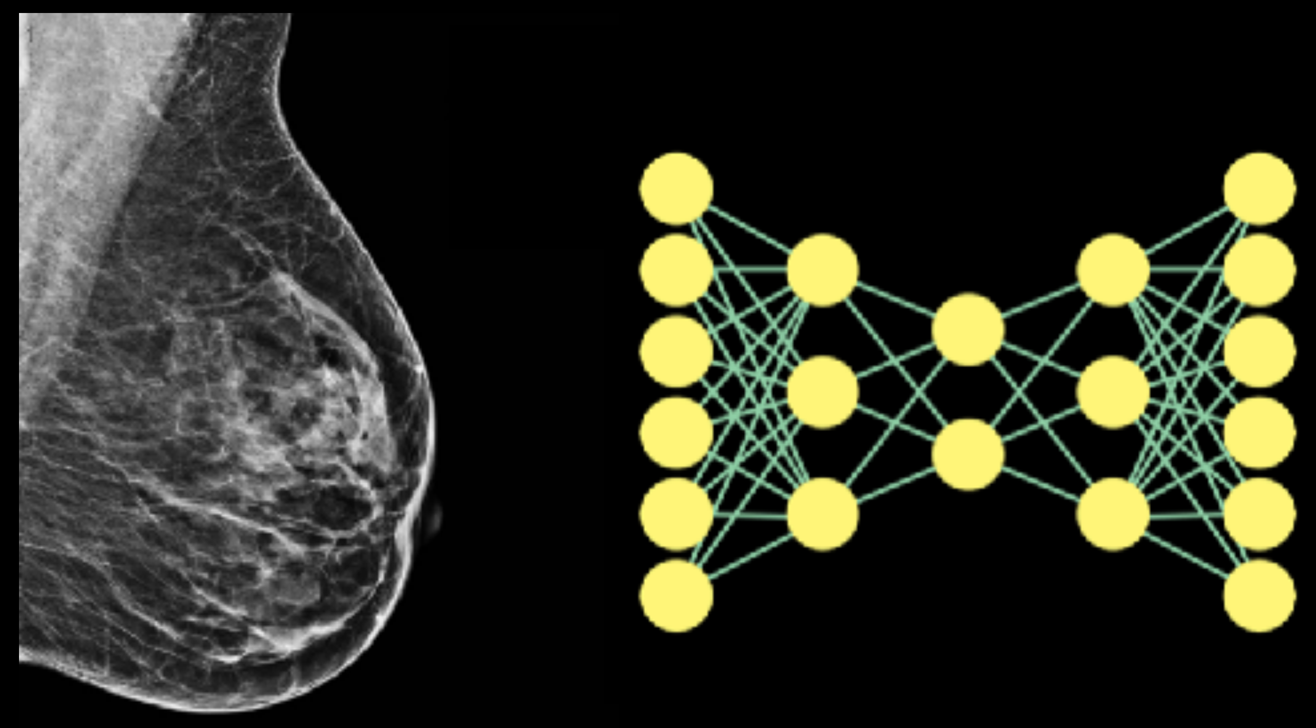
Augmentation:
Image
Rotations

Initialization:
Imagenet

Optimization:
Large Batches,
Normalization

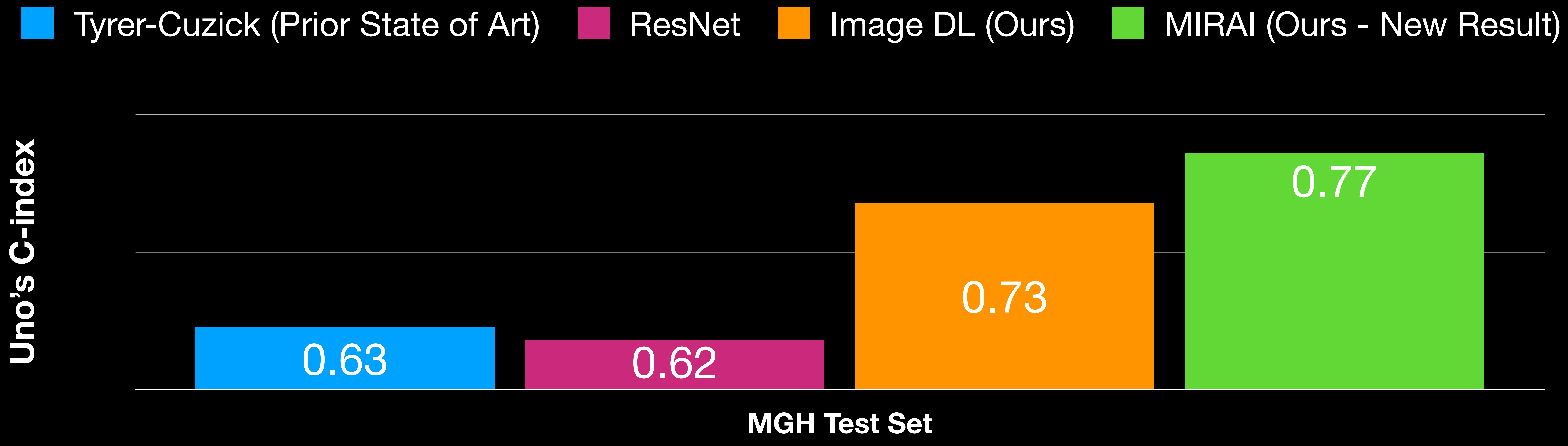

2. ResNet
Standard Architecture
Resnet-18
3. Image DL
Augmentation:
Image
Rotations
Initialization:
Imagenet
Optimization:
Large Batches,
Normalization
4. MIRAI
Advanced Modeling:
New Objective, Predicting Risk Factors, Multi-
Image Modeling, Device Invariance

MIRAI: Assessing Breast Cancer Risk

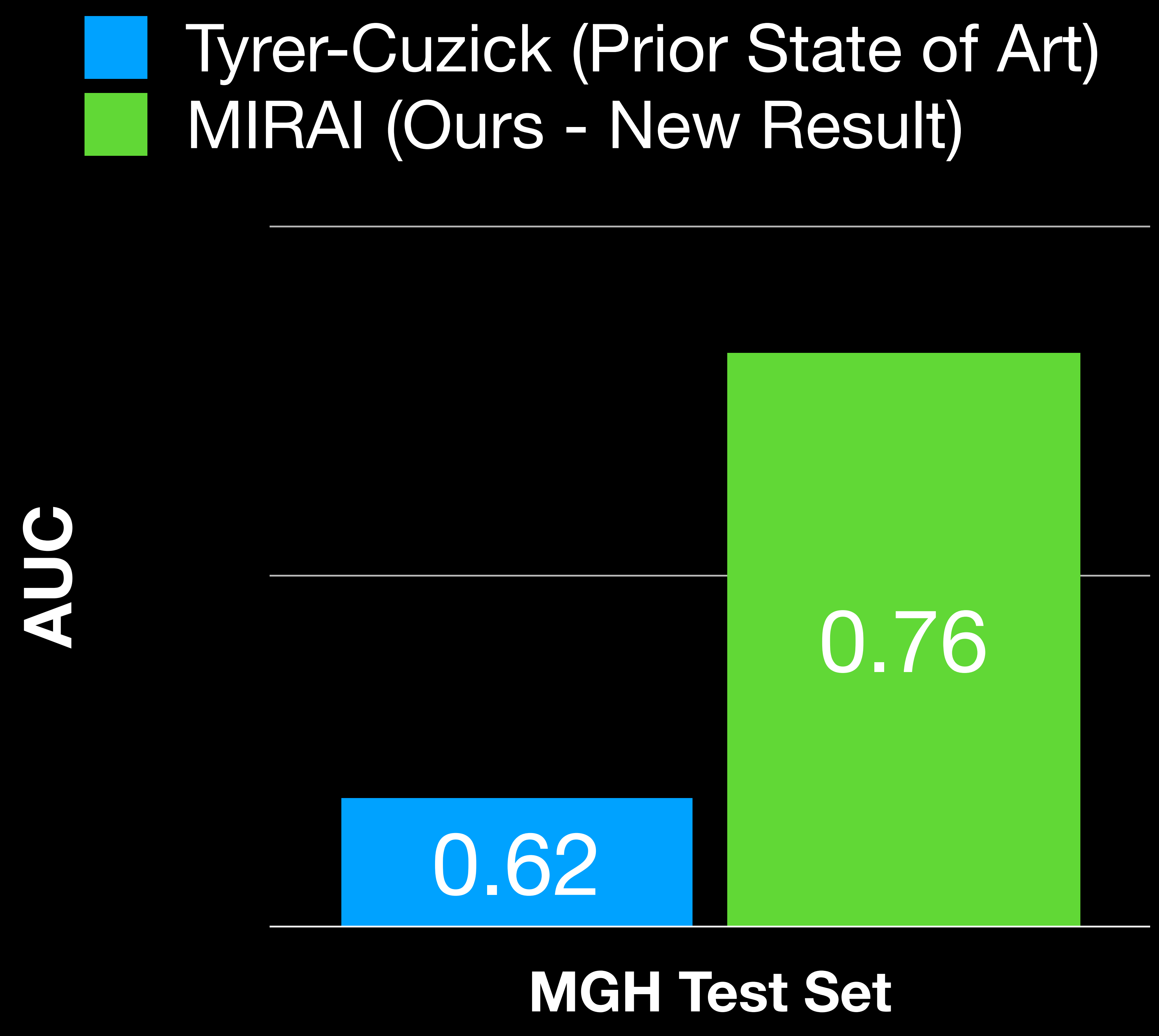


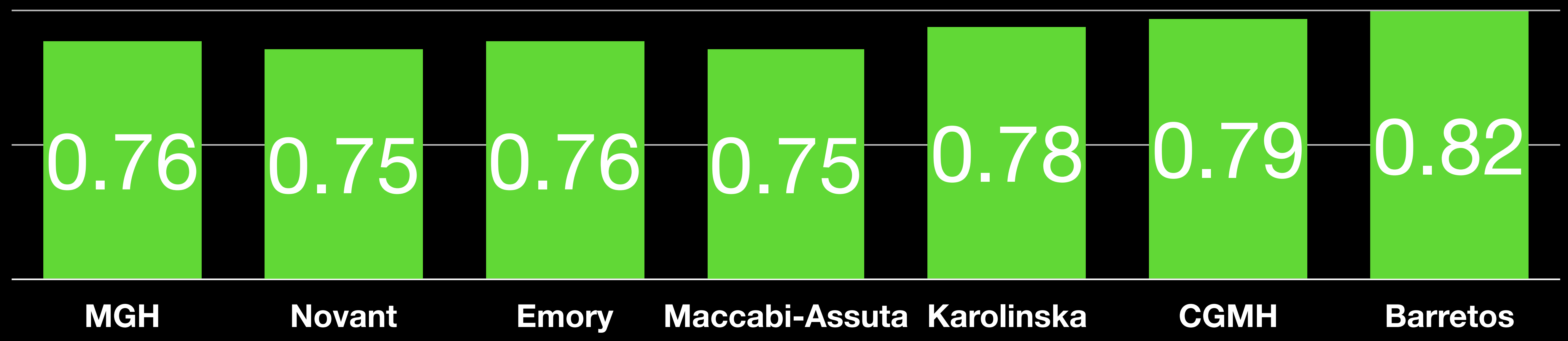
5 Year Breast Cancer Risk

Trained at MGH
Tested on 26,000 holdout exams

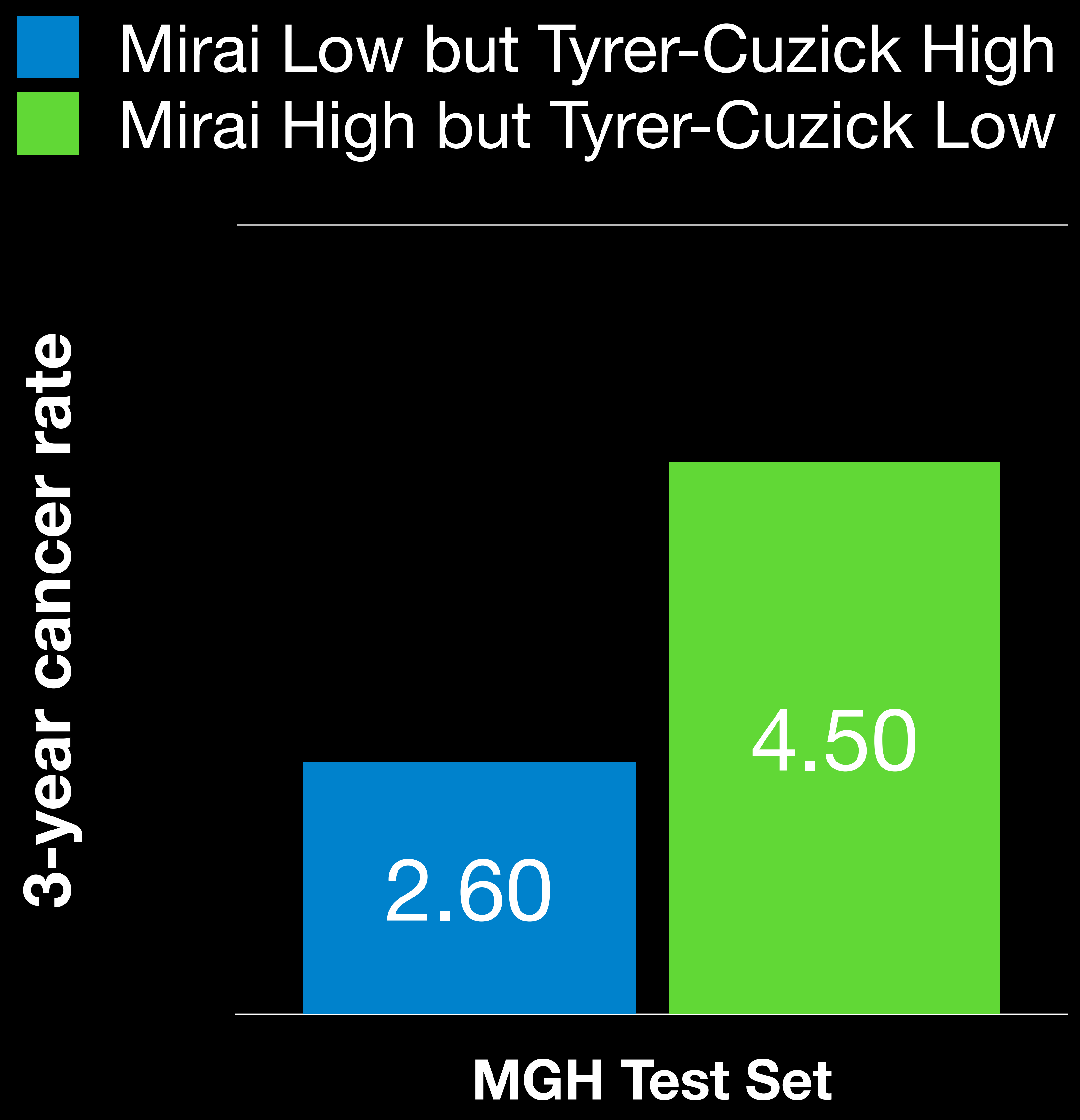


Maintains accuracy across diverse populations





Selecting patients for supplemental imaging



Retrospective analysis

1.7x higher cancer yield, same MRI volume as current care.

