GE 461: INTRODUCTION TO DATA SCIENCE Spring 2022-2023

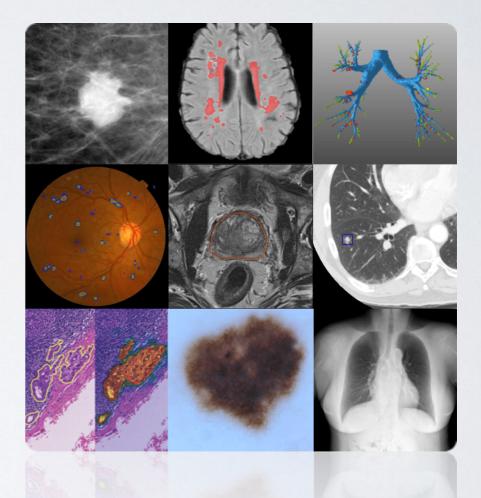


Machine Learning in Healthcare Tolga Çukur

Machine Learning in Healthcare

Outline

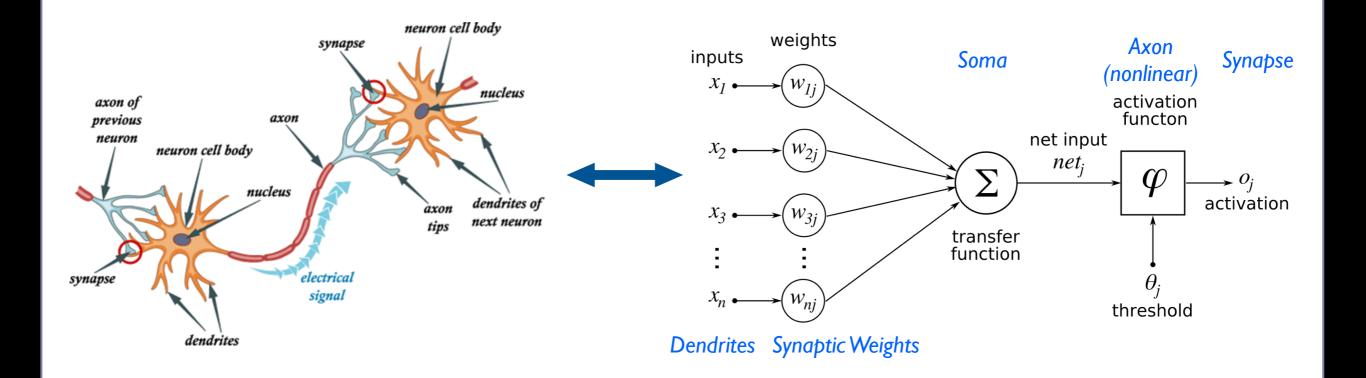
- Overview of Neural Networks
- History of Machine Learning in Medicine
- Big Data in Medical Applications
- Opportunities/Challenges in Healthcare
- Utility of Machine Learning in Medical Imaging
- Example Applications in Medical Imaging



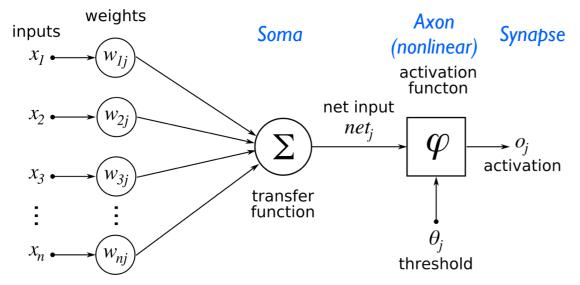
PART I: Overview of Neural Networks

Artificial Neuron: A mathematical abstraction

Perceptron Model (McCulloch-Pitts)



Single Neuron: A linear classifier



Dendrites Synaptic Weights

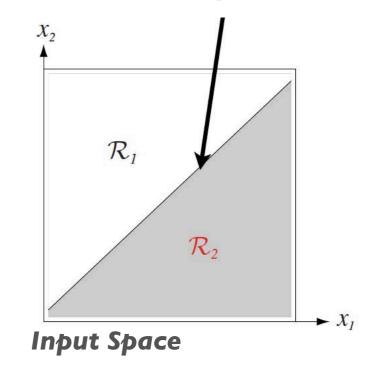
Output

 x_{l}

Inputs

 X_2

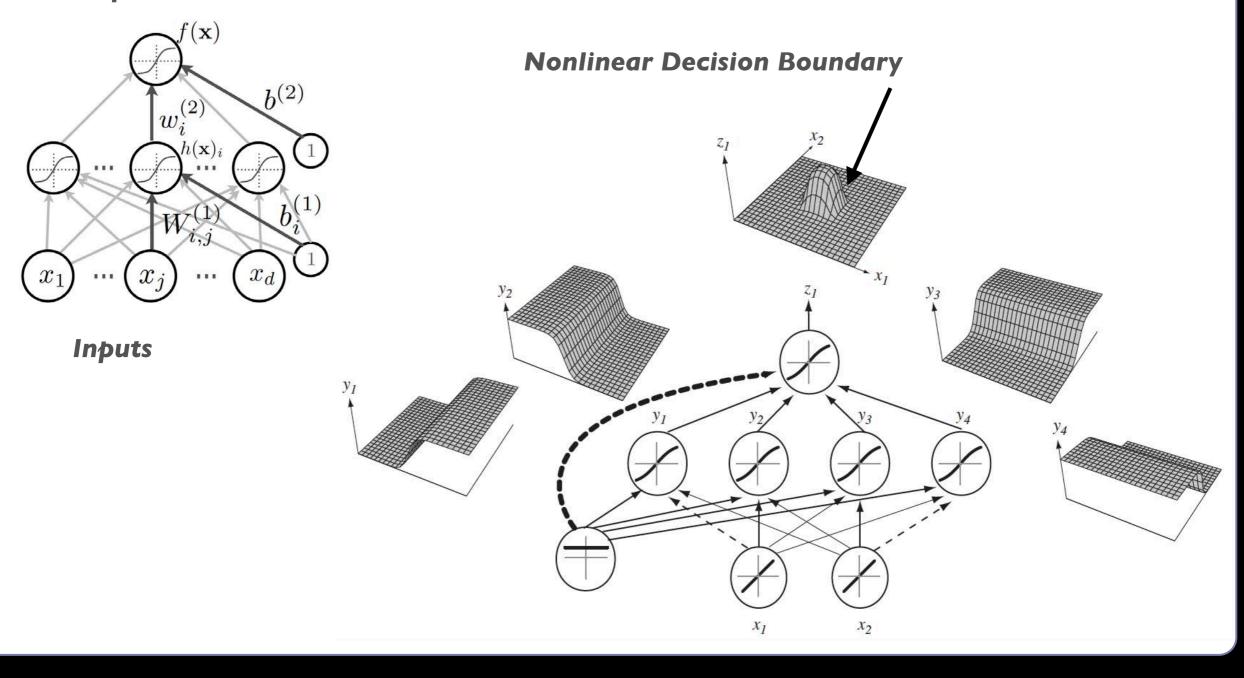
Linear Decision Boundary



Neural Network: Nonlinear mapping

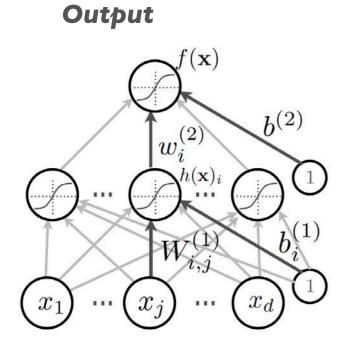
Single Hidden-Layer Network

Output



Universal Approximation Theorem

Multi-Layer Neural Network





- Early 1990s for single hidden-layer networks
- A universal approximator
- Model any continuous nonlinear function (given a sufficient number of neurons)
- <u>No guidance on how to find model parameters...</u>

Why is Deep Learning Hot Today?

Big Data Availability

faceleonix.

Walmart 2<mark>1</mark>5

You Auto

350 millions images uploaded per day 2.5 Petabytes of customer data hourly

100 hours of video uploaded every minute

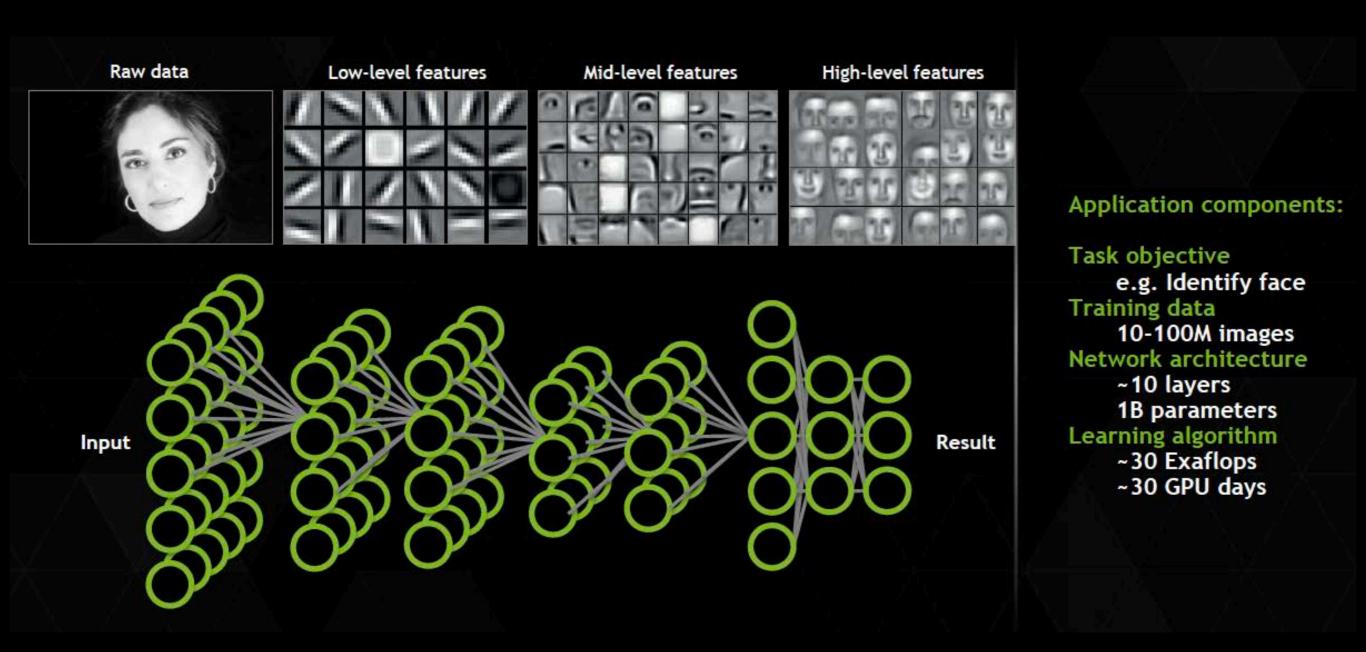
New DL Techniques



GPU acceleration



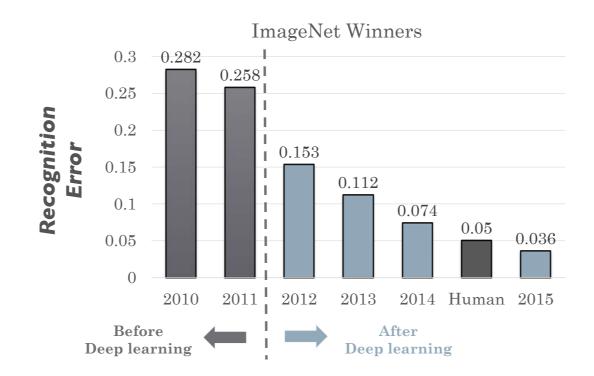
Deep Neural Networks



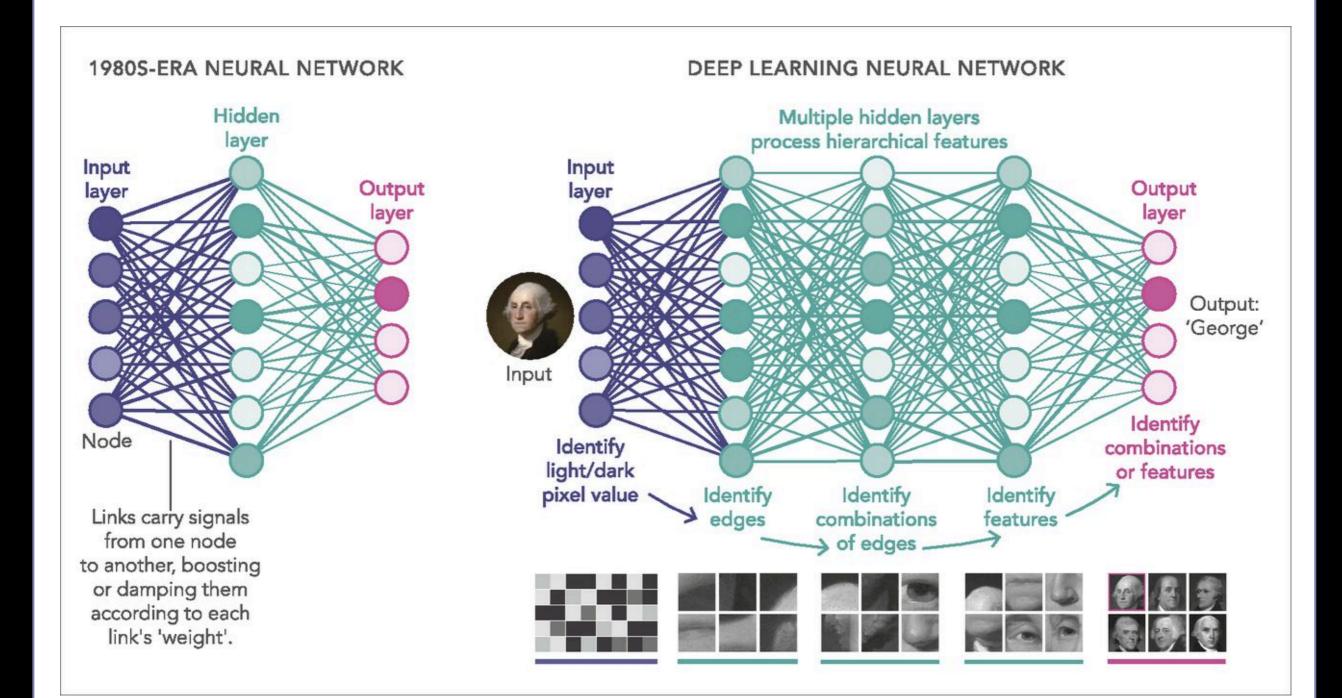
ImageNet Object Recognition Challenge



leopard	motor scooter	container ship	mite
leopard	motor scooter	container ship	mite
jaguar	go-kart	lifeboat	black widow
cheetah	moped	amphibian	cockroach
snow leopard	bumper car	fireboat	tick
Egyptian cat	golfcart	drilling platform	starfish



From Blackbox Models to Dark Magic?