Better early detection, same cost.

Mirai Use Cases

Organizing prospective trials for multiple use cases

Highlight: Prioritize screening from covid backlog at MGH

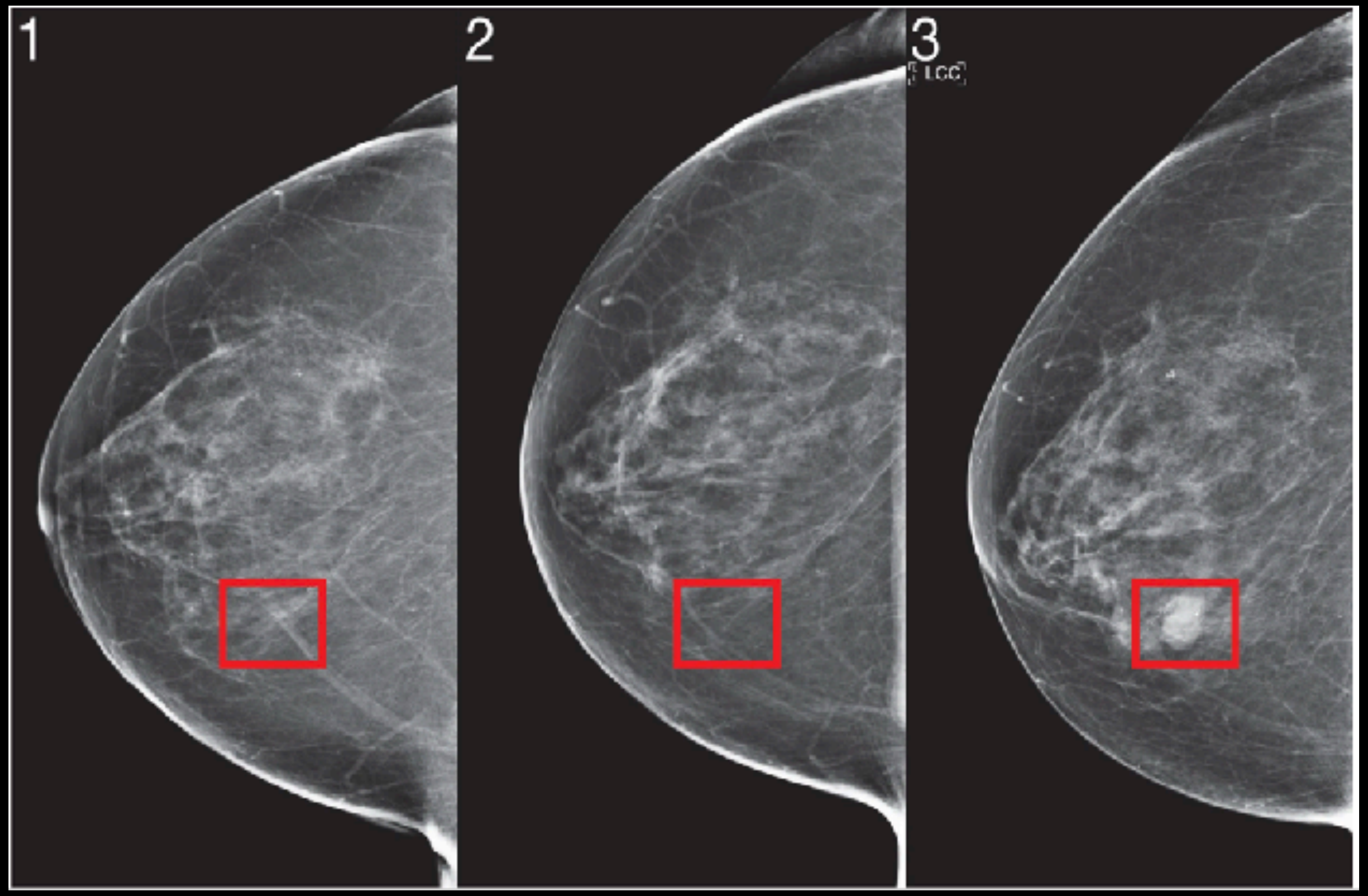

WIRED

BACKCHANNEL BUSINESS CULTURE GEAR IDEAS SCIENCE SECURITY

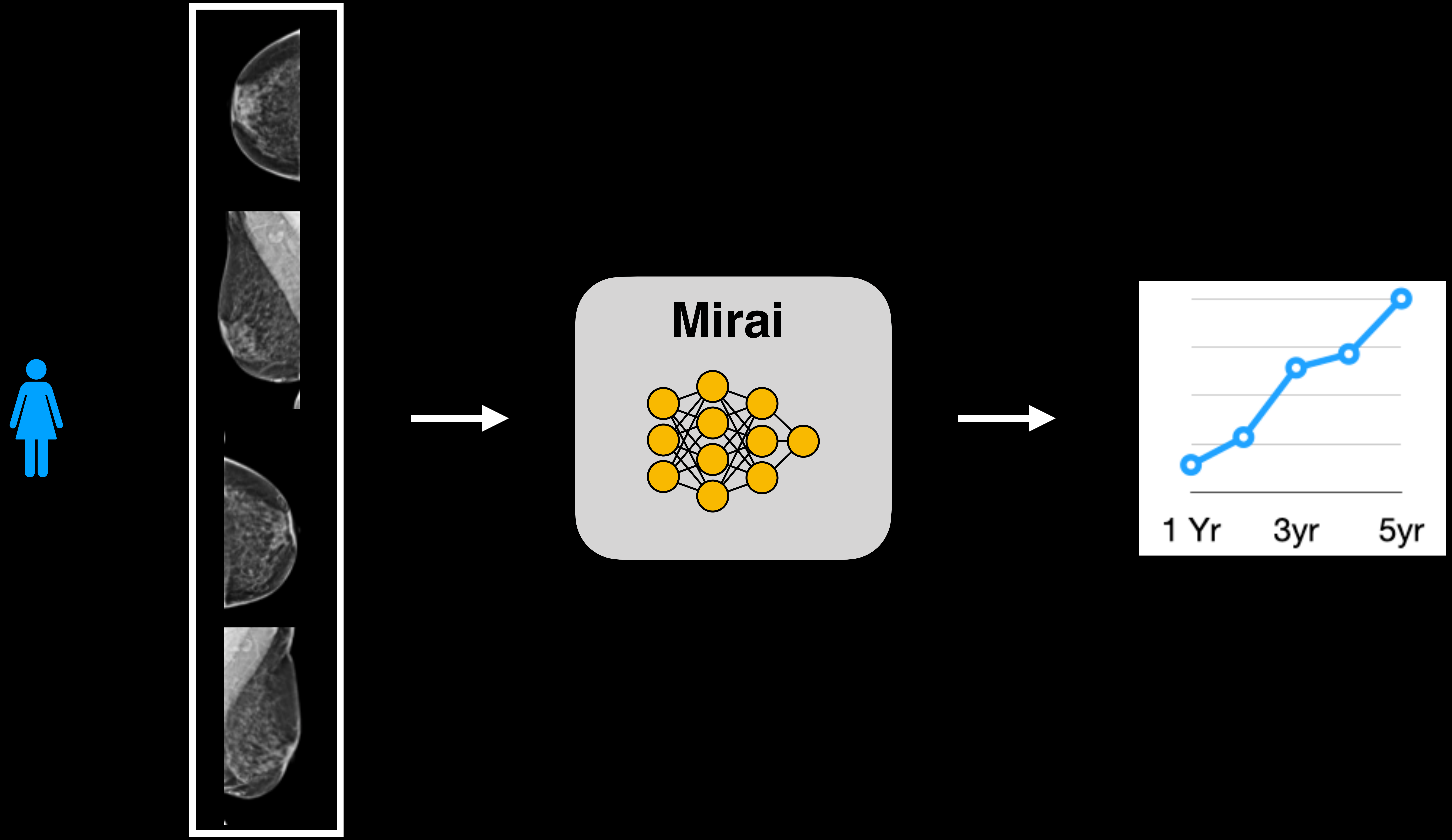
WILL KNIGHT BUSINESS

These Doctors Are Using AI to Screen for Breast Cancer

During the pandemic, thousands of women have skipped scans and check-ups. So physicians tapped an algorithm to predict those at the highest risk.