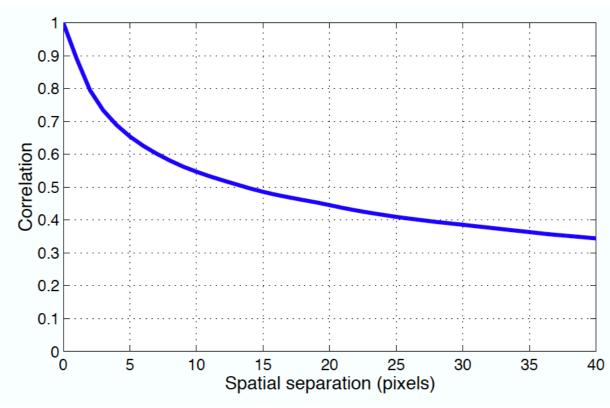
Task-Specific Priors

Task:





Priors:



Locally-Coded Features

Task-Specific Priors

Task:





Priors:

Spatially Invariant





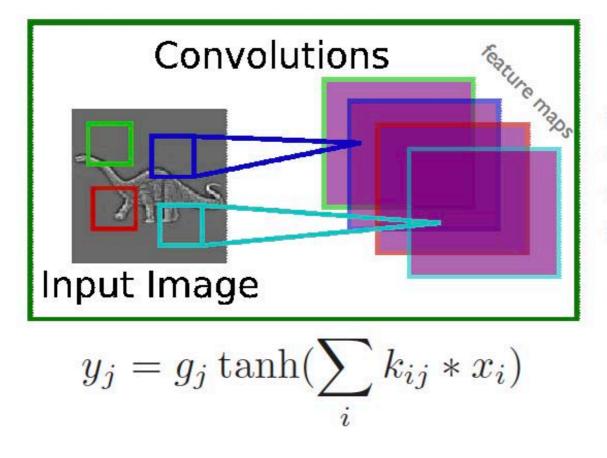
Scale Invariant





Ideas: Convolutional Layer

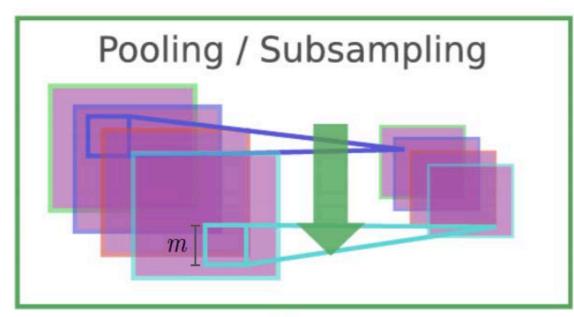
- Filtering with a kernel in small neighborhoods: local processing
- Parameter sharing: translation invariance



- x_i is the ith channel of input
- k_{ij} is the convolution kernel
- g_j is a learned scaling factor
- y_j is the hidden layer

Ideas: Pooling Layer

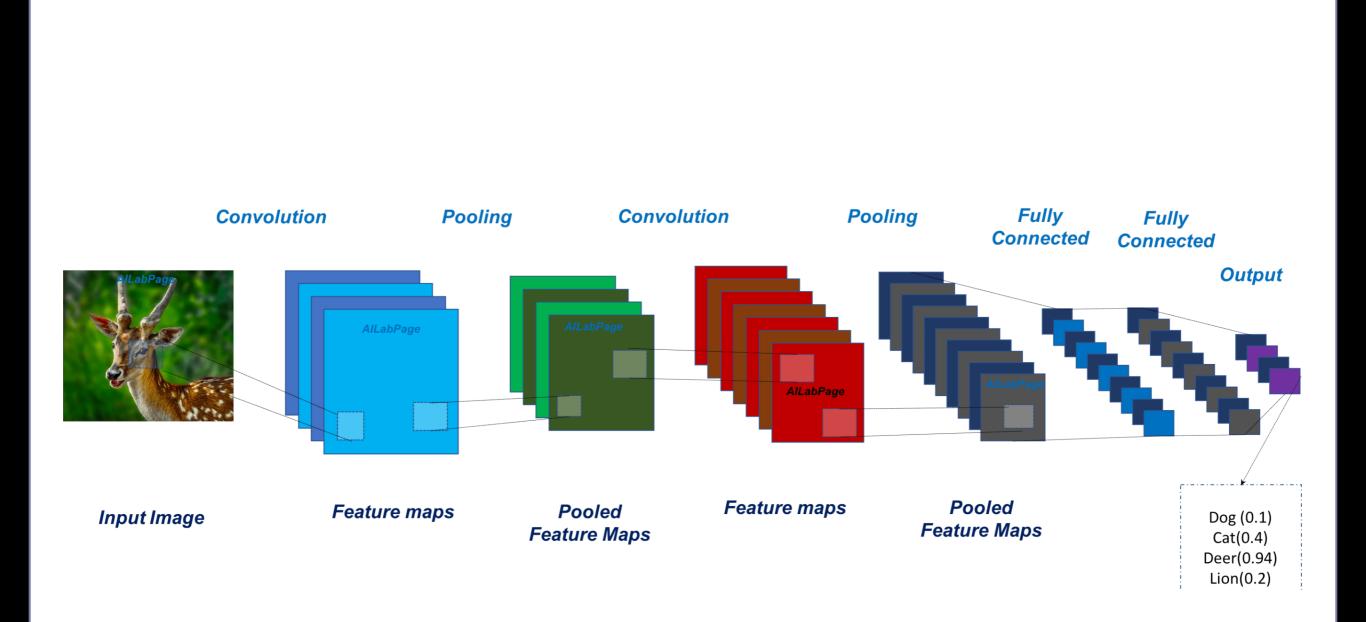
- Pooling hidden-unit responses: translation invariance
- Analysis at multiple-spatial scales: scale invariance



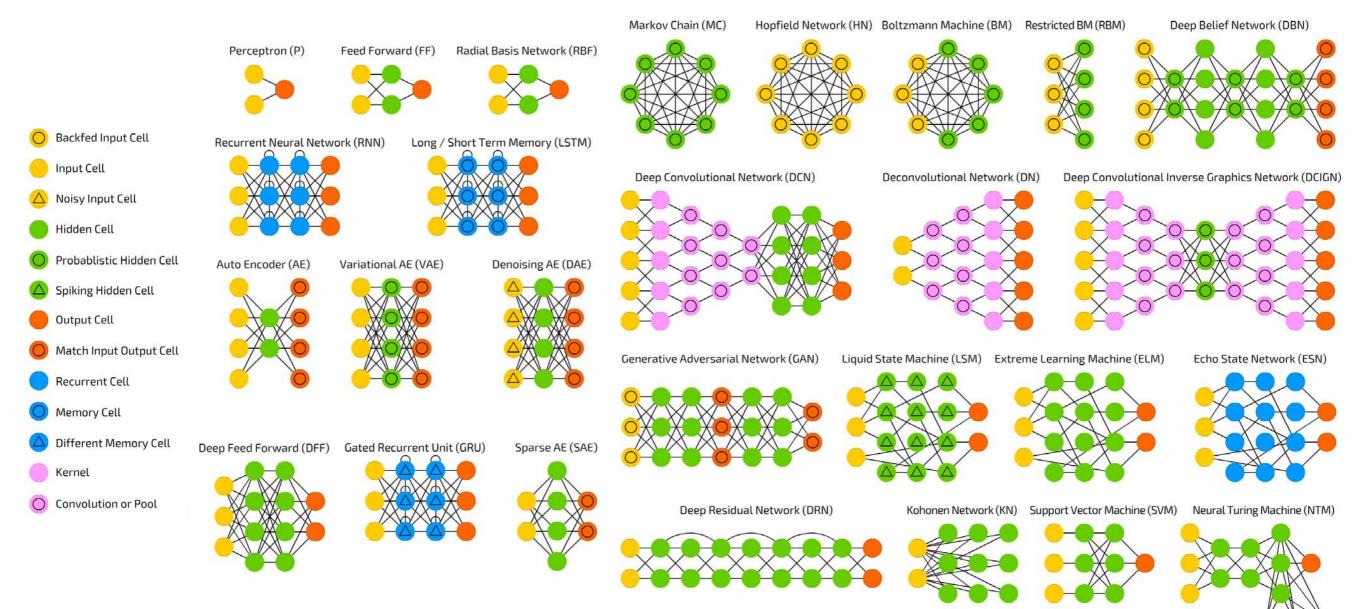
$$y_{ijk} = \frac{1}{m^2} \sum_{p,q} x_{i,j+p,k+q}$$

- x_{i,j,k} is value of the ith feature map at position j,k
- p is vertical index in local neighborhood
- q is horizontal index in local neighborhood
- y_{ijk} is pooled and subsampled layer
- *m* is the neighborhood height/width

Convolutional Neural Networks



Network Engineering



PART II: History of ML in Medicine

Expert Systems

1970's: MYCIN expert system

- 1970's (Stanford): MYCIN expert system for identifying bacteria causing severe infections
- Proposed a good therapy in ~69% of cases. Better than infectious disease experts

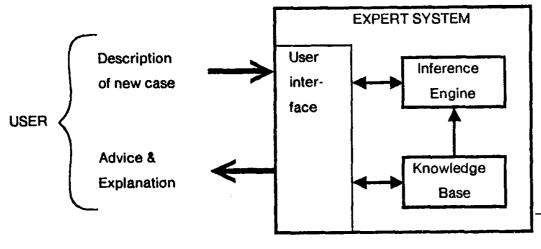


FIGURE 1-1 Major parts of an expert system. Arrows indicate information flow.

Dialogue interface

I am ready

** THIS IS A 26 YEAR OLD MALE PATIENT

My understanding is: The age of the patient is 26 The sex of the patient is male

** FIVE DAYS AGO, HE HAD RESPIRATORY-TRACT SYMPTOMS

What is his name?

** JO

My understanding is: The name of the patient is Jo Respiratory-tract is one of the symptoms that the patient had

** A COUPLE OF DAYS BEFORE THE ADMISSION, HE HAD A MALAISE

Please give me the date of admission

** MARCH 12, 1979

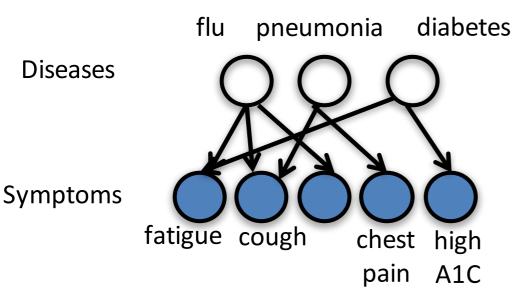
My understanding is: The patient was admitted at the hospital 3 days ago Malaise is one of the symptoms that the patient had 5 days ago

FIGURE 33-1 Short sample dialogue. The physician's inputs appear in capital letters after the double asterisks.

QMR Models

1980's: INTERNIST-1/QMR model

- 1980's (Univ. of Pittsburgh): INTERNIST-1/Quick Medical Reference
- Diagnosis for internal medicine



Probabilistic model relating:

570 binary disease variables4,075 binary symptom variables45,470 directed edges

Elicited from doctors: **15 person-years of work**

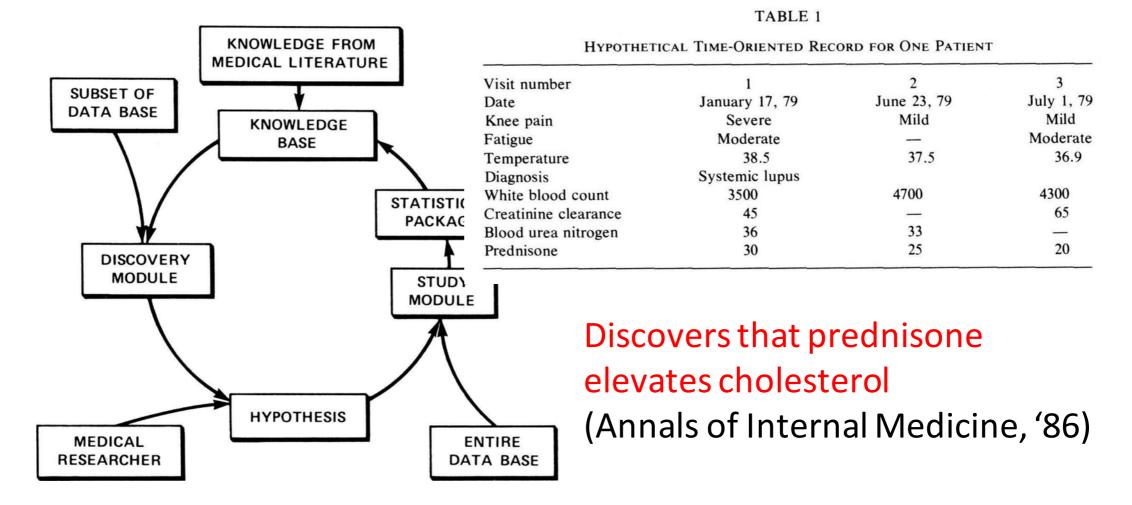
Led to advances in ML & AI (Bayesian networks, approximate inference)

Problems: 1. Clinicians entered symptoms *manually*2. Difficult to maintain, difficult to generalize

Drug Discovery

1980's: automating medical discovery

RX PROJECT: AUTOMATED KNOWLEDGE ACQUISITION



1990's: neural networks in medicine

- Neural networks with clinical data took off in 1990, with 88 new studies that year
- Small number of features (inputs)
- Data often collected by chart review

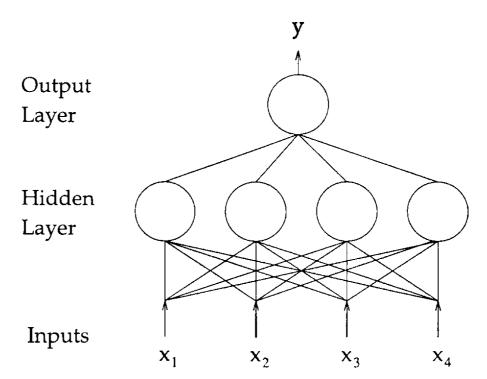


FIGURE 2. A multilayer perceptron. This is a two-layer perceptron with four inputs, four hidden units, and one output unit.