Mirai: Image-based Risk model



2. ResNet

Standard Architecture
Resnet-18

3. Image DL

Augmentation:
Image Rotations

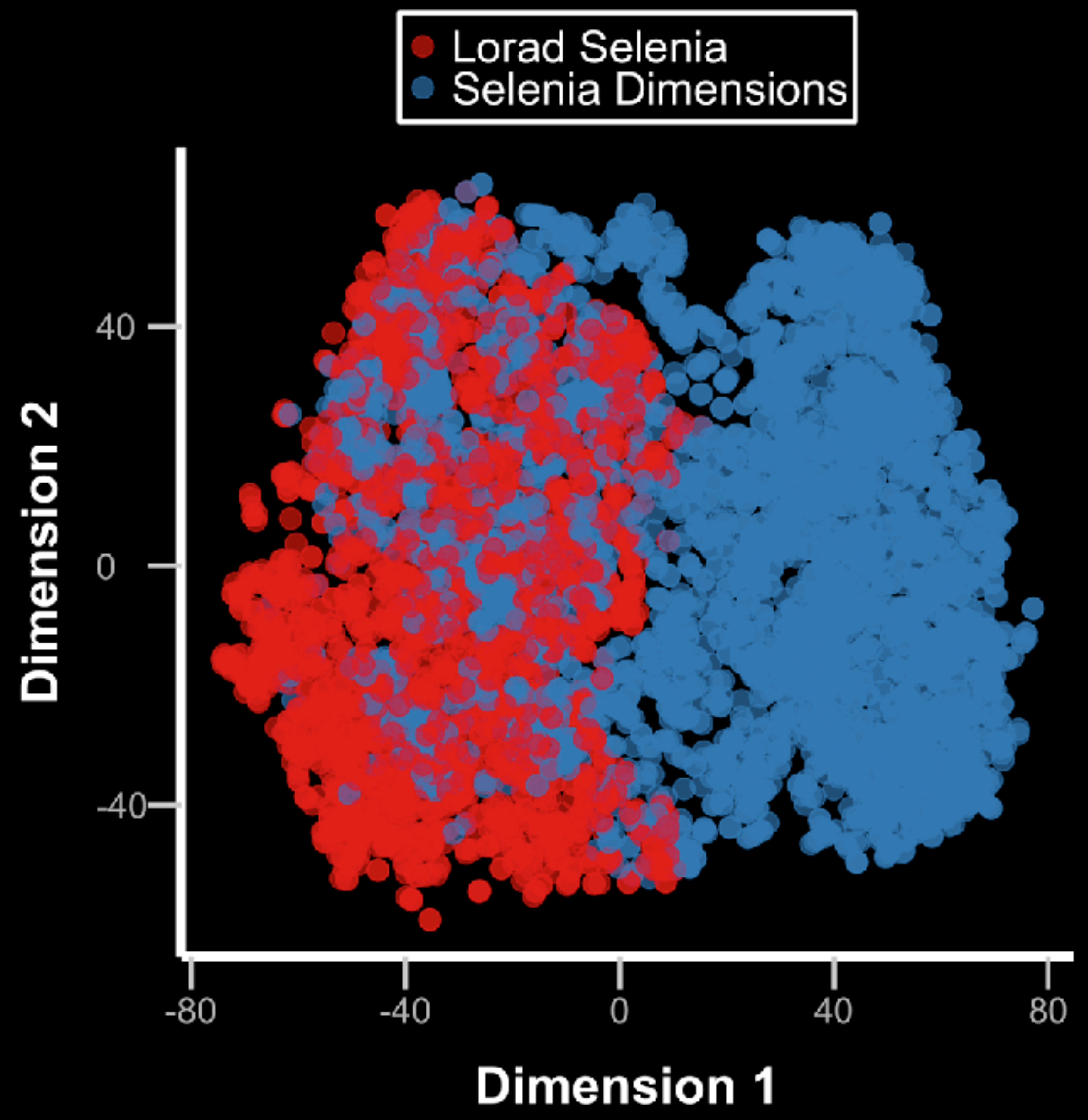
Initialization:
Imagenet

Optimization:
Large Batches, Normalization

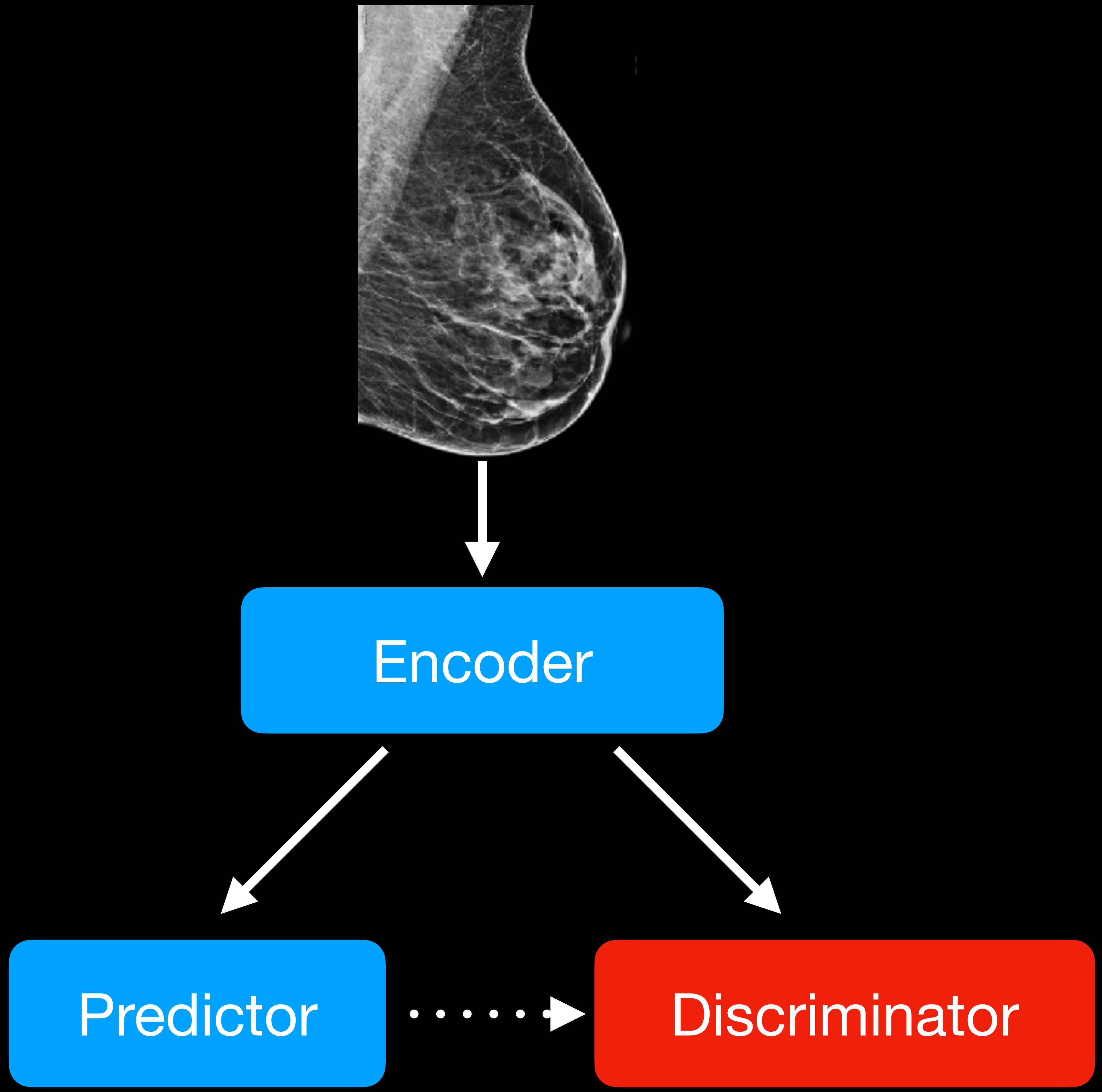
4. MIRAI

Advanced Modeling:
New Objective, Predicting Risk Factors, Multi-Image Modeling, Device Invariance

Problem 1: Device Invariance



Problem 1: Device Invariance

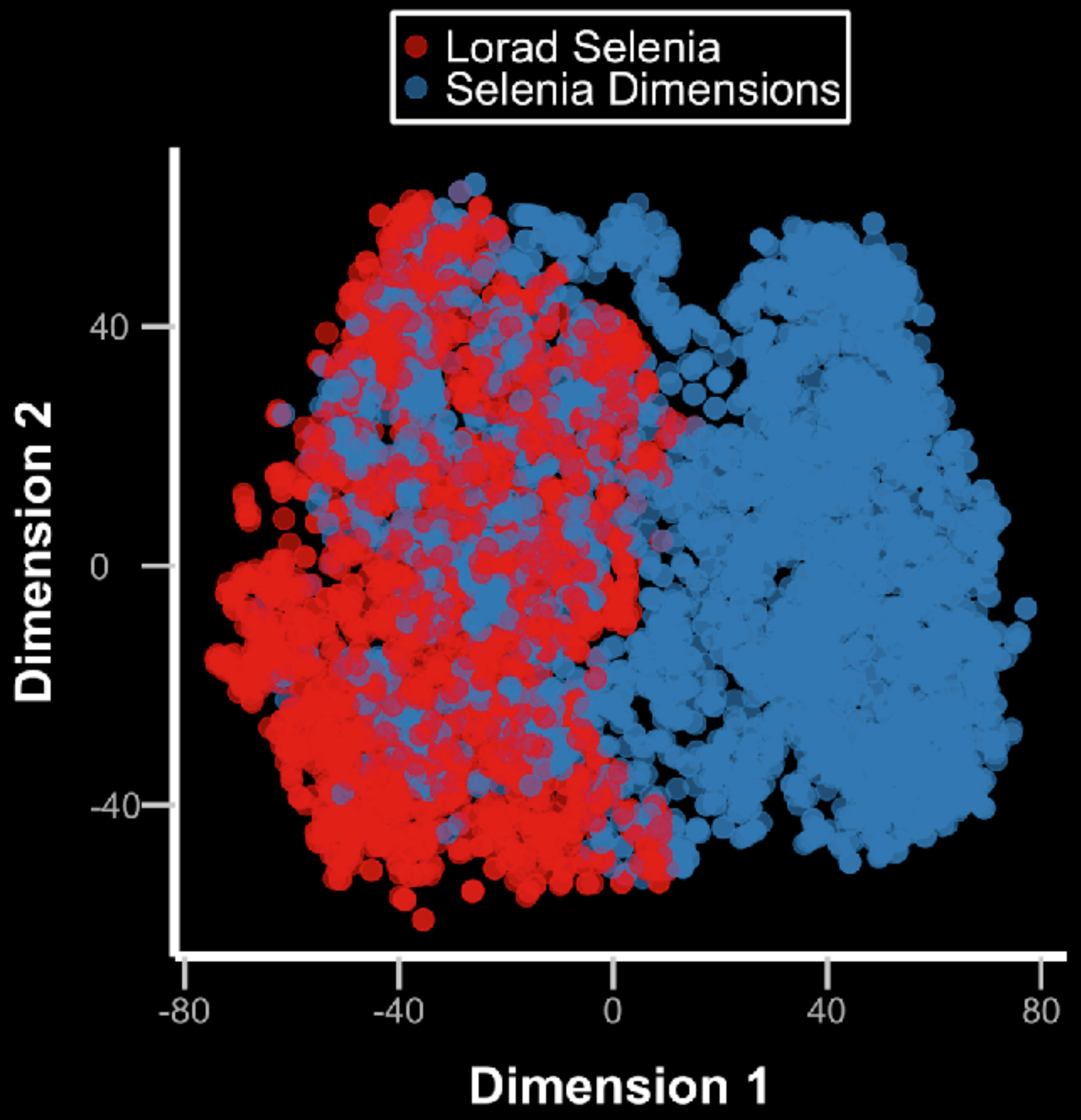


Objective:

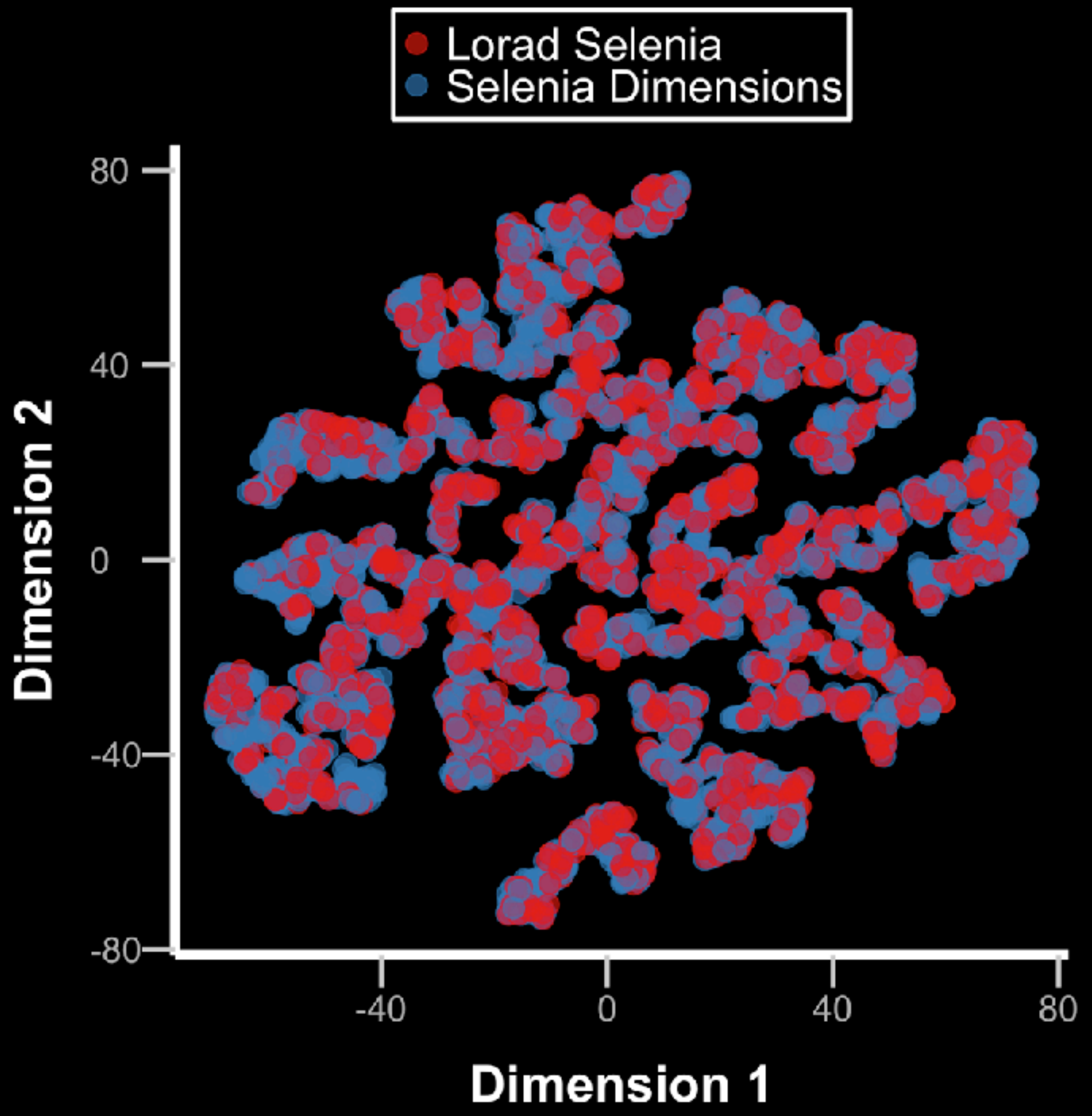
Max accuracy **Predictor**

Min accuracy **Discriminator**

Problem 1: Device Invariance

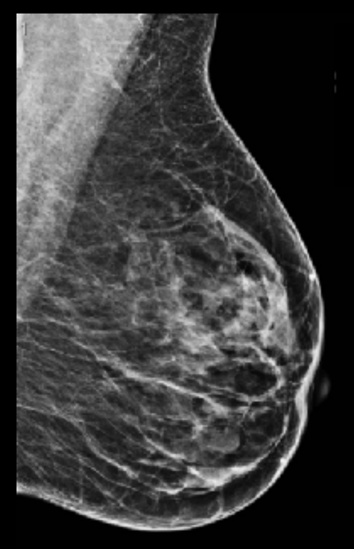


Without Adversary



With Adversary

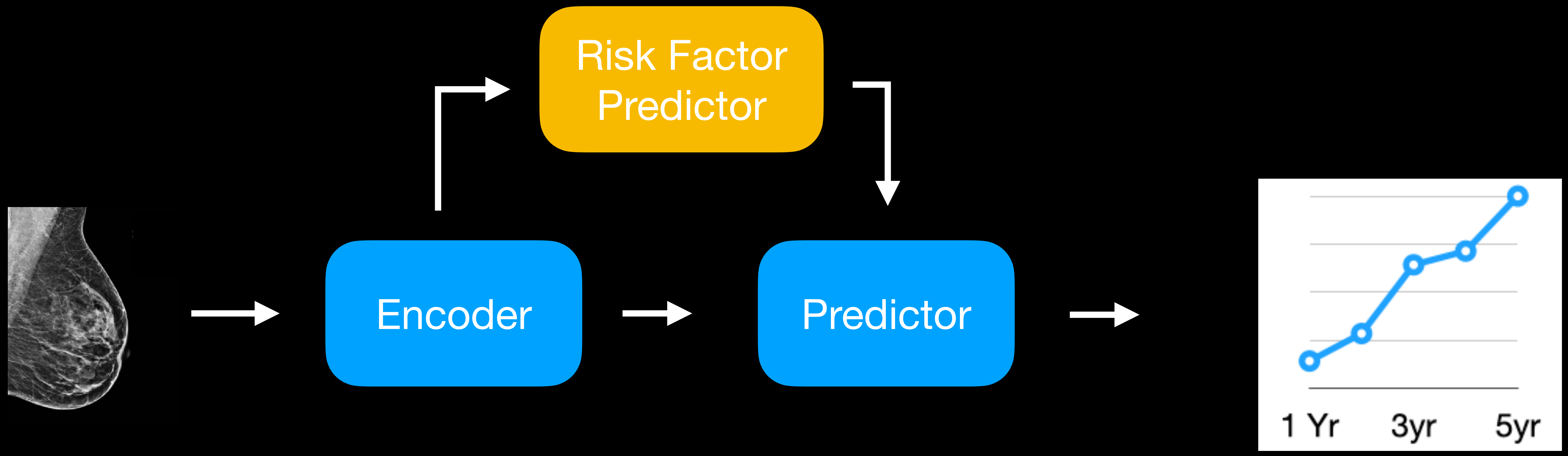
Problem 2: Missing risk factor data



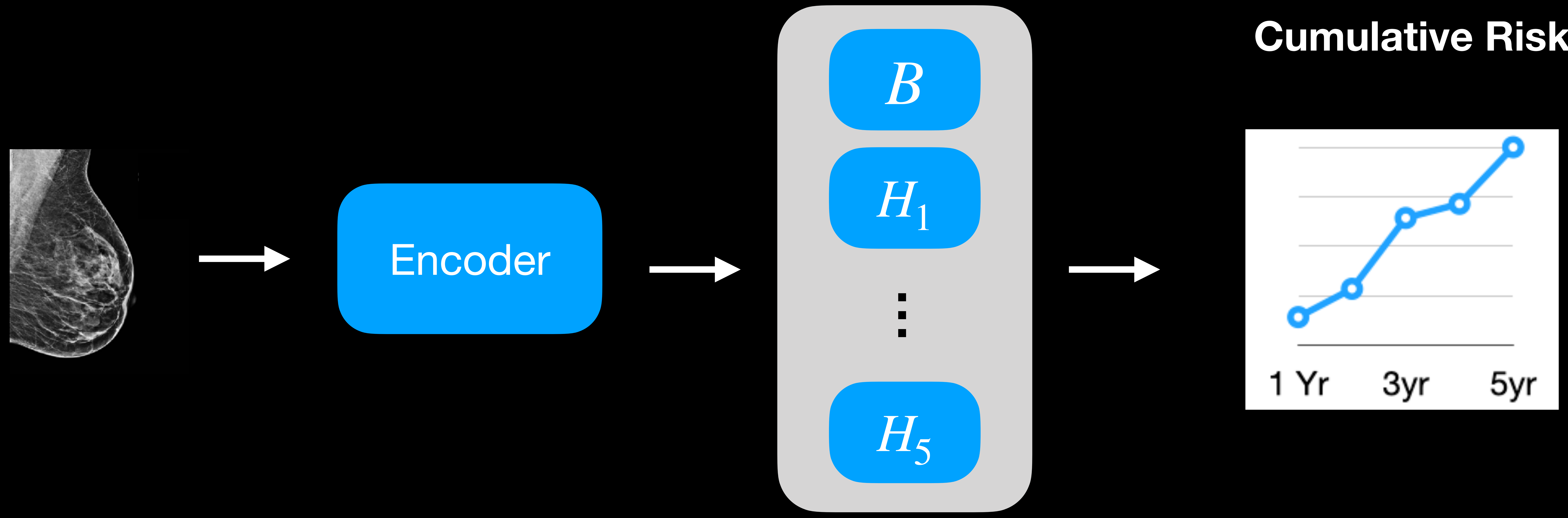
→ Encoder → Risk Factor Predictor →

Risk Factors
Age
Family History
Prior biopsy
Num children
⋮
Menopause status

Problem 2: Missing risk factor data

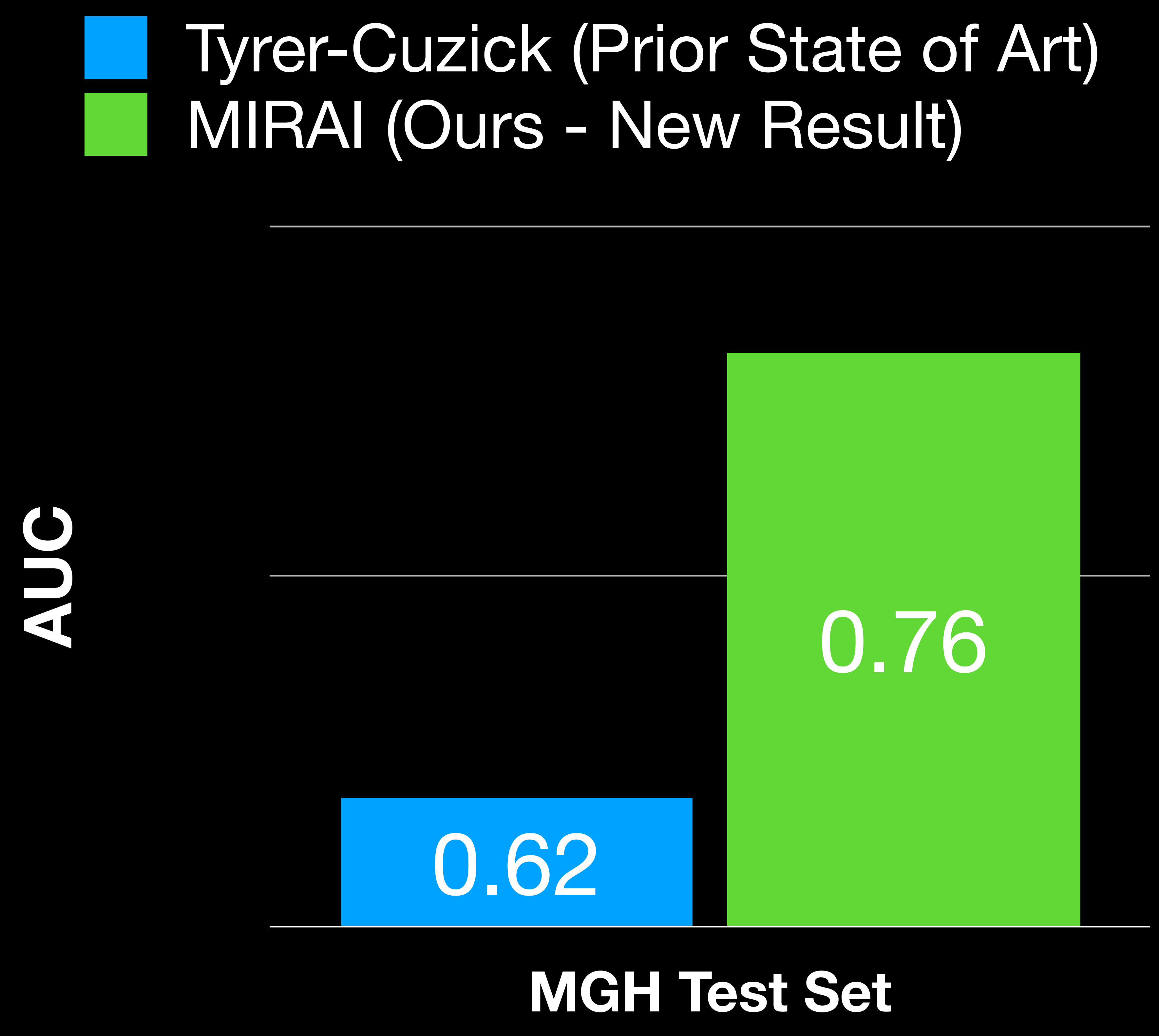


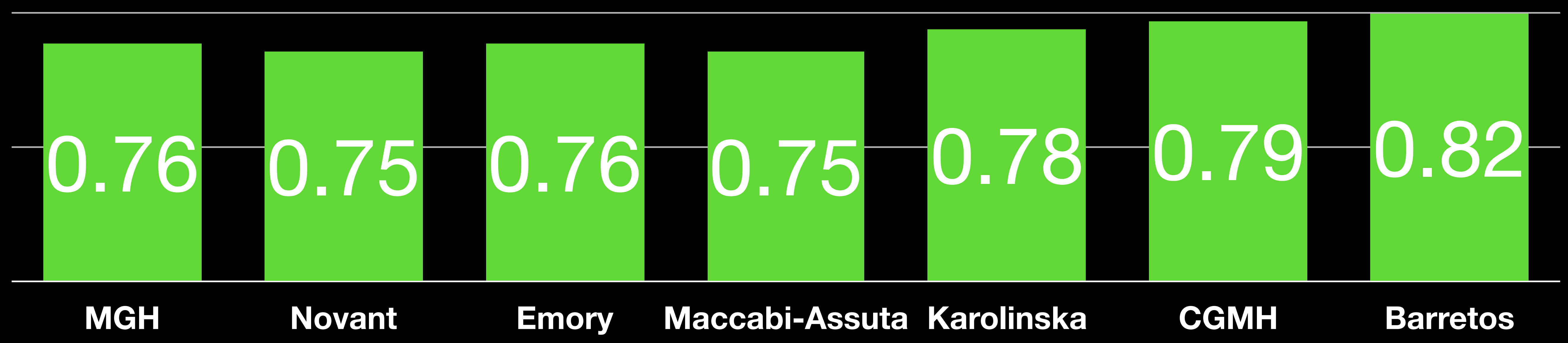
Problem 3: Modeling risk over time



$$P(t_{cancer} = k \mid x) = B(E(x)) + \sum H_i(E(x))$$

Maintains accuracy across diverse populations



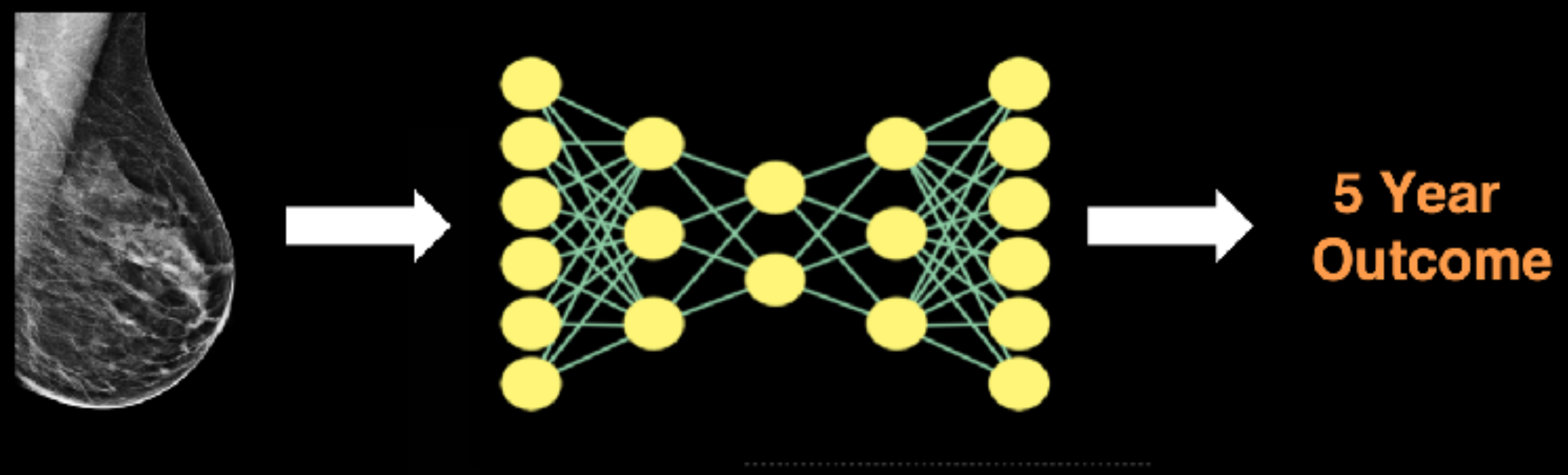


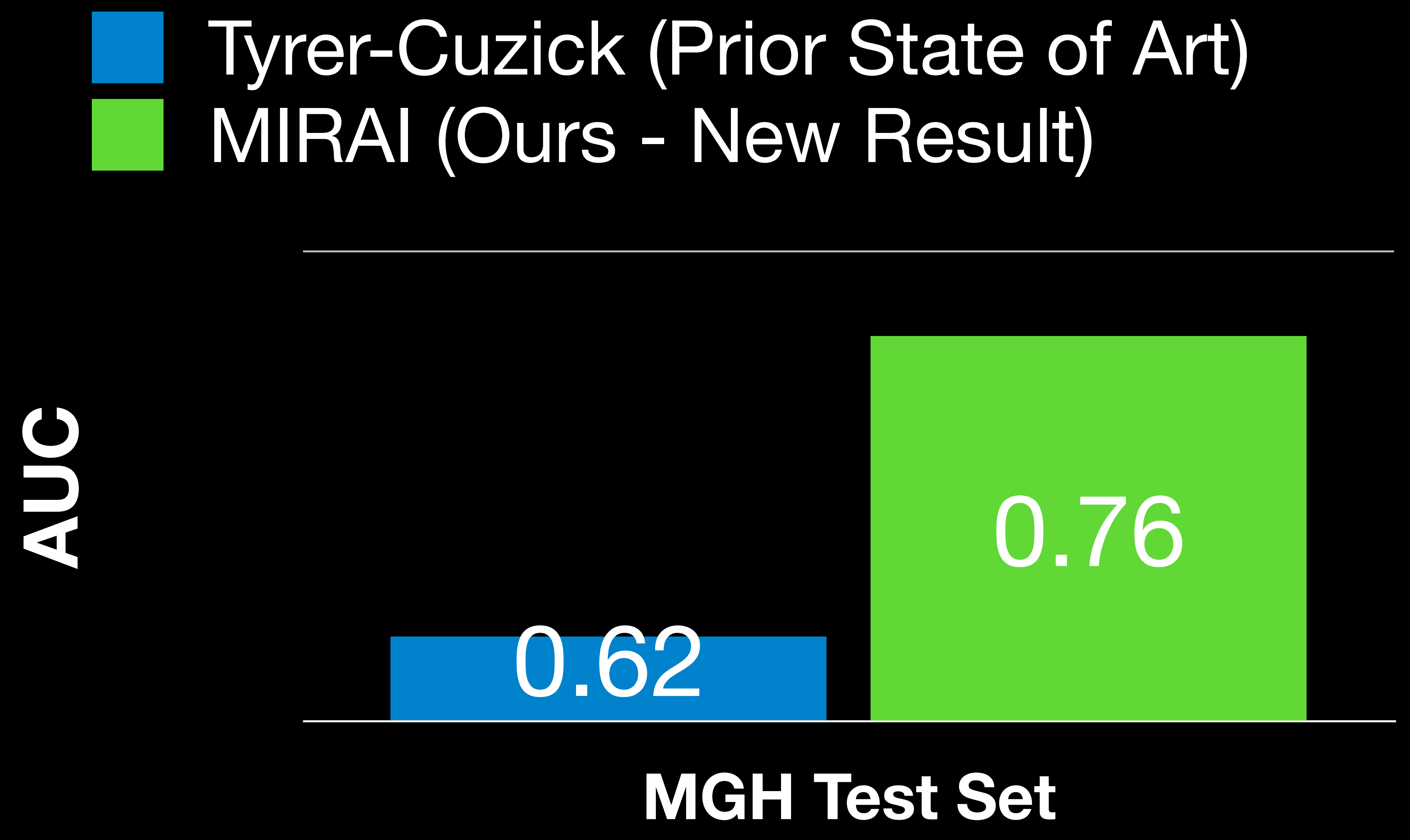