Problems: 1. Did not fit well into clinical workflow2. Poor generalization to new places

Disease Diagnosis

	No. of Ex	amples				Accu	racy§
Subject	Training	Test	P†	Network	D‡	Neural	Other
Breast cancer ⁴	57	20	60	9-15-2	0.6	80	75
Vasculitis ²	404	403	73	8-5-1	8.0	94	
Myocardial infarction ⁶	351	331	89	20-10 -10-1	1.1	97	84
Myocardial infarction ⁸	356	350	87	20-10-10-1	1.1	97	94
Low back pain ¹¹	100	100	25	50-48-2	0.2	90	90
Cancer outcome ¹³	5,169	3,102	_	54-40-1	1.4	0.779	0.776
Psychiatric length of stay ¹⁷	957	106	73	48-400-4	0.2	74	76
Intensive care outcome ²³	284	138	91	27-18-1	0.5	0.82	0.82
Skin tumor ²¹	150	100	80	18	<u> </u>	80	90
Evoked potentials ³⁵	100	67	52	14-4-3	3.8	77	77
Head injury ⁴⁷	500	500	50	6-3-3	20	66	77
Psychiatric outcome ⁵⁴	289	9 2	60	41-10-1	0.7	79	
Tumor classification55	53	6	38	8-9-3	1.4	99	88
Dementia ⁵⁷	75	18	19	80-10-7-7	0.6	61	
Pulmonary embolism59	607	606	69	50-4-1	2.9	0.82	0.83
Heart disease62	460	230	54	35-16-8-2	3	83	84
Thyroid function ⁶²	3,600	1,800	93	21-16-8-3	22	98	93
Breast cancer ⁶²	350	175	66	9-4-4-2	10	97	96
Diabetes ⁶²	384	192	65	8-4-4-2	12	77	75
Mycardial infarction63	2,856	1,429	56	291-1	9.8	85	
Hepatitis ⁶⁵	39	42	38	4-4-3	3.3	74	79
Psychiatric admission ⁷⁶	319	339	85	53-1-1	6.0	91	
Cardiac length of stay ⁸³	713	696	73	15-12-1	3.5	0.70	
Anti-cancer agents ⁸⁹	127	141	25	60-7-6	1.5	91	86
Ovarian cancer ⁹¹	75	98	_	6-6-2	2.6	84	81
Median value	350	175	71	20	2.8		

Table 1 • 25 Neural Network Studies in Medical Decision Making*

*For reference citations, see the reference list

†P = prior probability of most prevalent category.

‡D = ratio of training examples to weights per output.

§A single integer in the accuracy column denotes percentage overall classification rate and a single real number between 0 and 1 indicates the AUROCC value. Neural = accuracy of neural net, Other = accuracy of best other method.

PART III: Big Data in Medical Applications

Sources of Medical Data



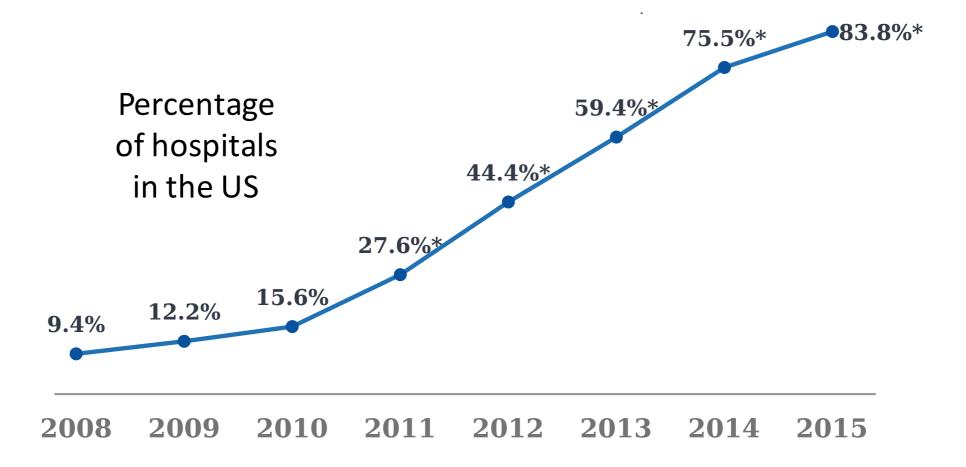
Sources of Medical Data

Diversity of digital health data



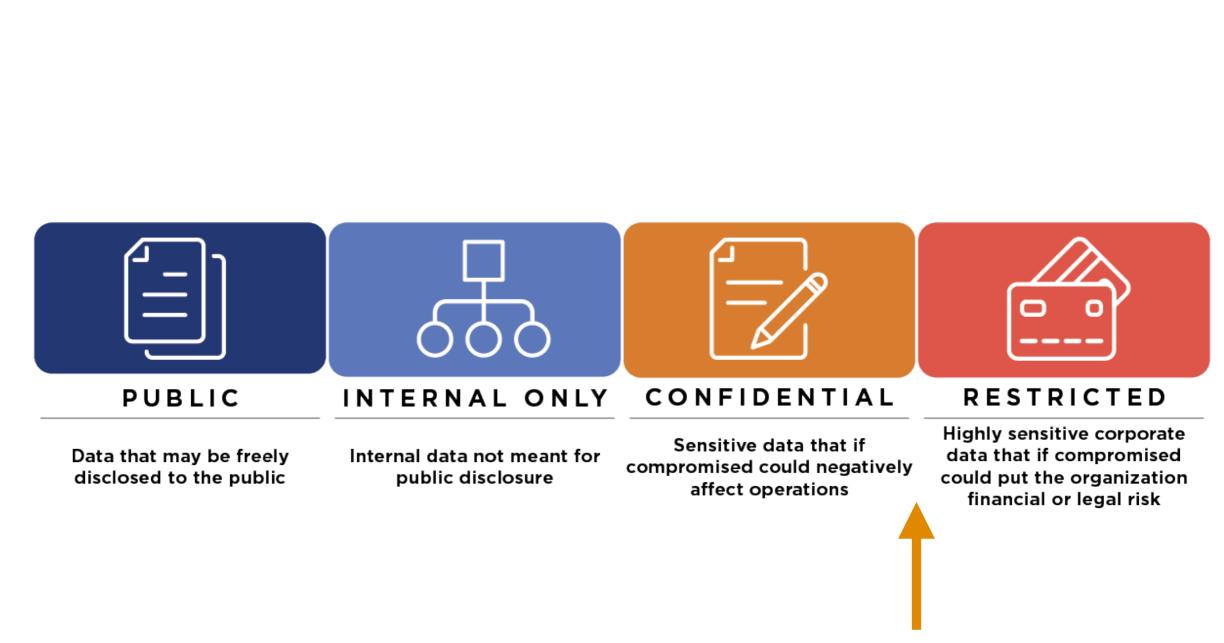
Availability of Medical Data

Adoption of Electronic Health Records (EHR) has increased 9x since 2008



[Henry et al., ONC Data Brief, May 2016]

Standard Classes of Data





Large, Public Databases are Emerging



UK Biobank is a national and international health resource with unparalleled rest aims to improve the prevention, diagnosis and treatment of a wide range of seric diabetes, arthritis, osteoporosis, eye disorders, depression and forms of dement and provides health information, which does not identify them, to approved reset please ensure you read the <u>background materials</u> before registering. To our part health. Without you, none of the research featured on this website would



Laboratory for Computational Physiology

De-identified health data from ~40K critical care patients

Demographics, vital signs, laboratory tests,

Read more about Biobank UK



UK BIOBANK MAKES INFECTION AND HEALTH DATA AVAILABI TO TACKLE COVID-19

THE PRECISION MEDICINE INITIATIVE



- Baseline health exam
- Clinical data derived from electronic health records (EHRs)
- Healthcare claims
- Laboratory data

PART IV: Opportunities/Challenges in Healthcare

ML/DL in Biomedical Domain

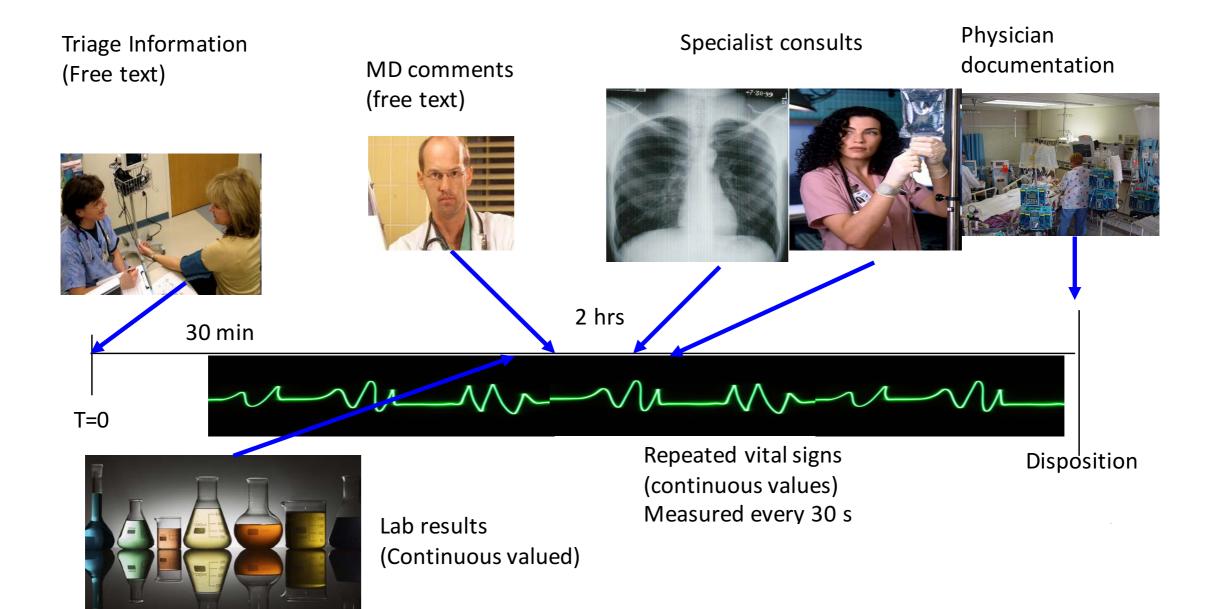


Example: Emergency Departments



- Limited resources
- Time sensitive
- Critical decisions

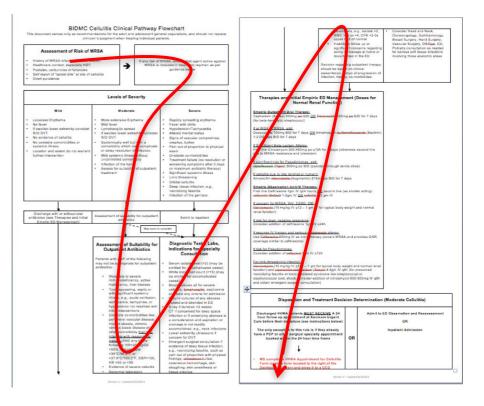
Example: Emergency Departments



How can Machine Learning Help?

- Triggering clinical pathways
- Context-specific displays
- Risk stratification
- Improving clinical documentation

BIDMC Cellulitis Clinical Pathway Flowchart



Pathways have been shown to reduce in-hospital complications without increasing costs [Rotter et al 2010]

Automatic Protocol Selection

Triggering clinical pathways

- Context-specific displays
- Risk stratification
- Improving clinical documentation

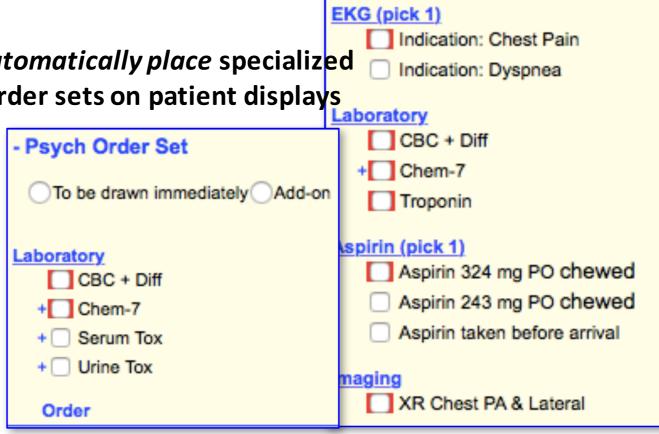
<u>Our task:</u> Determine whether a patient has or is suspected to have cellulitis Automating triggers Don't rely on the user's knowledge that the pathway exists!

		Enroll in pathway		
		Decline		
Van		and for the second		
You	an include a comn	nent for the revie	vers: Mandator	y if Declining

Disease-specific Recommender Systems

- Triggering clinical pathways
- Context-specific displays
- Risk stratification
- Improving clinical documentation Automatically place specialized Indication: Chest Pain order sets on patient displays

<u>Our task:</u> Determine whether patient complained of chest pain, or is a psych patient



Initial

Chest Pain Order Set

Place IV (saline lock); flush per protocol

To be drawn immediately Add-on

Continuous Cardiac monitoring

Minimizing Risk

- Triggering clinical pathways
- Context-specific displays
- Risk stratification
- Improving clinical documentation

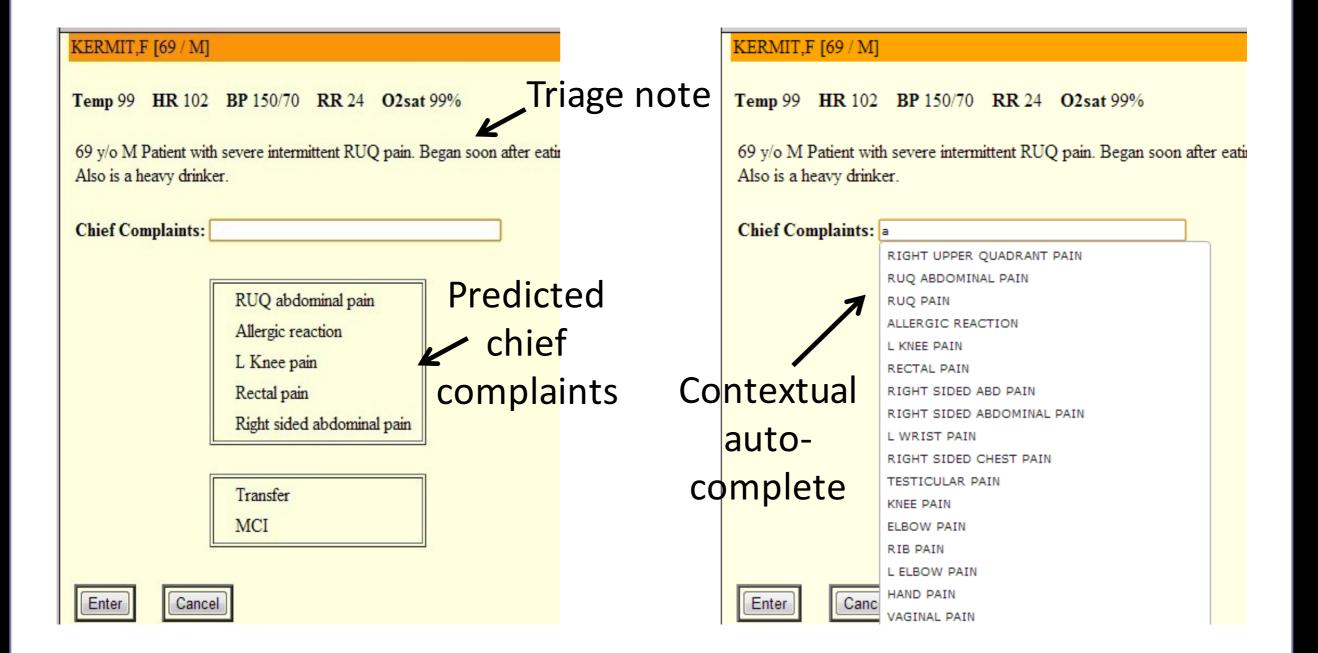
Ex 1: Likelihood of mortality or admission to ICU

Ex 2: Early detection of severe sepsis

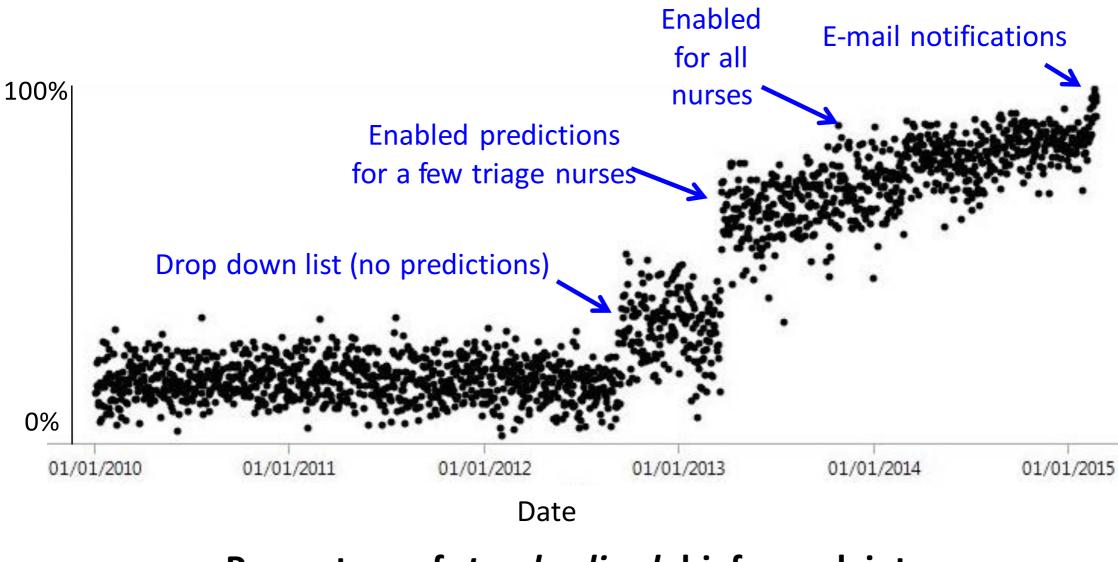
Real-time Disease Prediction

History	Acute	Deep vein thrombosis	Laceration
Alcoholism	Abdominal pain	Employee exposure	Motor vehicle accident
Anticoagulated	Allergic reaction	Epistaxis	Pancreatitis
Asthma/COPD	Ankle fracture	Gastroenteritis	Pneumonia
Cancer	Back pain	Gastrointestinal bleed	Psych
Congestive heart	Bicycle accident	Geriatric fall	Obstruction
failure	Cardiac etiology	Headache	Septic shock
Diabetes	Cellulitis	Hematuria	Severe sepsis
HIV+	Chest pain	Intracerebral	Sexual assault
Immunosuppressed	Cholecystitis	hemorrhage	Suicidal ideation
Liver malfunction	Cerebrovascular	Infection	Syncope
	accident	Kidney stone	Urinary tract infection

Improving Clinical Documentation

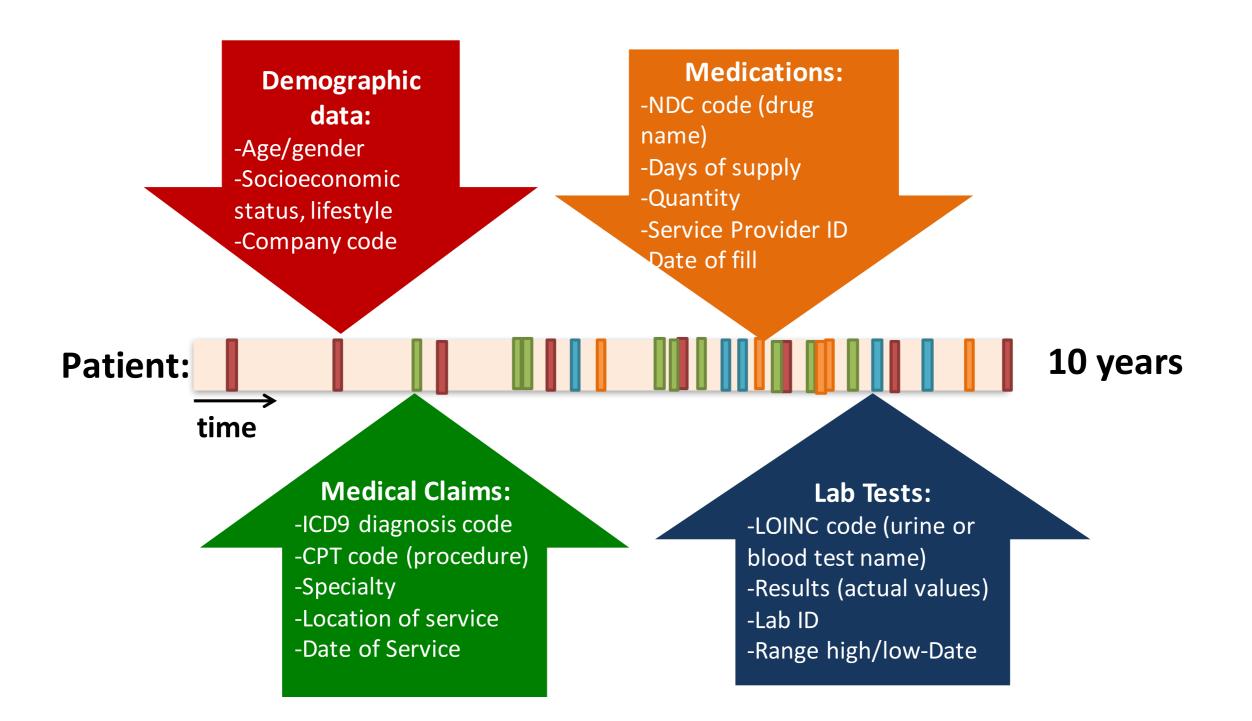


Improving Clinical Documentation



Percentage of *standardized* chief complaints (per week)

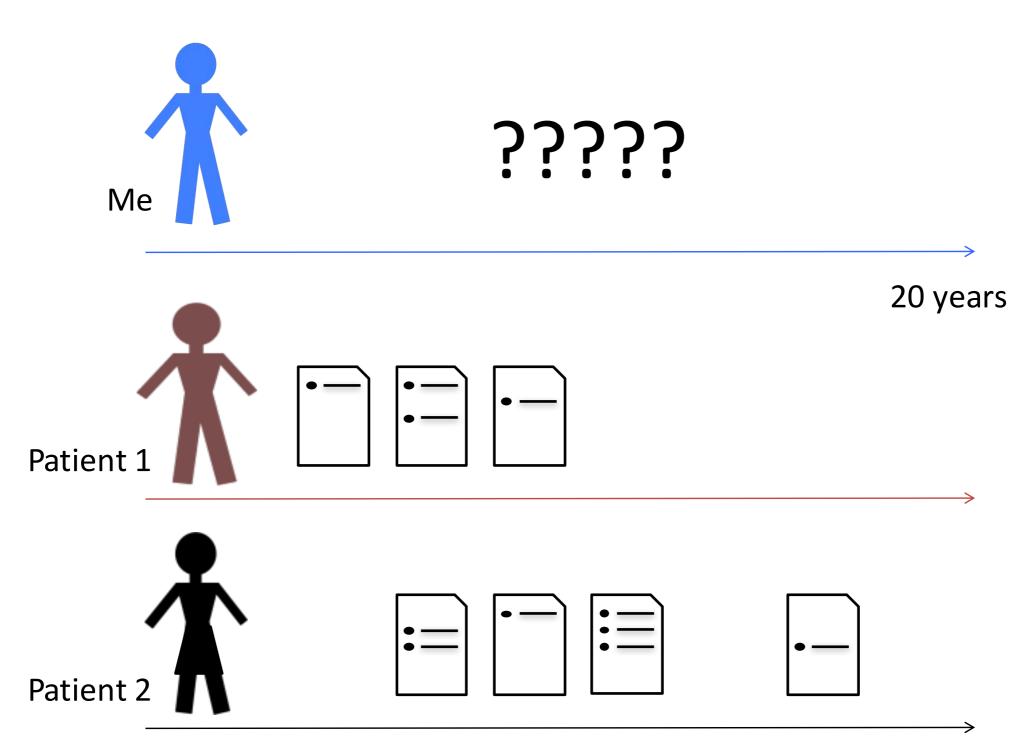
At a Broader Time Scale...



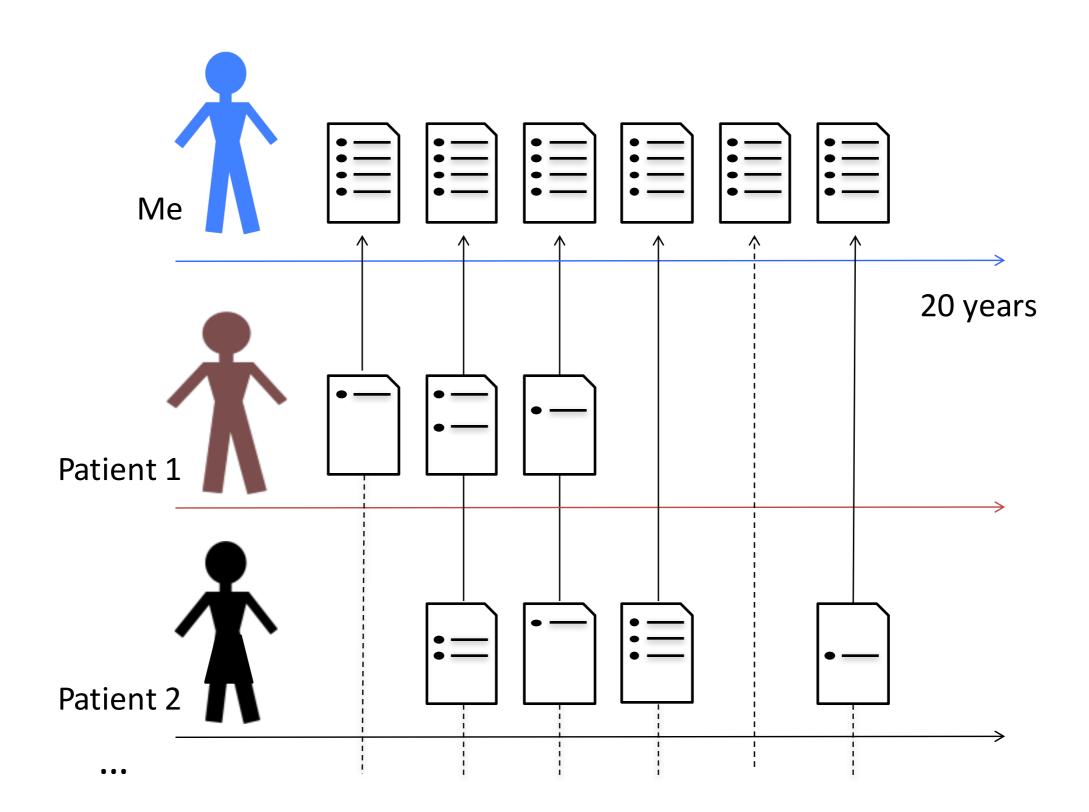
Temporal Modeling of Disease Progression

- Find markers of disease stage and progression, statistics of what to expect when
 - What is the "typical trajectory" of a female diagnosed with Sjögren's syndrome at the age of 19?
- Estimate a patient's future disease progression
 - When will a specific individual with smoldering multiple myeloma (a rare blood cancer) transition to full-blown multiple myeloma?
 - Which second-line diabetes treatment should we give to a patient?