SCIENCE TRANSLATIONAL MEDICINE | RESEARCH ARTICLE

CANCER

Toward robust mammography-based models for breast cancer risk

Adam Yala[1,2*], Peter G. Mikhael[1,2], Fredrik Strand[3,4], Gigin Lin[5], Kevin Smith[6,7], Yung-Liang Wan[5], Leslie Lamb[8], Kevin Hughes[9], Constance Lehman[8†], Regina Barzilay[1,2†]

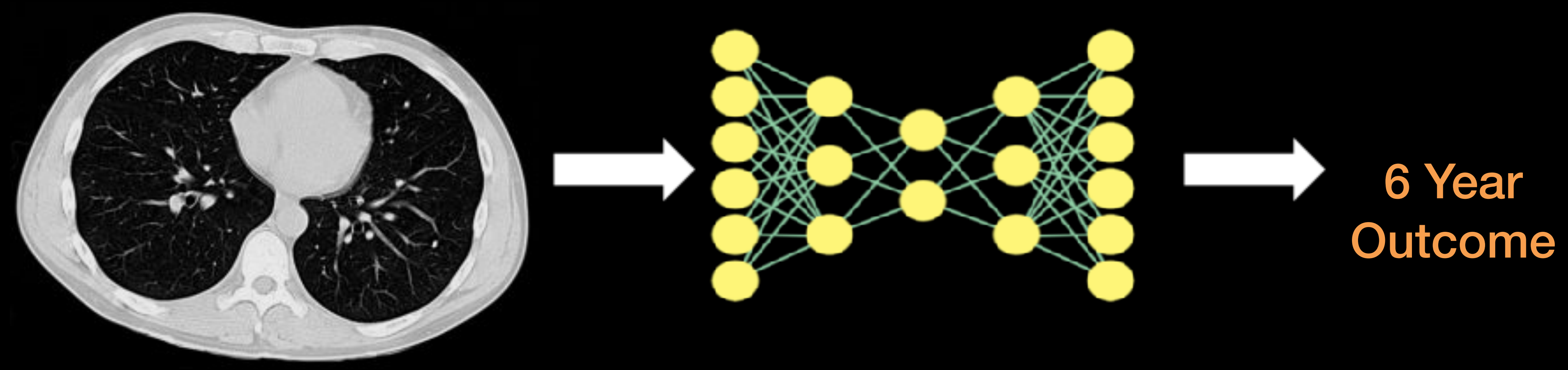


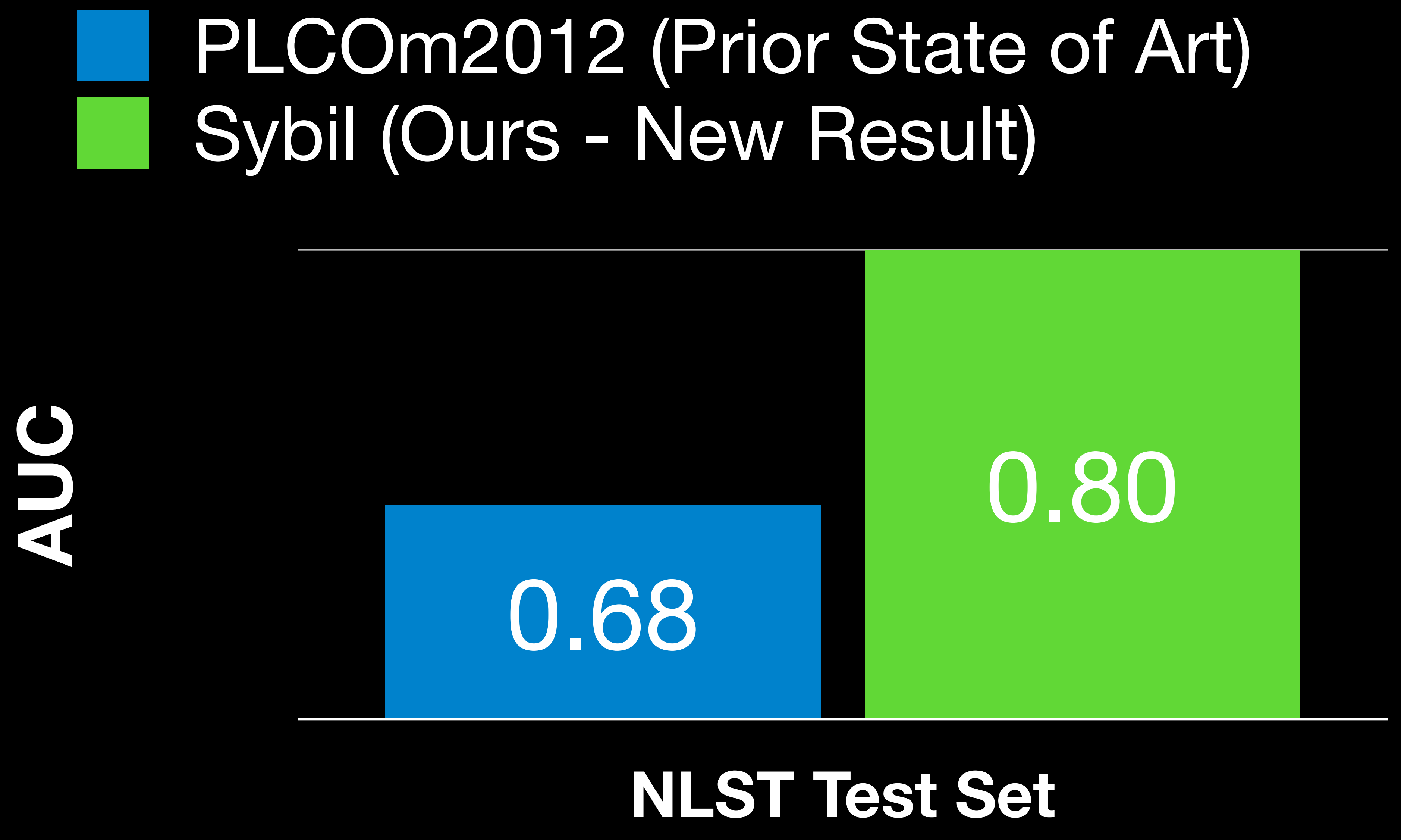


Under Review

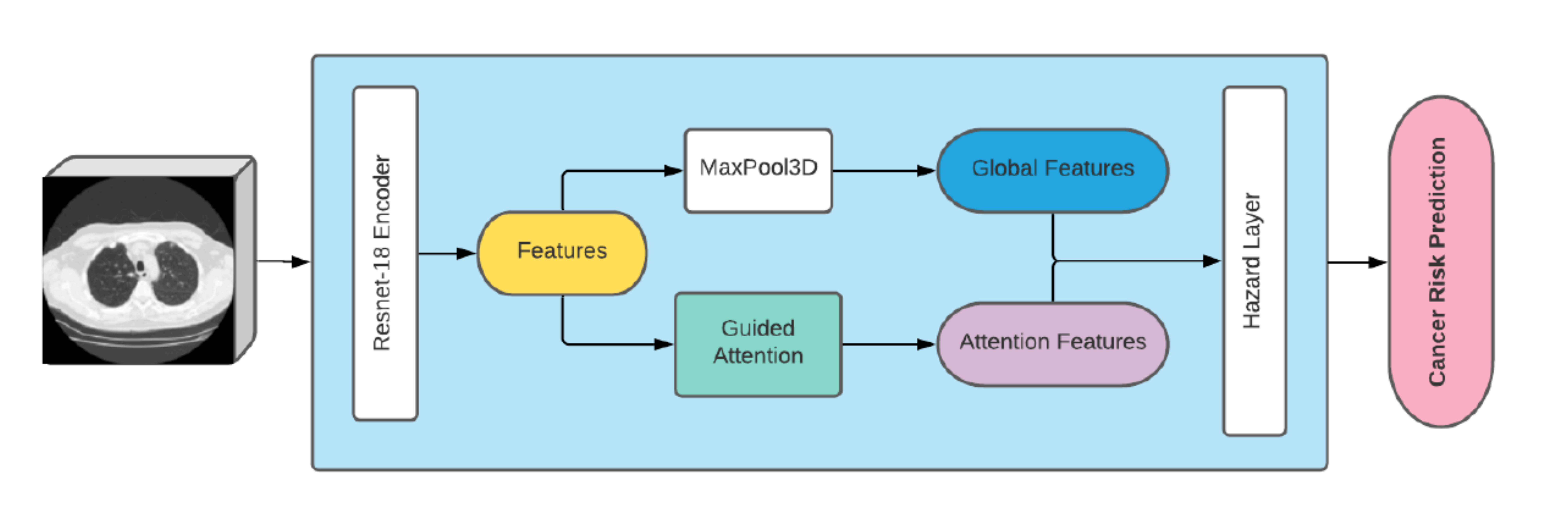
Ask Sybil: Predicting Lung Cancer Risk with Low-dose Chest Computed Tomography

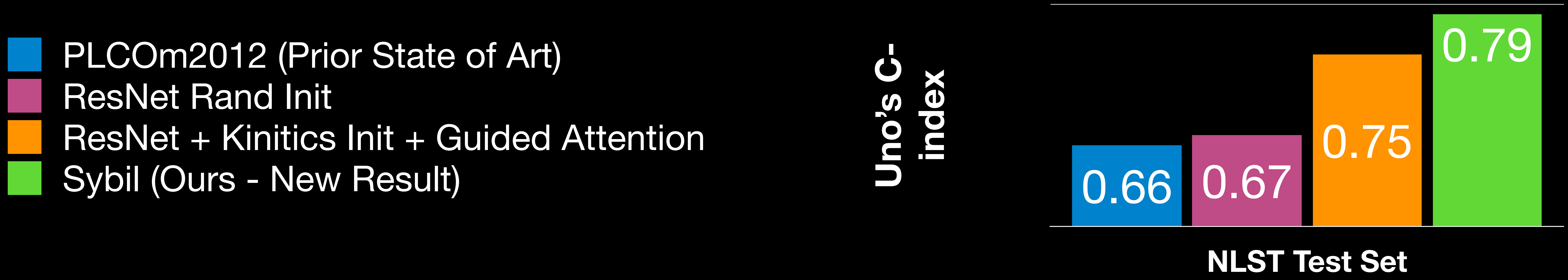
Peter G. Mikhael[1,2,†,*], Jeremy Wohlwend[1,2,†], Adam Yala[1,2], Justin Xiang[1,2], Angelo K. Takigami[3,4], Patrick P. Bourgouin[3,4], PuiYee Chan[5], Sofiane Mrah[4], Lecia V. Sequist[3,5], Florian J. Fintelmann[3,4,‡], Regina Barzilay