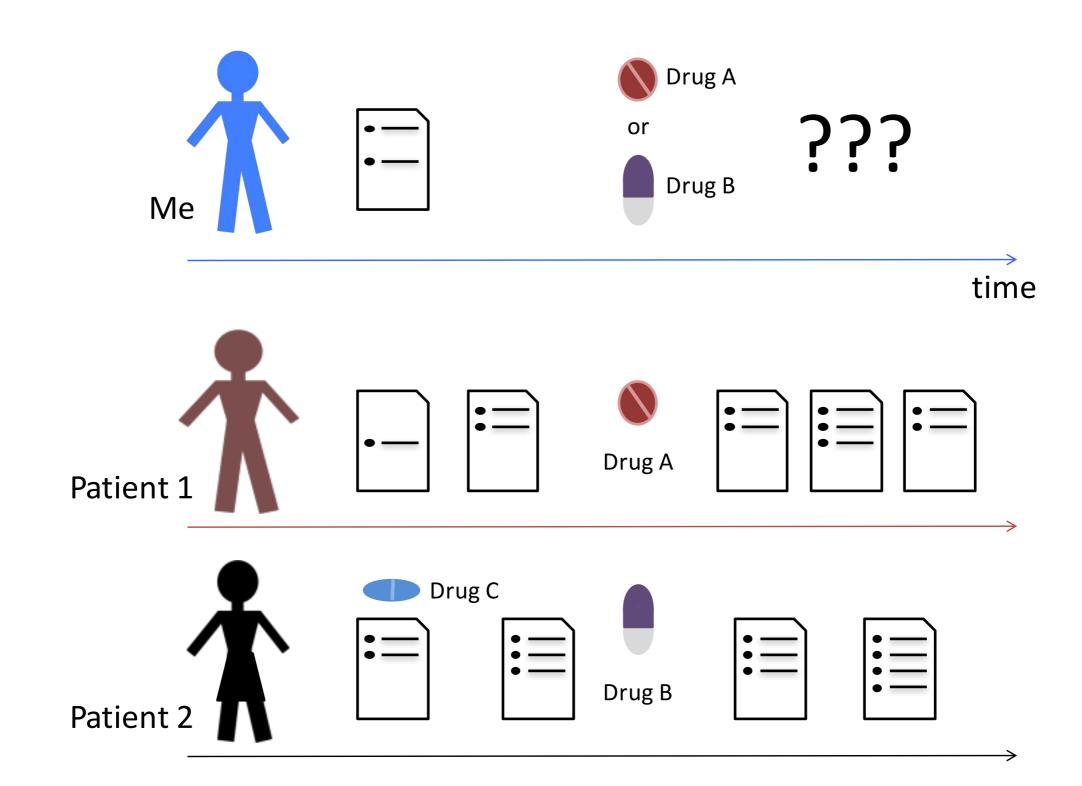
Personalized Medicine



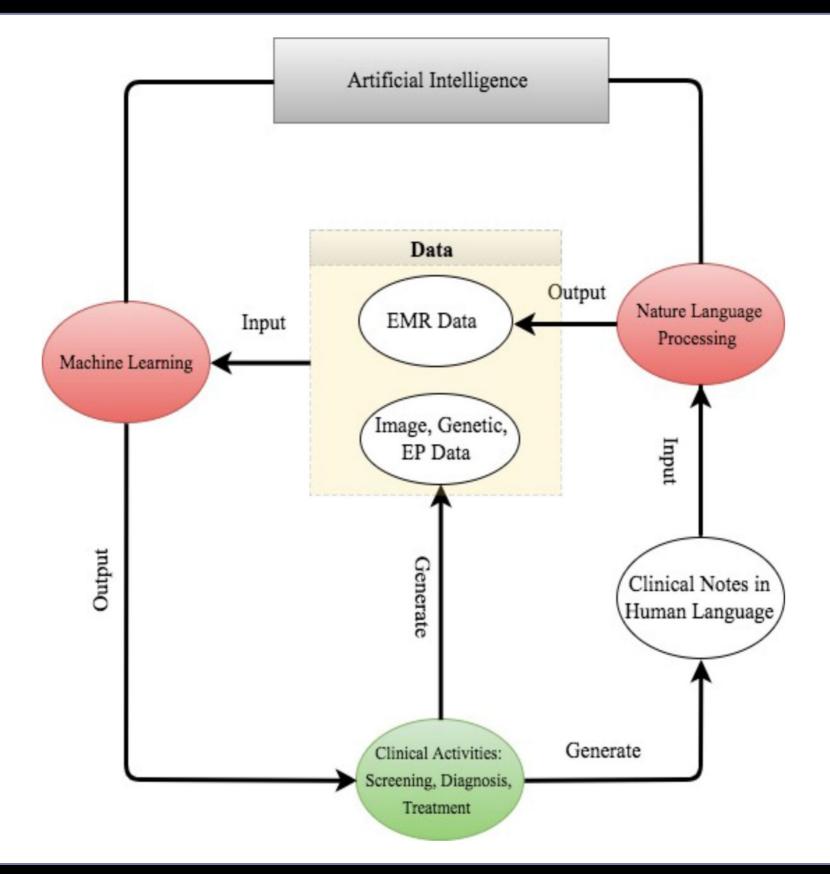
Prediction of Health Status



Personalized Prescriptions



From Data Generation to Decision Making



Many Challenges Unique to Medicine

- Life or death decisions
 - Need **robust** algorithms
 - Checks and balances built into ML deployment
 - (Also arises in other applications of AI such as autonomous driving)
 - Need fair and accountable algorithms
- Many questions are about unsupervised learning
 - Discovering disease subtypes, or answering question such as "characterize the types of people that are highly likely to be readmitted to the hospital"?
- Many of the questions we want to answer are *causal*
 - Naïve use of supervised machine learning is insufficient

Problems with "Data"

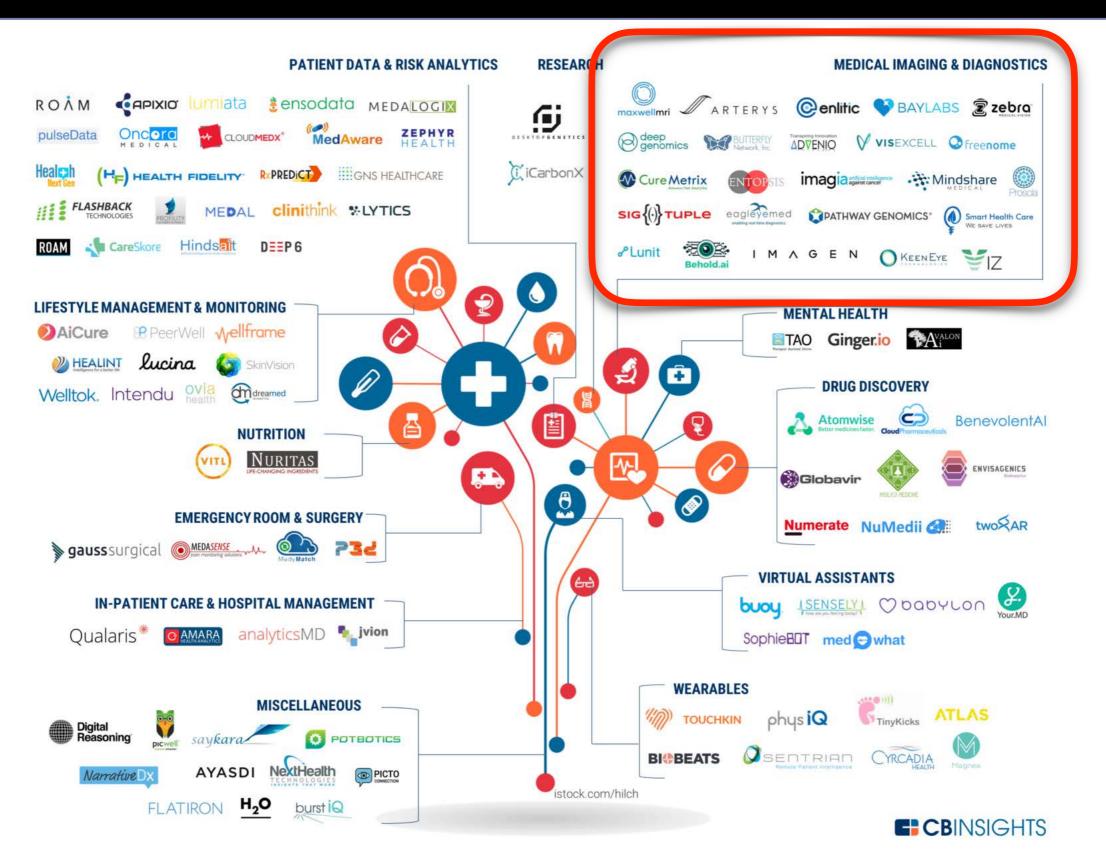
- Often very little labeled data (e.g., for clinical NLP)
 - Motivates semi-supervised learning algorithms
- Sometimes small numbers of samples (e.g., a rare disease)
 - Learn as much as possible from other data (e.g. healthy patients)
 - Model the problem carefully
- Lots of missing data, varying time intervals, censored labels

Problems with Clinical Integration

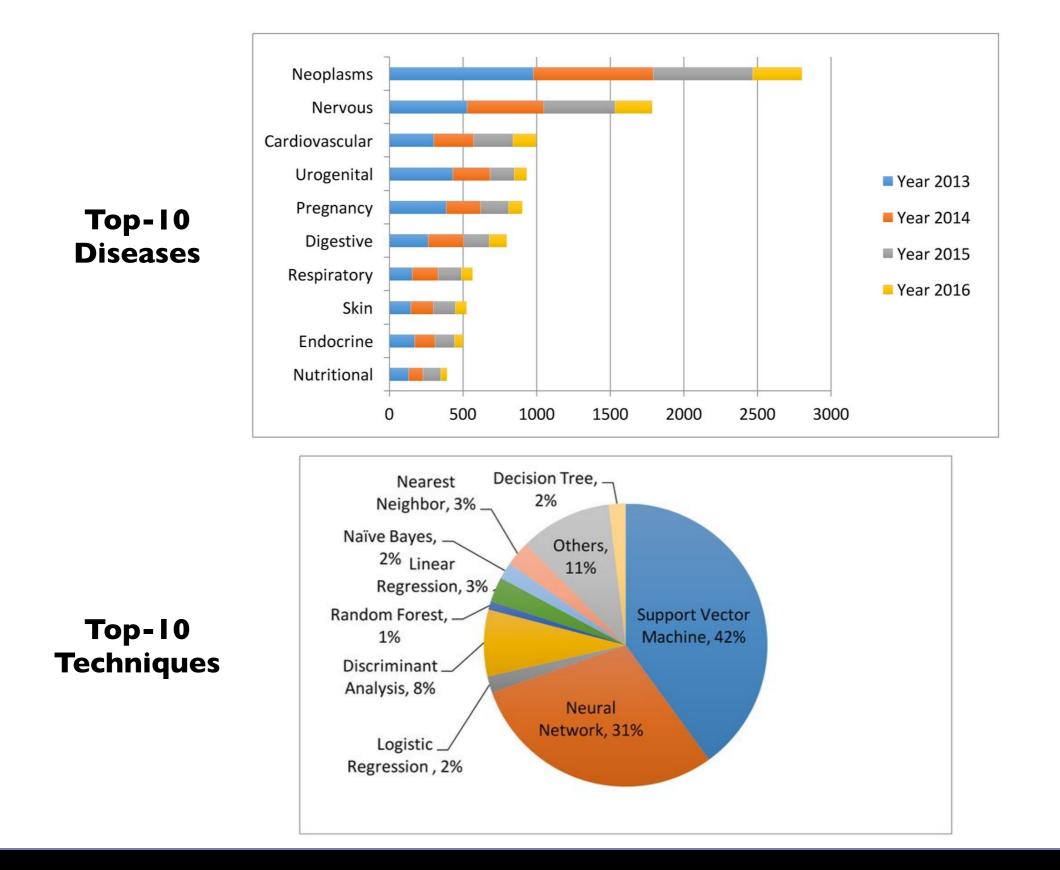
- Difficulty of de-identifying data
 - Need for data sharing agreements and sensitivity
- Difficulty of deploying ML
 - Commercial electronic health record software is difficult to modify
 - Data is often in silos; everyone recognizes need for interoperability, but slow progress
 - Careful testing and iteration is needed

PART V: Utility of ML in Medical Imaging

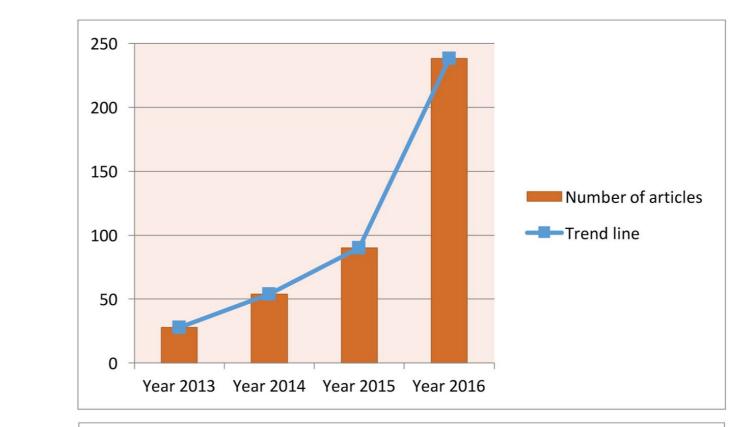
Medical Uses of Deep Learning



Machine Learning for Diagnosis

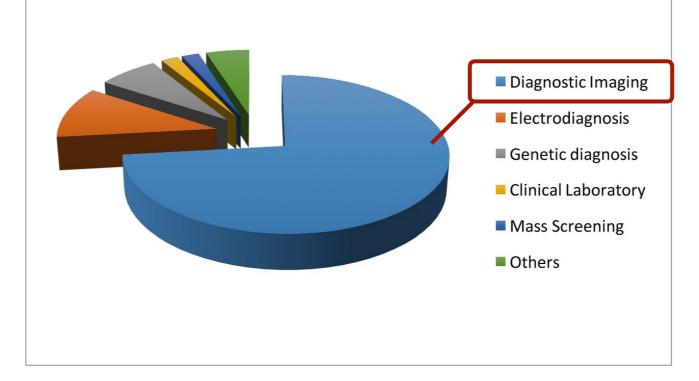


Deep Learning on the Rise

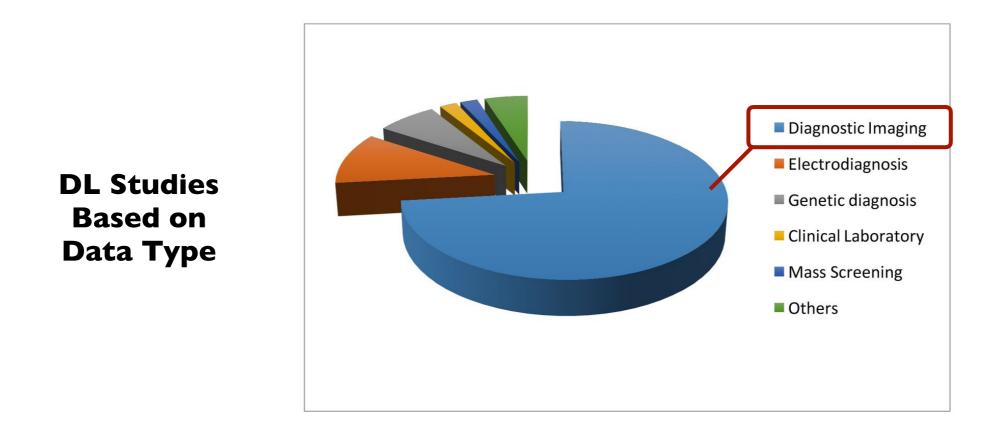


Number of **DL Studies**



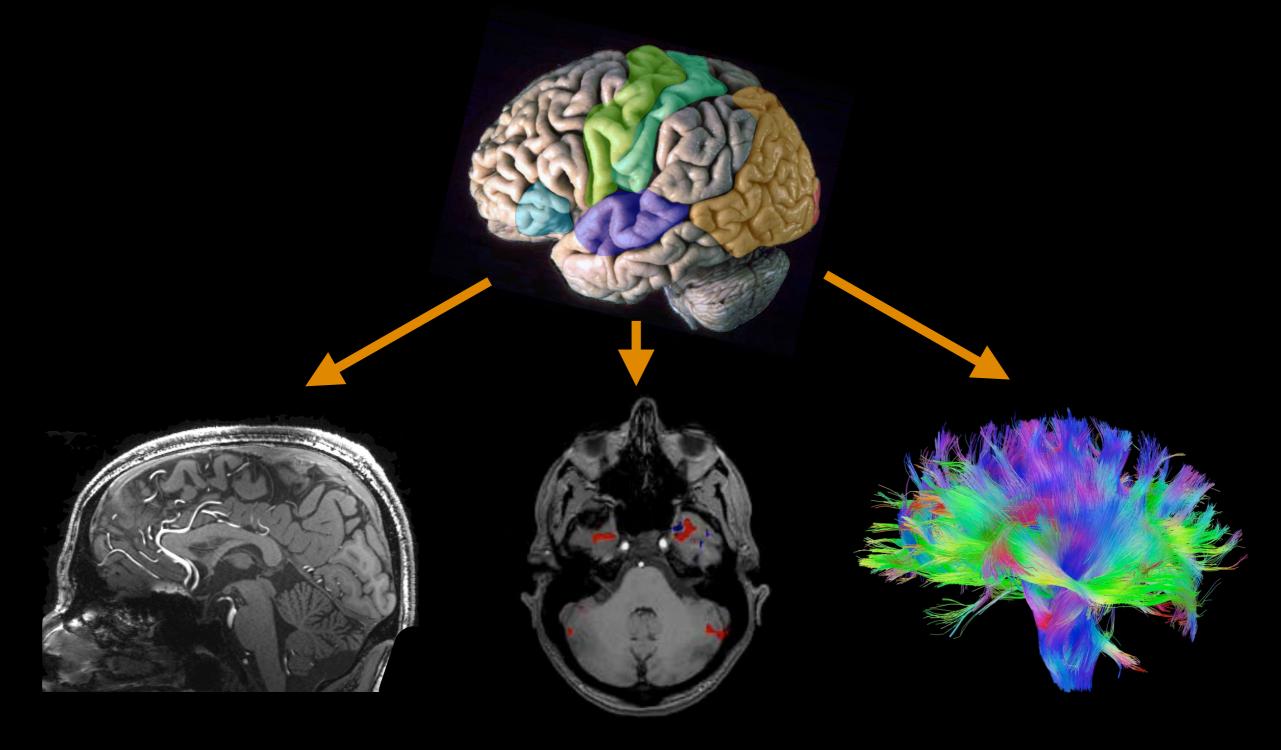


Deep Learning for Medical Imaging



- Medical images are high-dimensional (volumetric and temporal)
- Medical images are mostly interpreted by radiologists (manual labor)
- Humans are quite poor in seeing fine-grained patterns in static images
- Similarities medical-natural images (closely tied to computer vision)

Imaging Morphology and Function



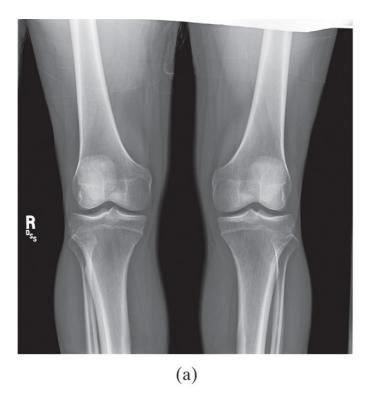
Anatomical

Functional

Diffusion

Modern Imaging Modalities

X-ray





Nuclear Medicine

MRI

(b)

Ultrasound

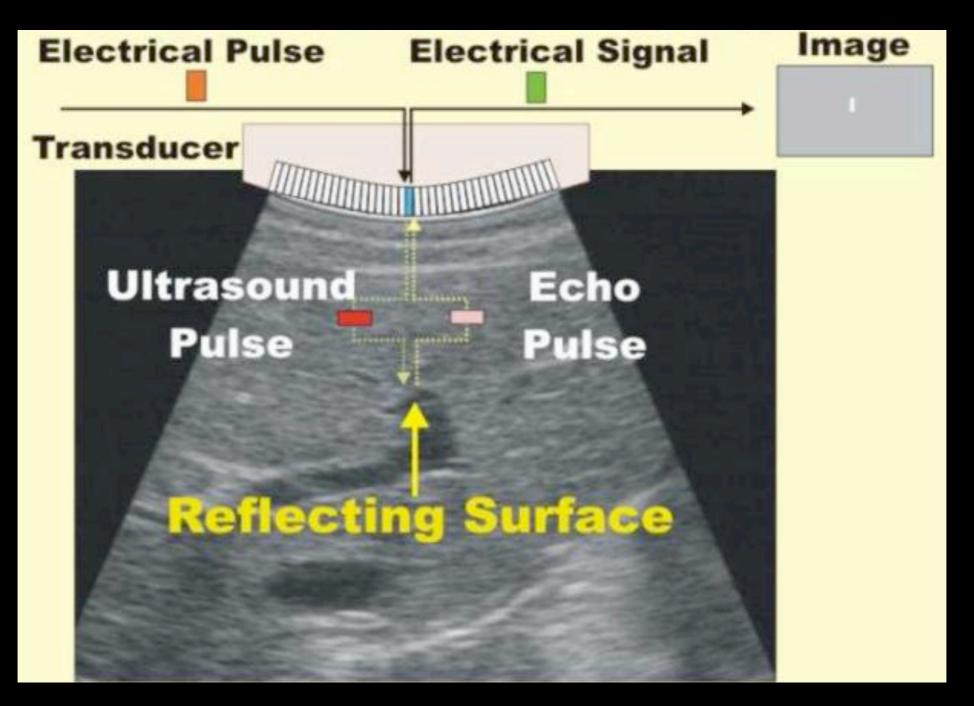




(d)

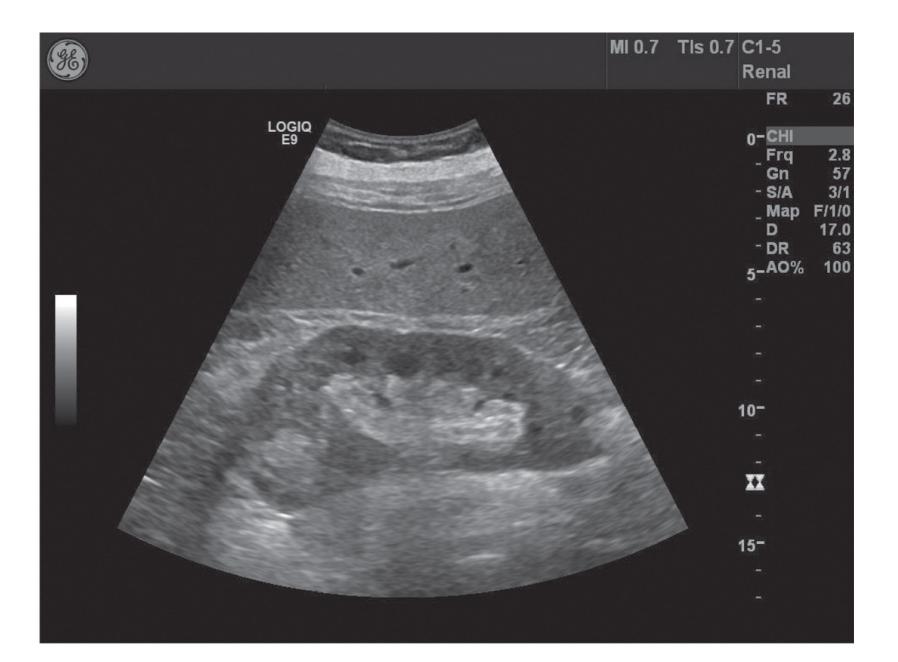
Ultrasound

- Uses sound pressure waves
- We cannot hear these sounds: > 20 kHz
- Typically 2-18 MHz



Ultrasound



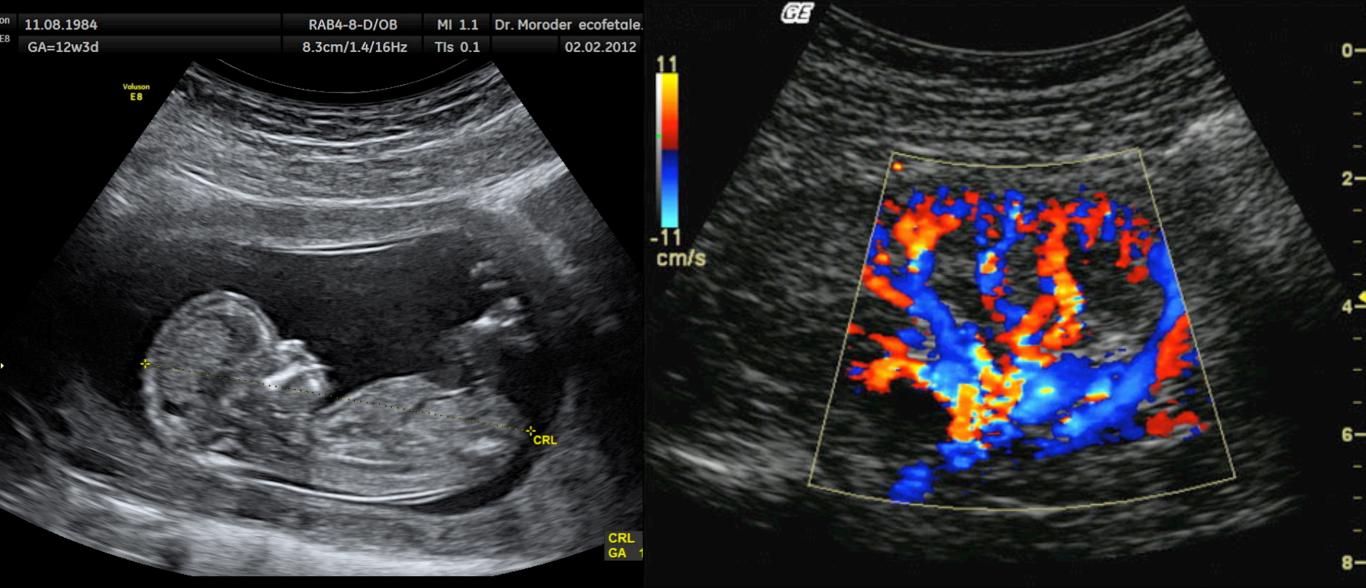


(a)

(b)

Ultrasound

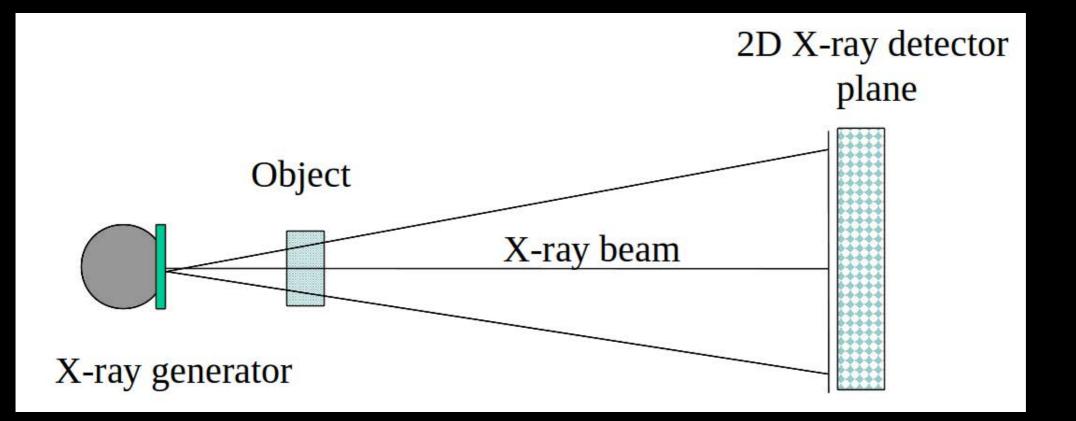
- Higher resolution with increasing frequency
- Cannot image too deep if the frequency is high
- Lots of application areas: cardiology, urology, obstetrics,...



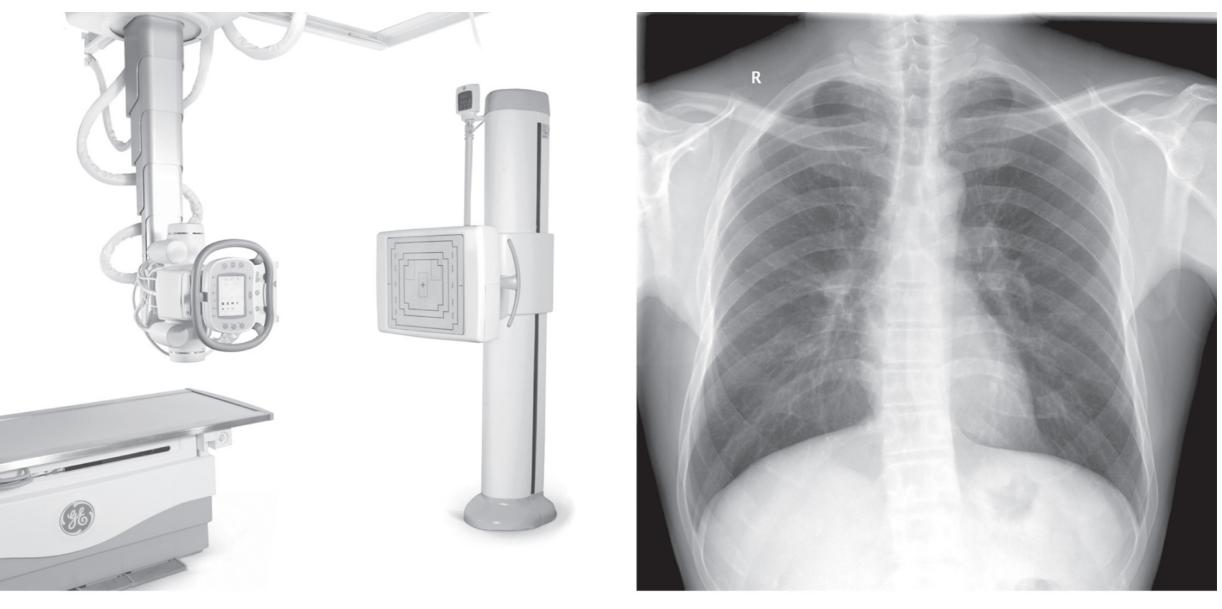
Doppler Ultrasound

X-ray

- Uses X-ray photons
- Photons at VERY high frequency: ~10¹⁶-10¹⁹ Hz
- Different tissues attenuate photons differently \rightarrow contrast
- Very good at detecting bone structure
- Projection images



Chest X-ray



(a)

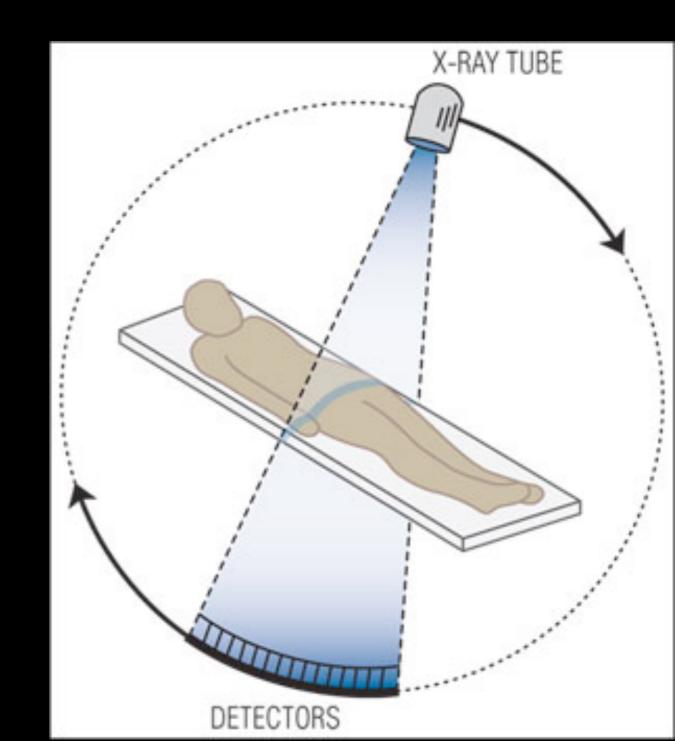
(b)