Sybil Architecture

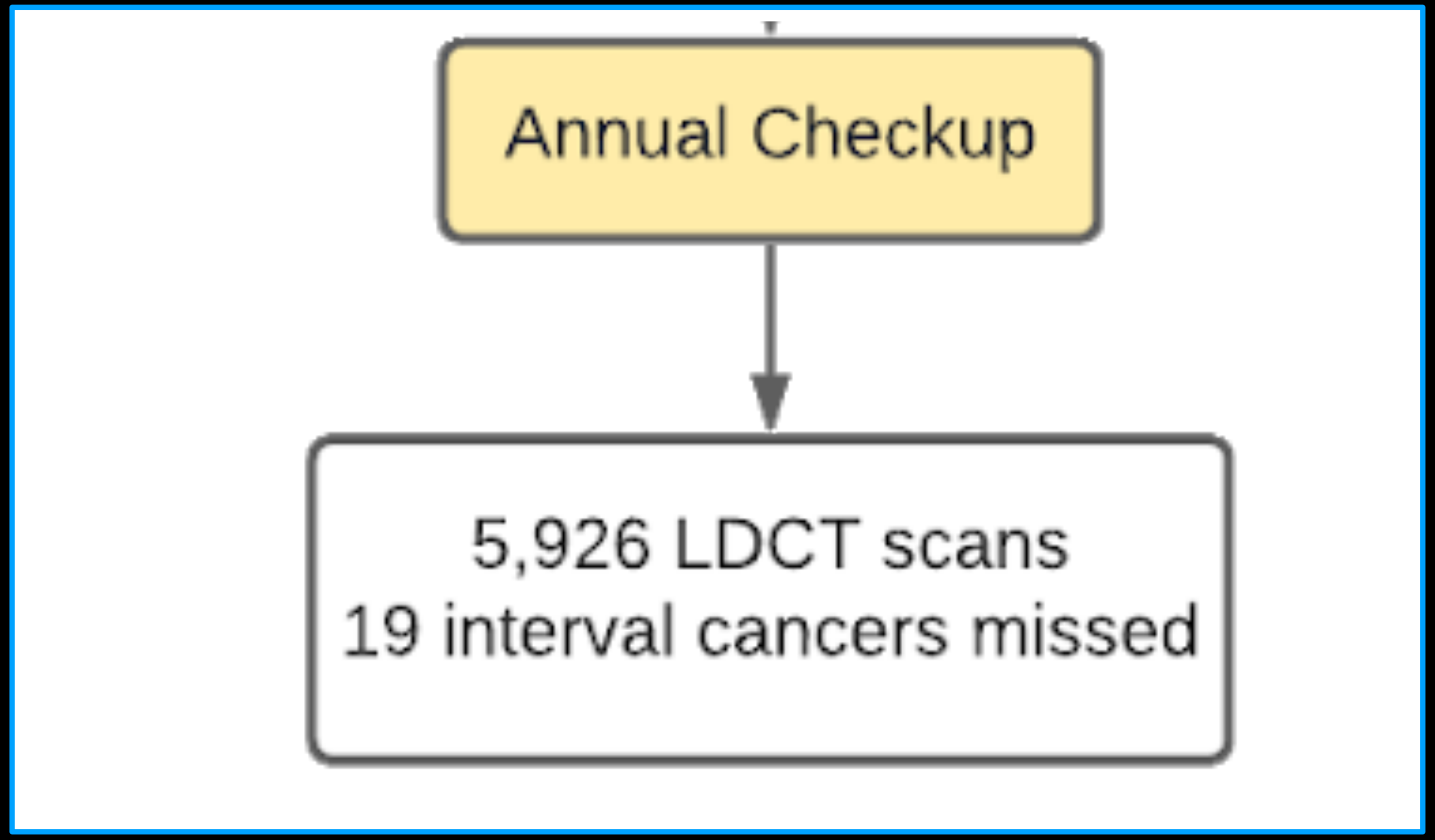




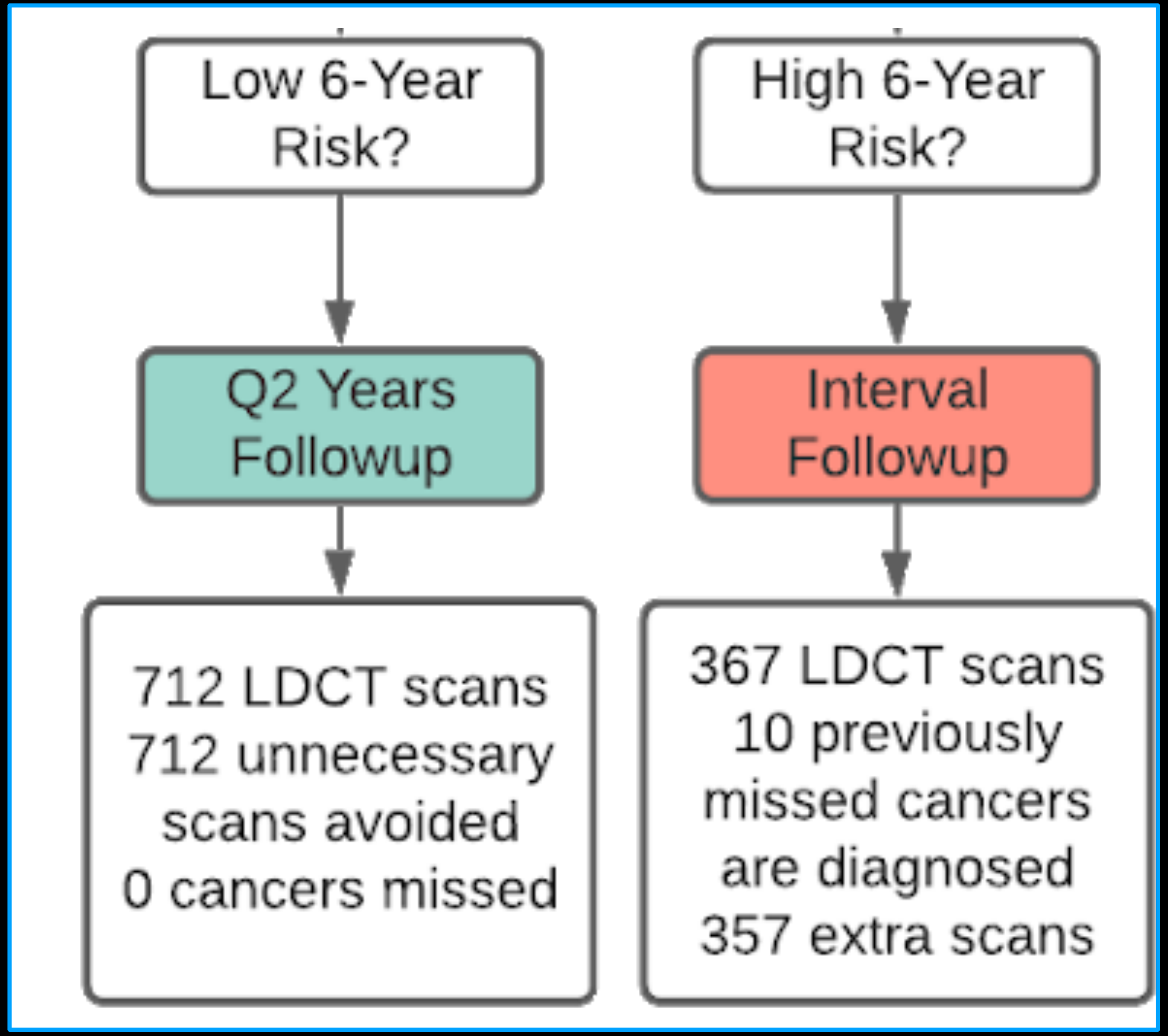
Sybil Clinical Impact : Workflow

Improve early detection and lower screening cost with risk-based followup.