Projection Through the Body

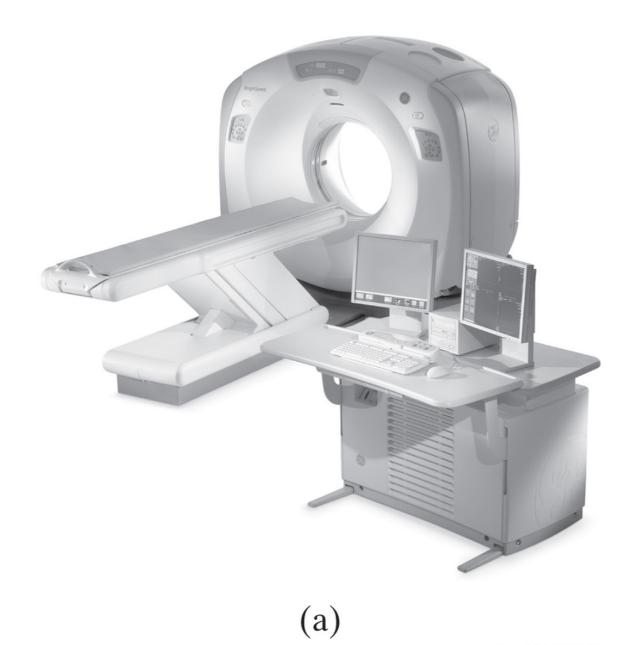


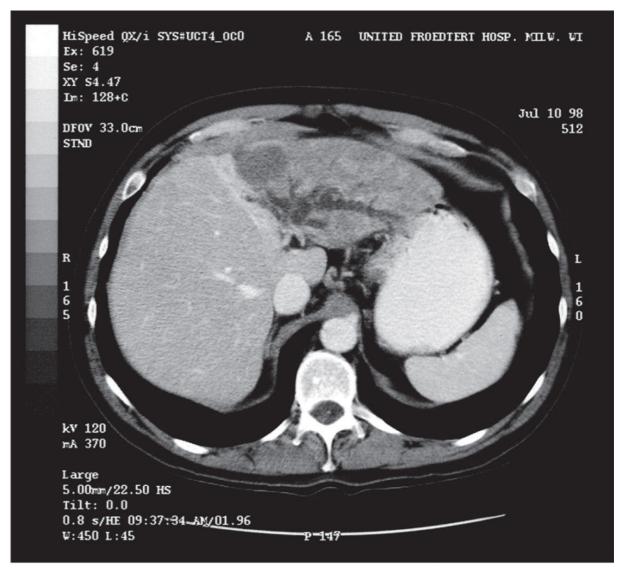
Computerized Tomography

- Extends X-ray imaging to 3D format
- Rotate X-ray source and detectors all together



Computerized Tomography

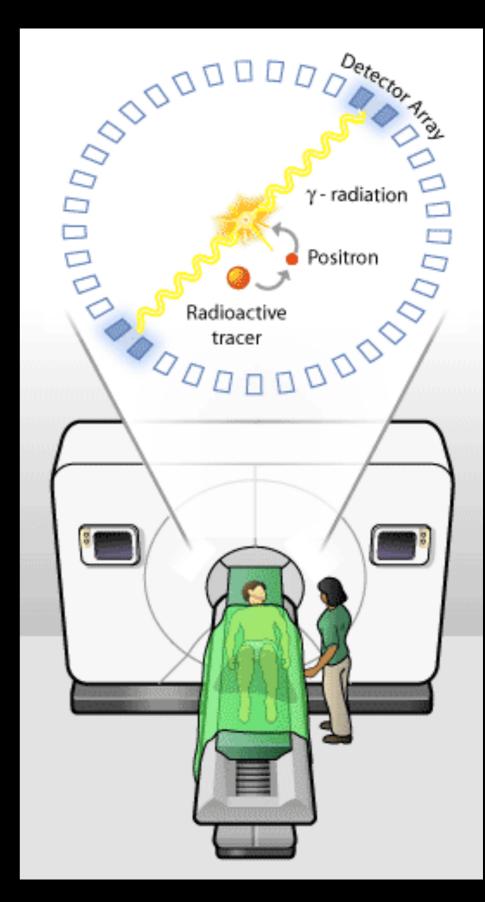




(b)

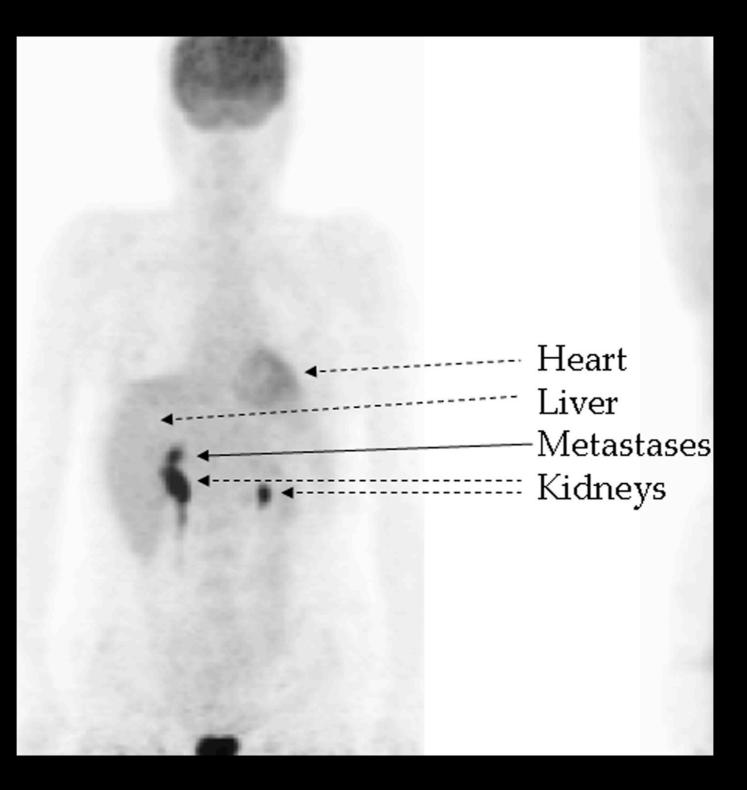
Nuclear Medicine

- A radionuclide is injected into the blood stream
- Goes to cancer locations before cancer cells are hyperactive
- Emitted Gamma rays are detected for imaging



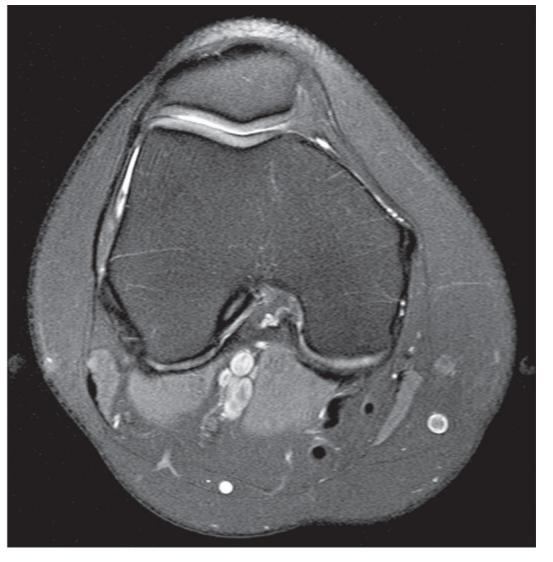
Positron Emission Tomography

Resolution is not good, but very sensitive to cancer



MRI



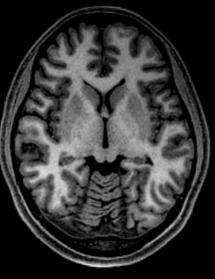


(b)

(a)

Magnetic Resonance Imaging (MRI)

- FLEXIBLE CONTRAST
- Arbitrary geometries
- Non-invasive, non-ionizing imaging
- Relatively slow imaging

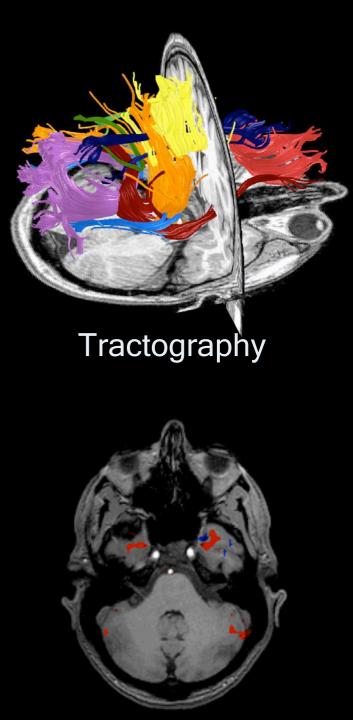


Anatomy

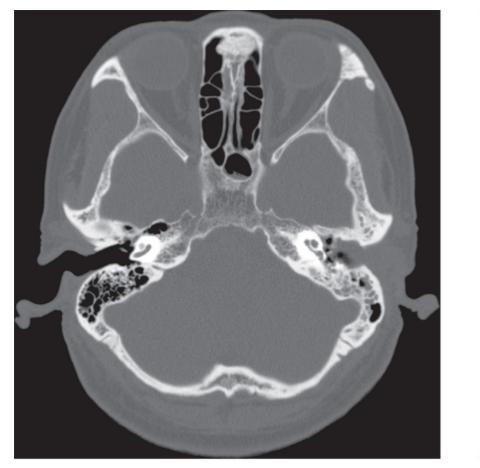


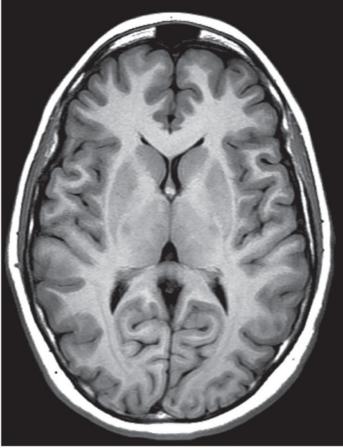


Angiography



functional MRI







(a) CT (b)

MRI



Medical Imaging Pipeline

 Acquisition

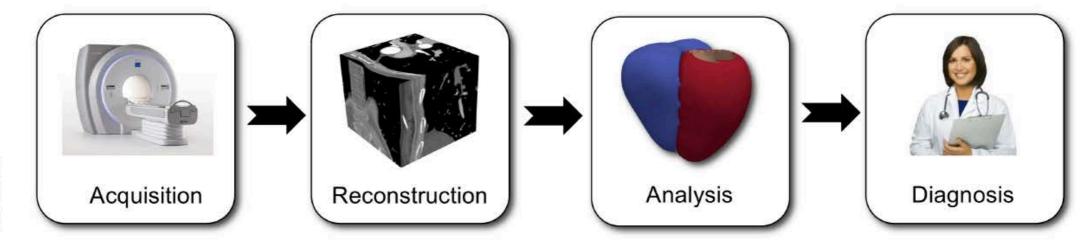
End-to-end integrated medical imaging pipeline

Current

Future

Motivation

Serial medical imaging pipeline

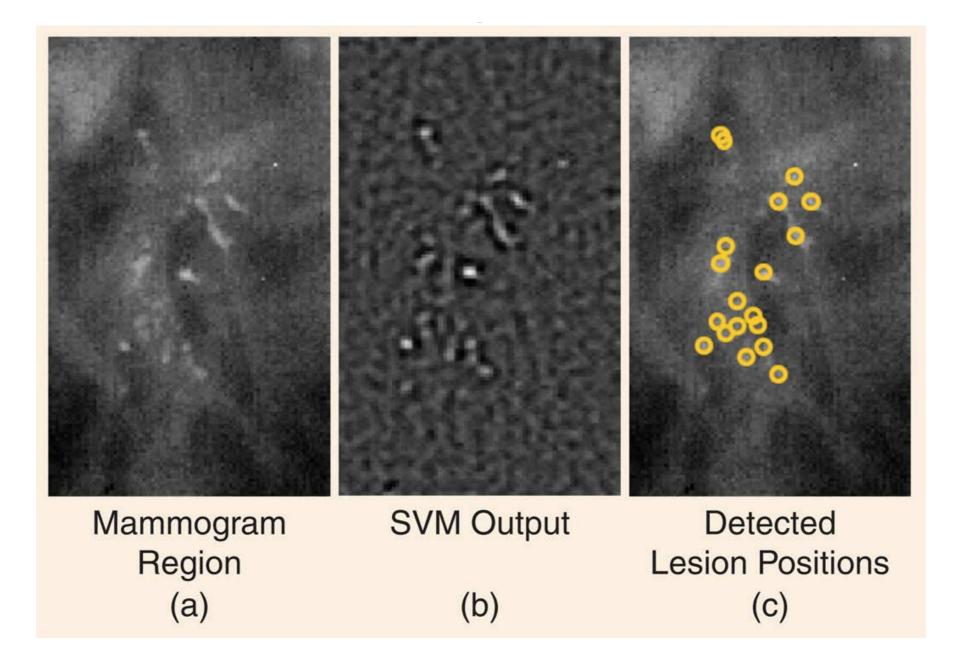


Current

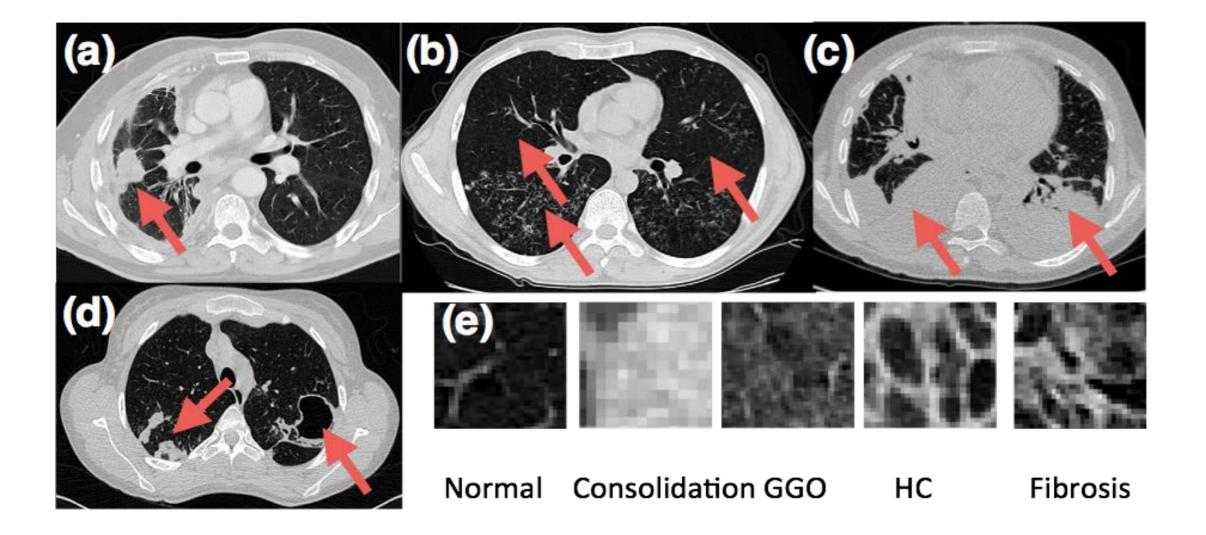
- Radiologists need to interpret an excessively large number of images
- Their capacity to correctly interpret images is overwhelmed
- Automated image analysis systems are needed for error reduction
- Machine learning underpins the algorithms for such systems