Standard of Care

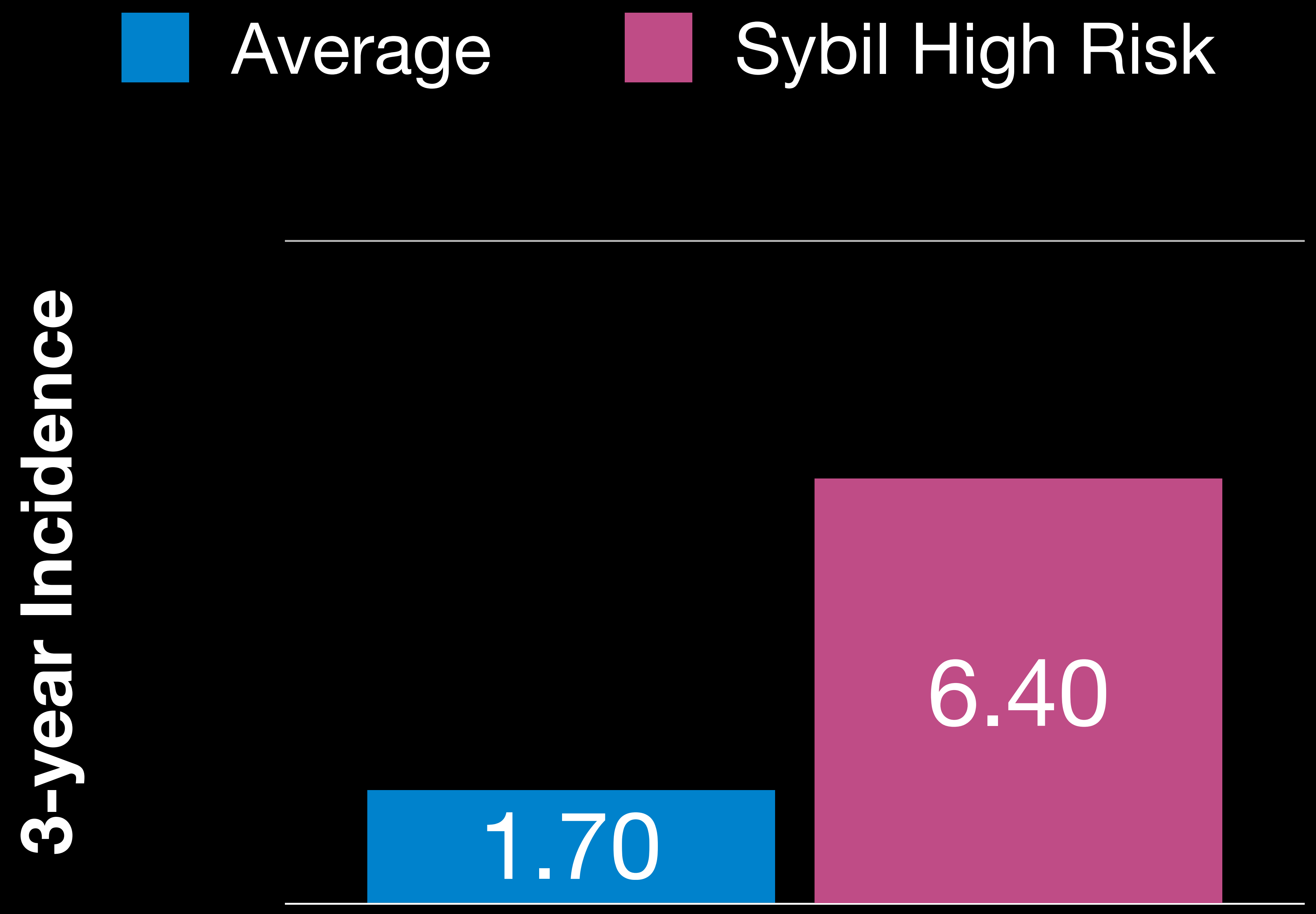


Sybil risk based screening



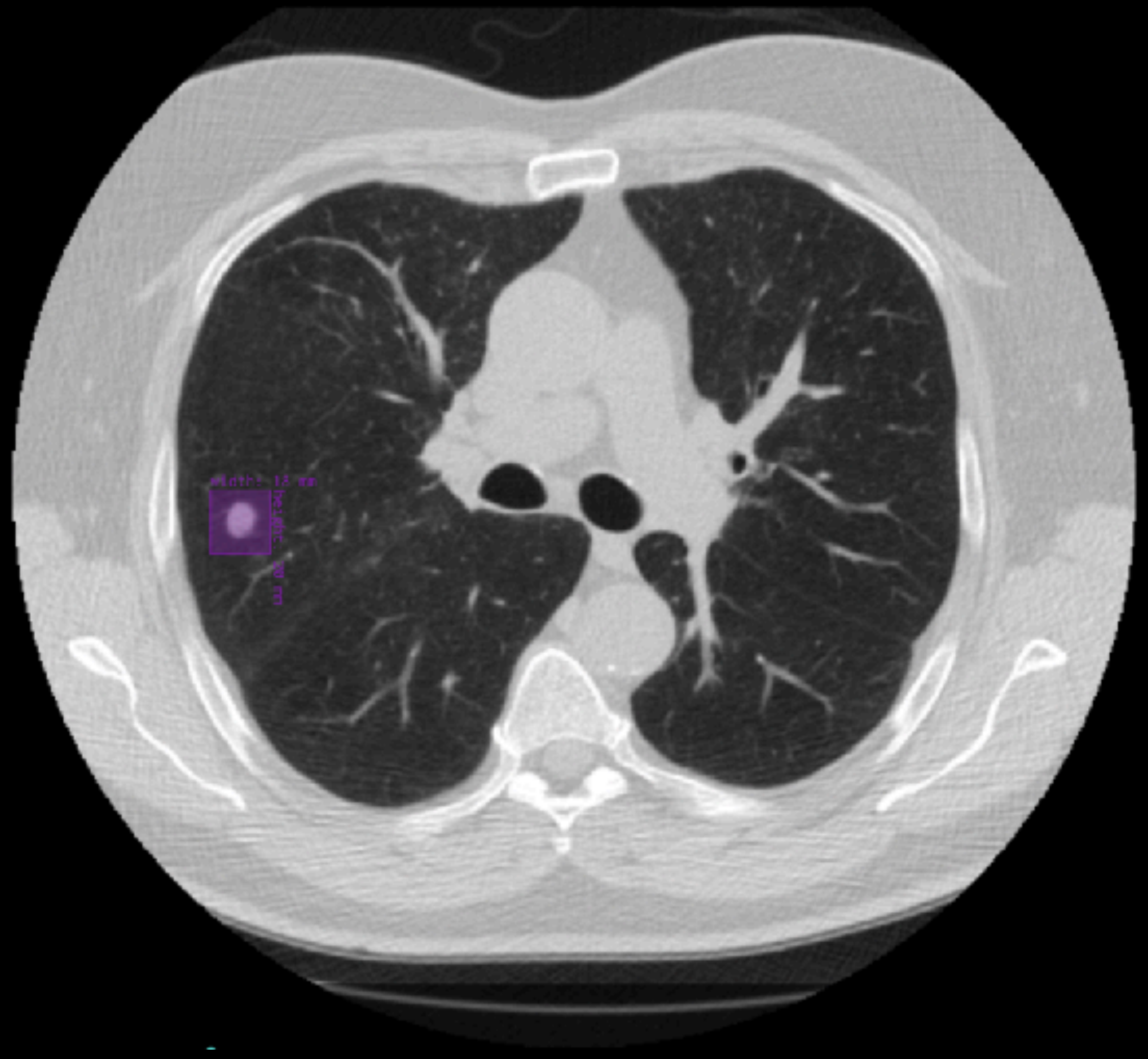
Sybil Clinical Impact : Prevention

Identifying high risk cohorts for clinical trails



Identifying future cancer location

83% accurate predicting future cancer side



Agenda

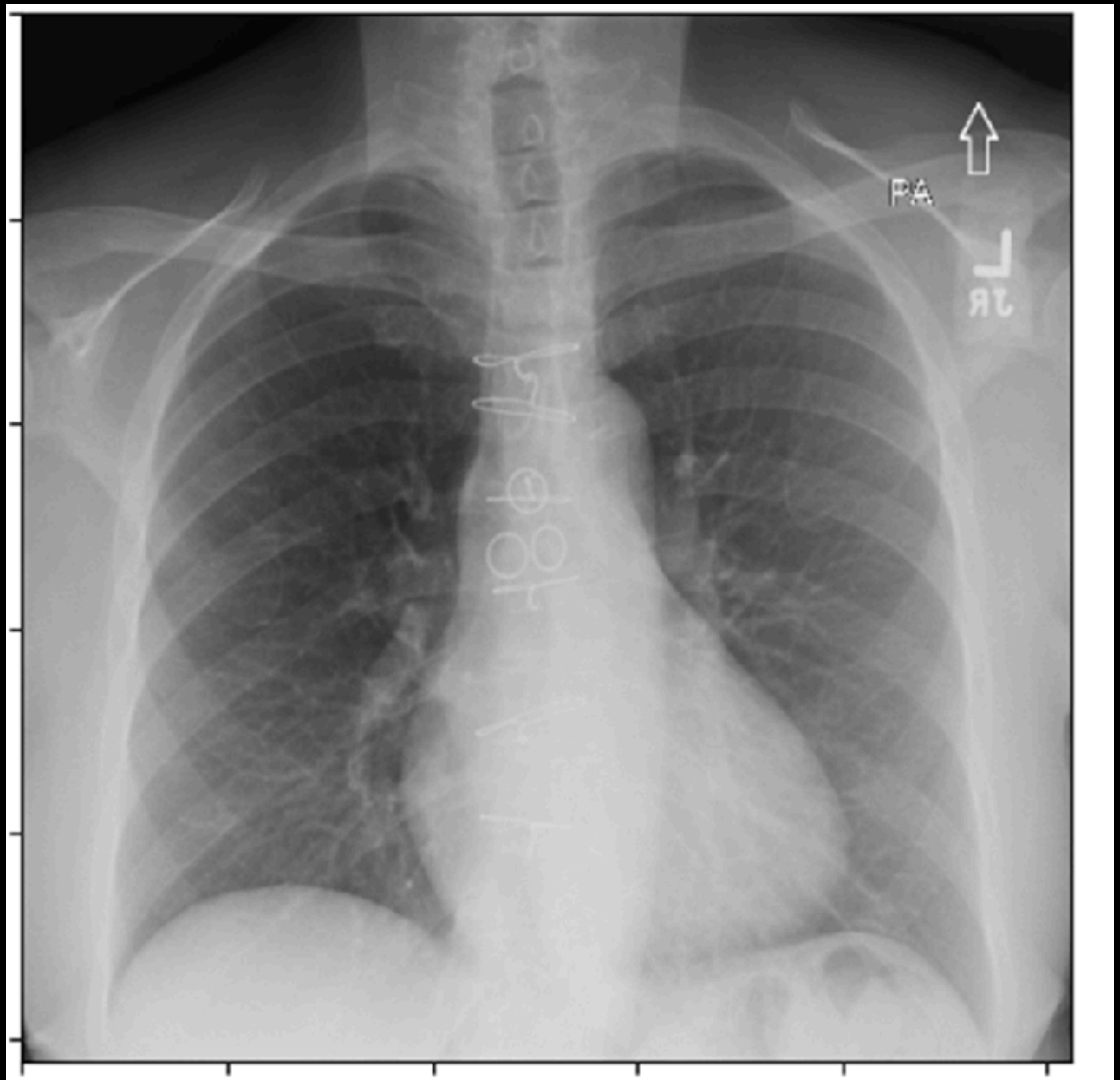
Data: What is medical imaging?

Method Foundations: How do we build models on imaging data?

Applications: How can we catch cancer earlier?

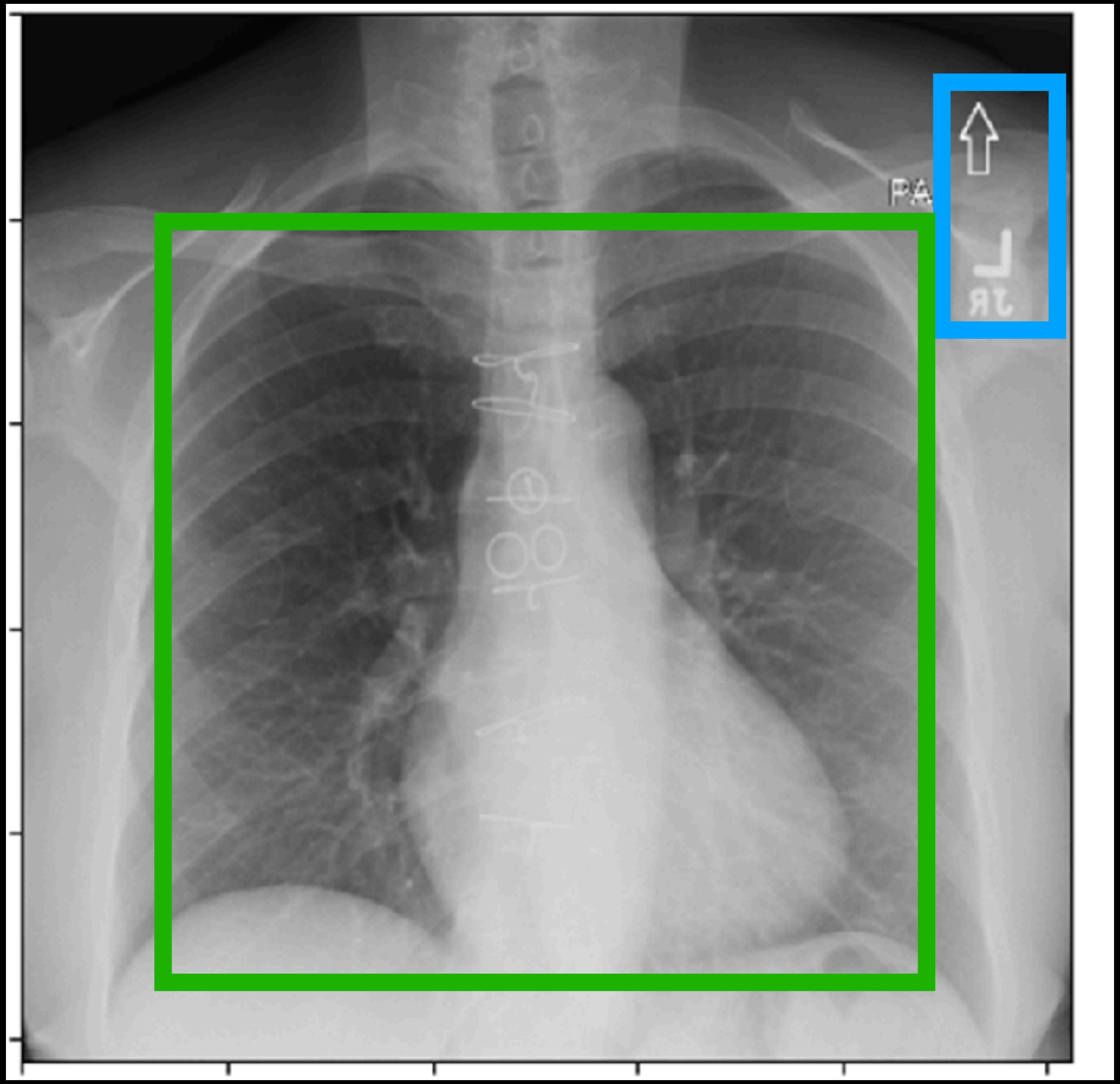
Interpretation: How can we audit our models?

Visualizing model behavior



Question:
What is the model looking at?

Visualizing model behavior



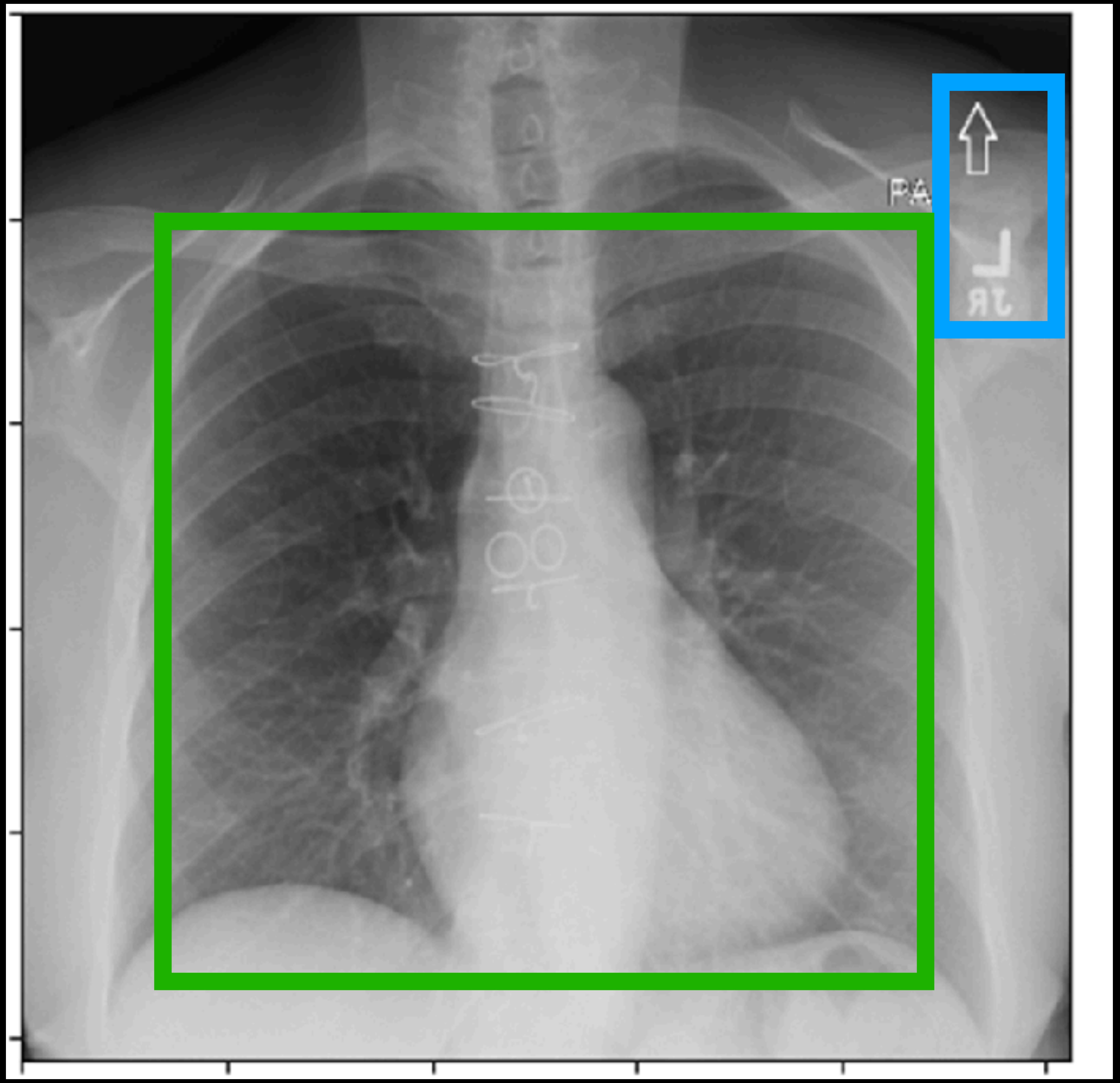
Question:
What is the model looking at?

Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study

John R. Zech, Marcus A. Badgeley, Manway Liu, Anthony B. Costa, Joseph J. Titano, Eric Karl Oermann

Published: November 6, 2018 • https://doi.org/10.1371/journal.pmed.1002683

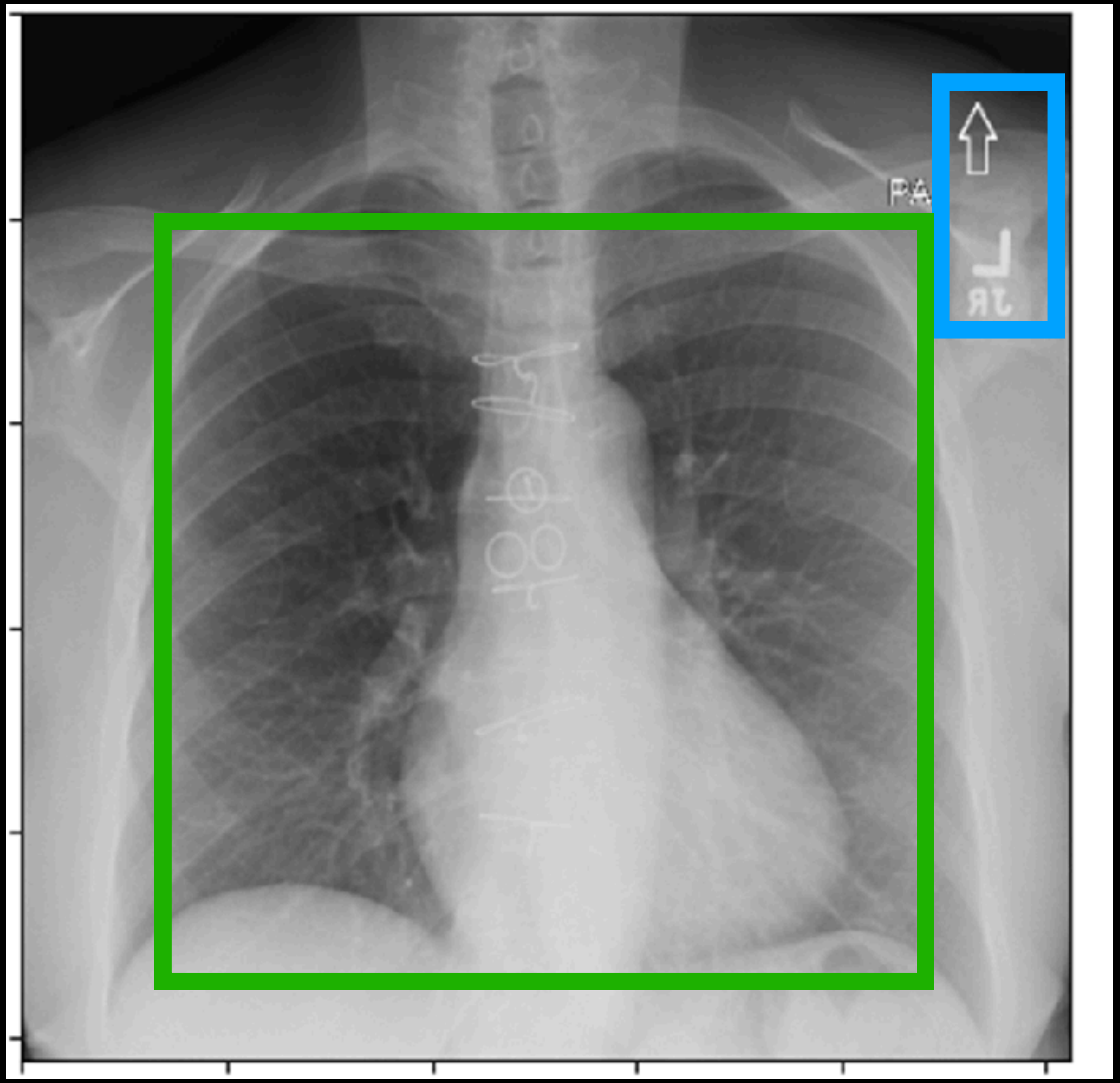
Visualizing model behavior



Question:
What is the model looking at?

Key idea: *Saliency Maps*
What inputs changing would change model predictions?

Visualizing model behavior

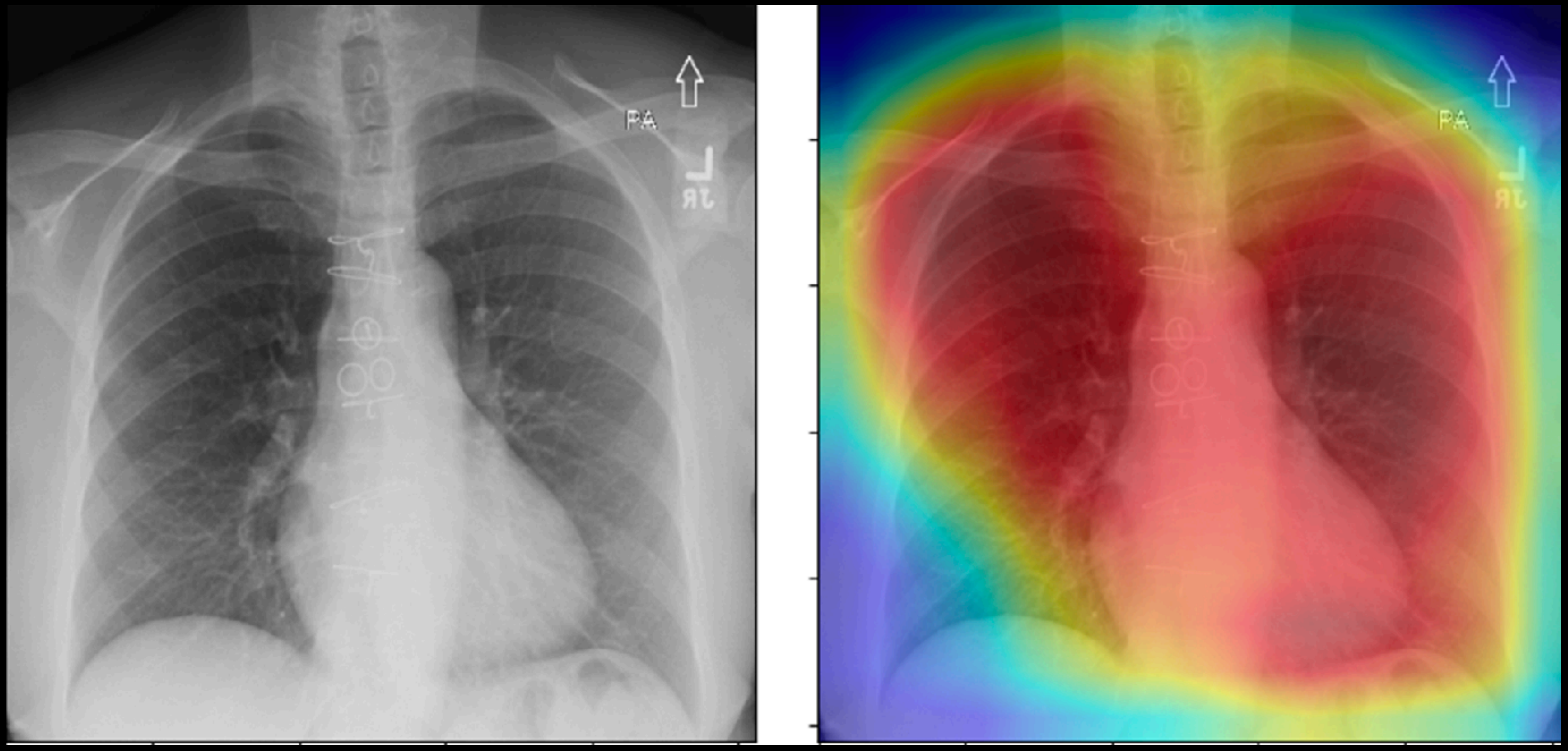


Question:
What is the model looking at?

Key idea: *Saliency Maps*
What inputs changing would change model predictions?

Compute gradient of predict in respect to **input**

Saliency Maps



Science Contents News Careers Journals

SHARE

RESEARCH ARTICLE

Dissecting racial bias in an algorithm used to manage the health of populations

Ziad Obermeyer[1,2,*], Brian Powers[3], Christine Vogeli[4], Sendhil Mullainathan[5,*,†]

+ See all authors and affiliations

Science 25 Oct 2019:
Vol. 366, Issue 6464, pp. 447-453
DOI: 10.1126/science.aax2342

What software was this? The researchers didn't say, but the Washington Post identifies it as Optum, owned by insurer UnitedHealth. It says its product is used to "manage more than 70 million lives." Though the researchers only focused on one particular tool, they identified the same flaw in the 10 most widely used algorithms in the industry. Each year, these tools are collectively applied to an estimated 150 to 200 million people in the US.

Handling Diversity

A.I. Could Worsen Health Disparities

In a health system riddled with inequity, we risk making dangerous biases automated and invisible.

By Dhruv Khullar

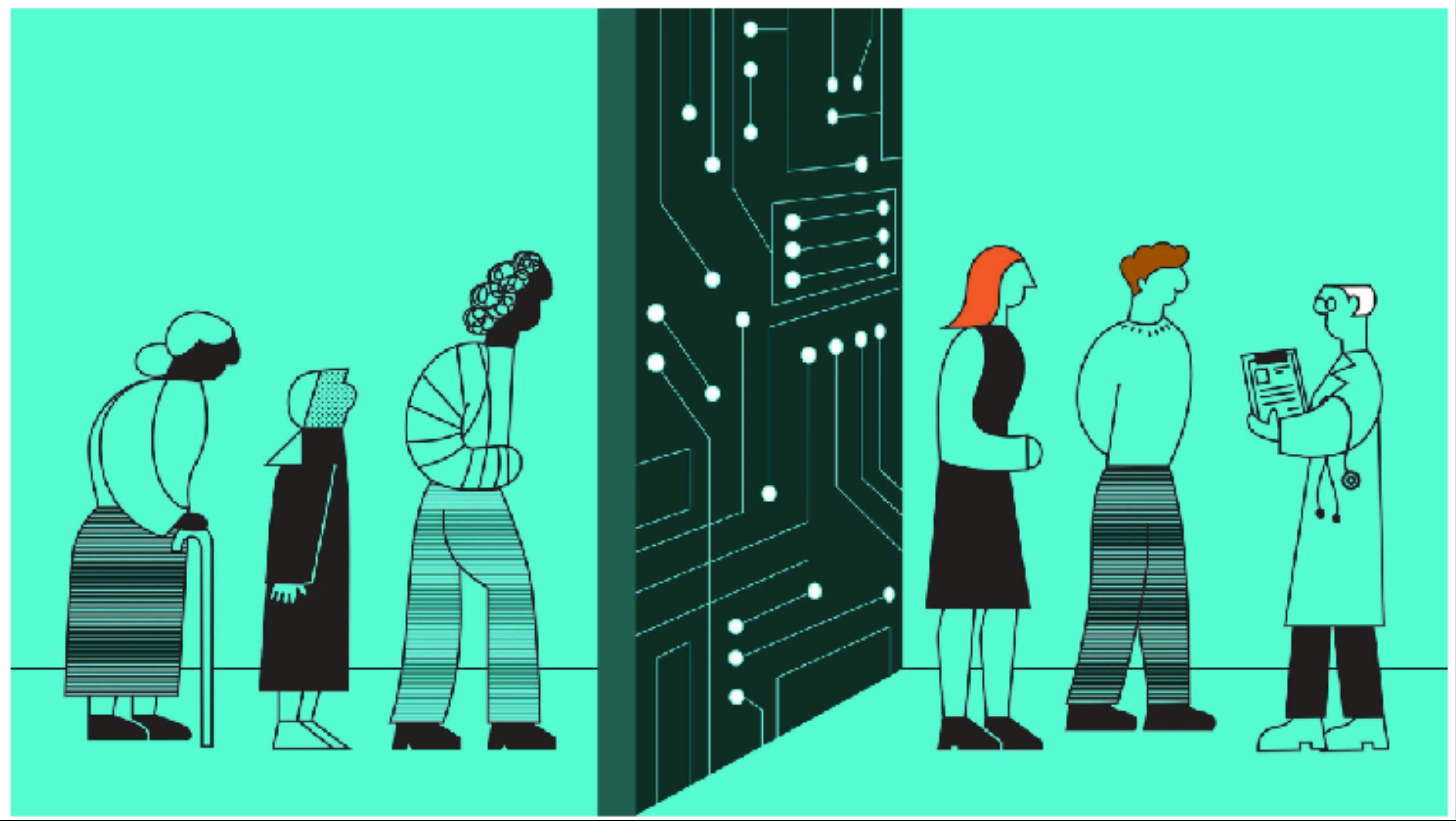
Dr. Khullar is an assistant professor of health care policy and research.

HEALTH TECH — STAT+

AI systems are worse at diagnosing disease when training data is skewed by sex

By REBECCA ROBBINS @rebeccadrobbins / MAY 25, 2020

Reprints



Performance Audits

Validate model performance across diverse populations

Test model performance by demographic group

Test model performance by imaging device, clinical setting, etc.

Summary

Data: Tissue response to generator energy

Method Foundations: Convolutional Neural Networks

Applications: Predicting future disease in breast and lung cancer

Interpretation: Saliency Maps and Performance Audits

Questions?