PART V: Example Applications in Medical Imaging

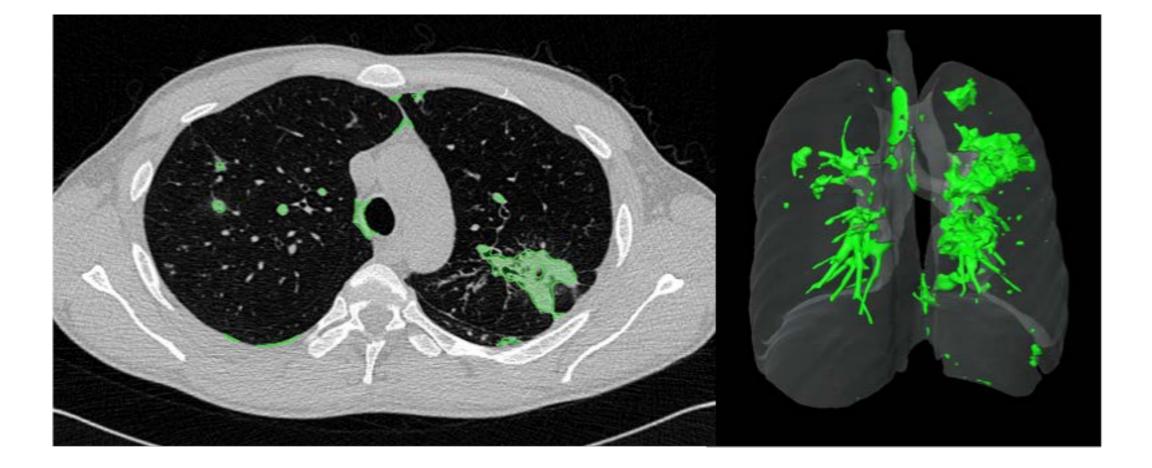
Examples: Detecting Micro-calcifications

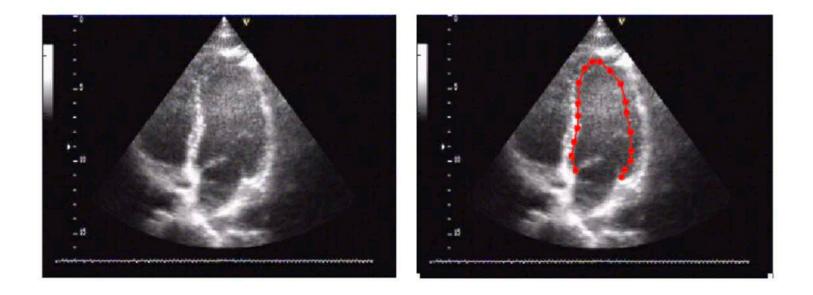


Examples: Detecting Pulmonary Abnormalities



Examples: Detecting Pulmonary Abnormalities





- Goals
 - Automated functional analysis of the heart
 - Improve workflow, reduce user variability
- Challenges
 - Low signal-to-noise ratio, edge dropout, shadows
 - Training set (machine learning methods need lots of annotated images)

DEEP BELIEF NETWORK

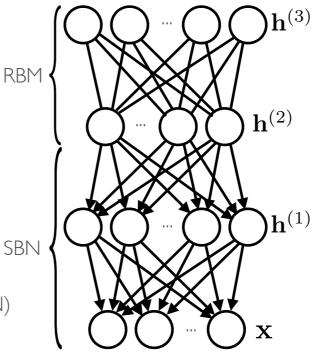
Topics: deep belief network

- The idea of pre-training came from work on deep belief networks (DBNs)
 - it is a generative model that mixes undirected and directed connections between variables
 - top 2 layers' distribution $p(\mathbf{h}^{(2)}, \mathbf{h}^{(3)})$ is an RBM
 - other layers form a Bayesian network:
 - the conditional distributions of a layers given the one above it are

$$p(h_j^{(1)} = 1 | \mathbf{h}^{(2)}) = \text{sigm}(\mathbf{b}^{(1)} + \mathbf{W}^{(2)\top} \mathbf{h}^{(2)})$$
$$p(x_i = 1 | \mathbf{h}^{(1)}) = \text{sigm}(\mathbf{b}^{(0)} + \mathbf{W}^{(1)\top} \mathbf{h}^{(1)})$$

- this is referred to as a **sigmoid belief network** (SBN)
- a DBN is not a feed-forward network

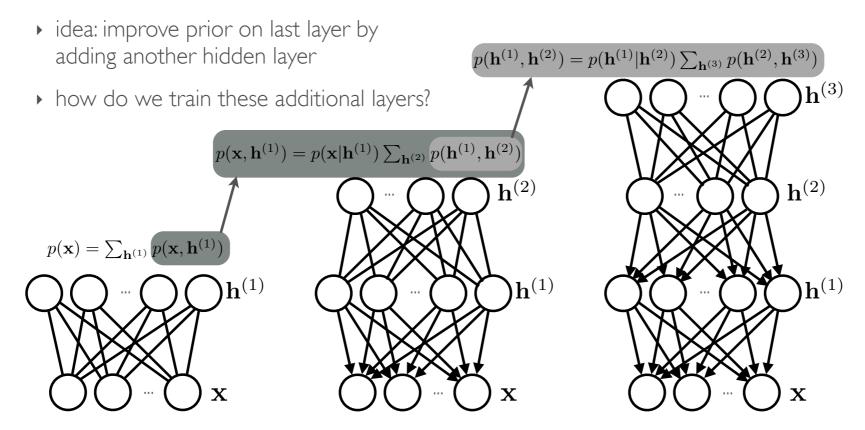
DBN's graphical model



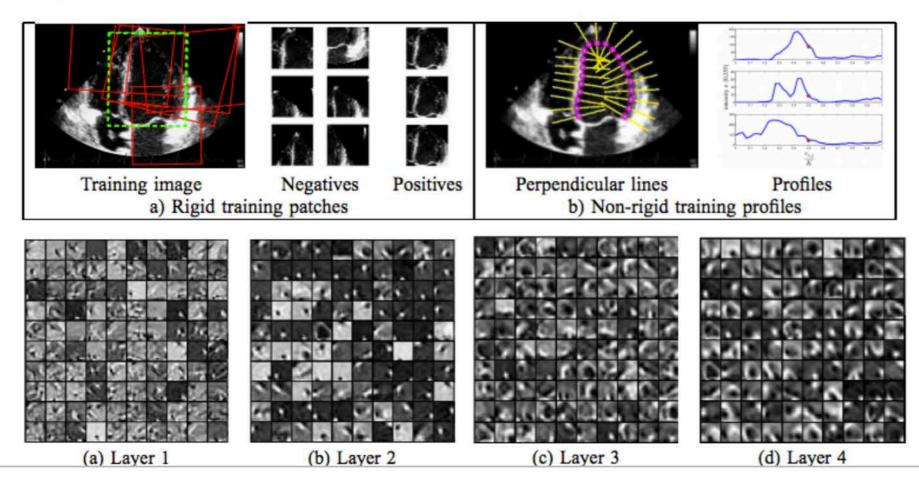
DEEP BELIEF NETWORK

Topics: deep belief network

• This is where the RBM stacking procedure comes from



- Coarse to fine search strategy (3 scales)
- ROI detected from sampling initial distribution (fewer initial points, compared to grid search)
- Gradient-based search in fine stages (less computation than grid-based search)



Importance

The volume of the hippocampus is an important trait for early diagnosis of neurological diseases (e.g., Alzheimer's disease)

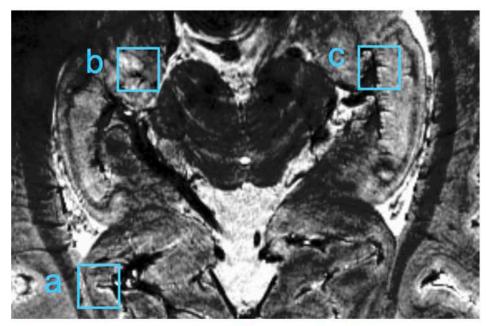
Challenges

- The hippocampus is small ($\approx 35 \times 15 \times 7$ mm³)
- The hippocampus is surrounded by complex structures
- Low imaging resolution (≈1×1×1mm³) of 1.5T or 3T MRI scanners

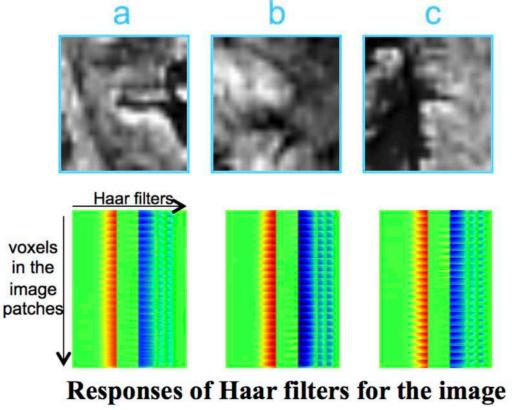
Hippocampus

Hand-Crafted Features

Limited discriminative power



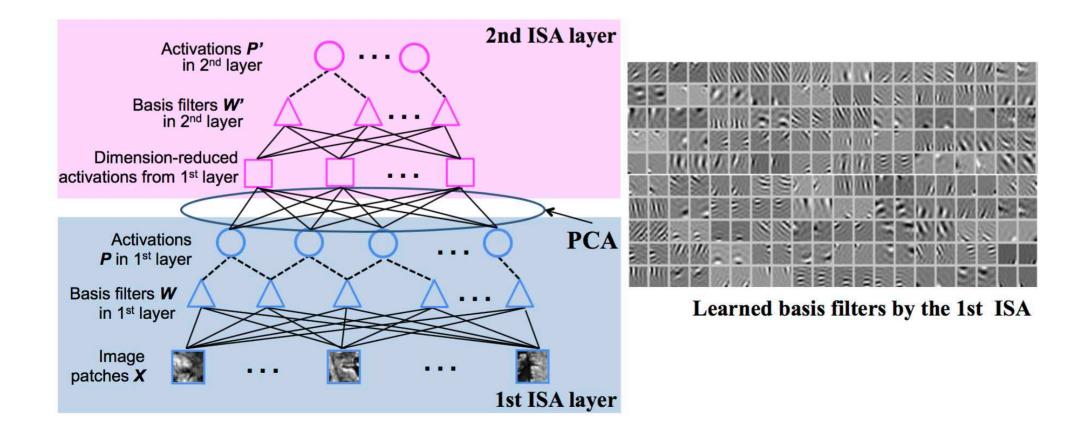
Extracting patches from a 7T MR image



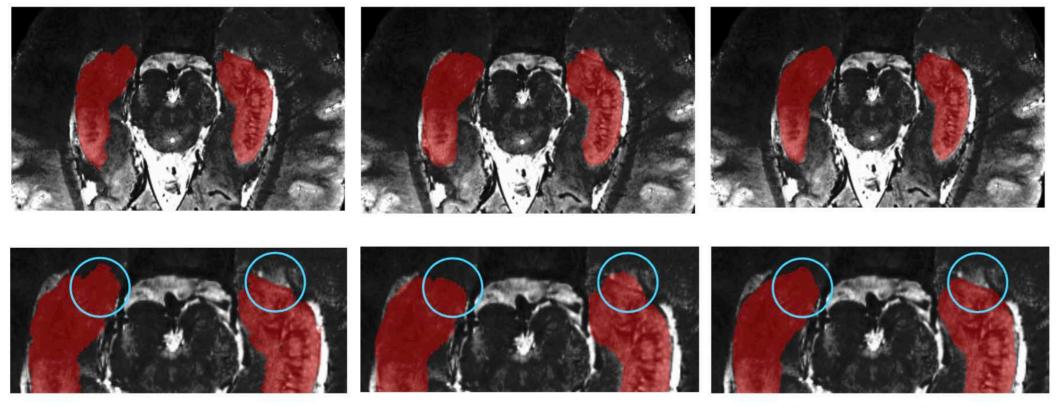
patches in a-c

Hierarchical Feature Extraction

Stacked two-layer convolutional ISA (Independent Subspace Analysis)



Qualitative Evaluations



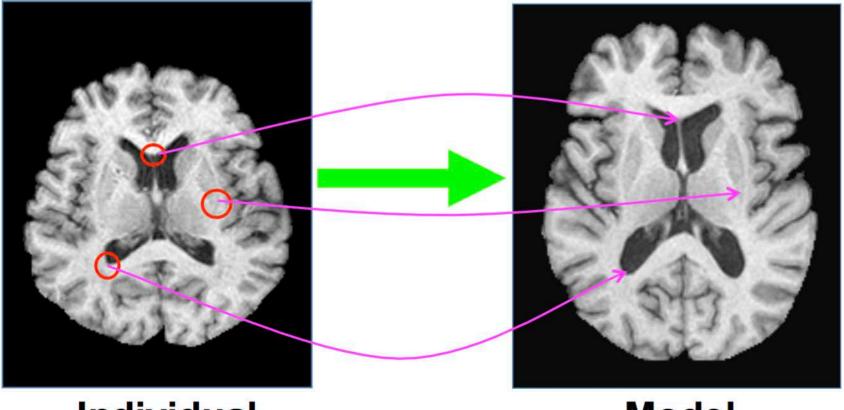
Ground Truth

Haar + Texture Features

Hierarchical Features

Examples: Image Registration

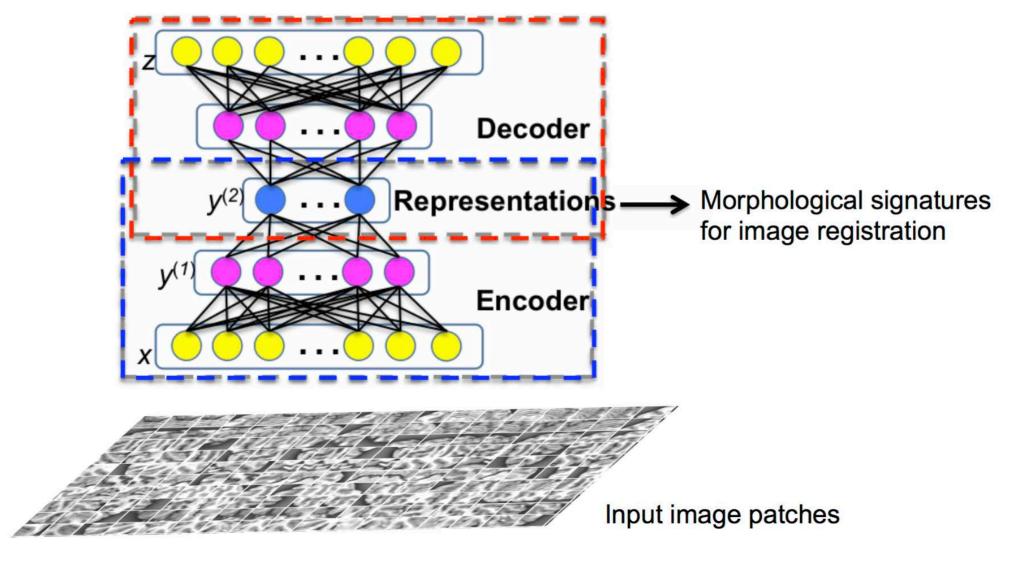
Determine accurate correspondences between images



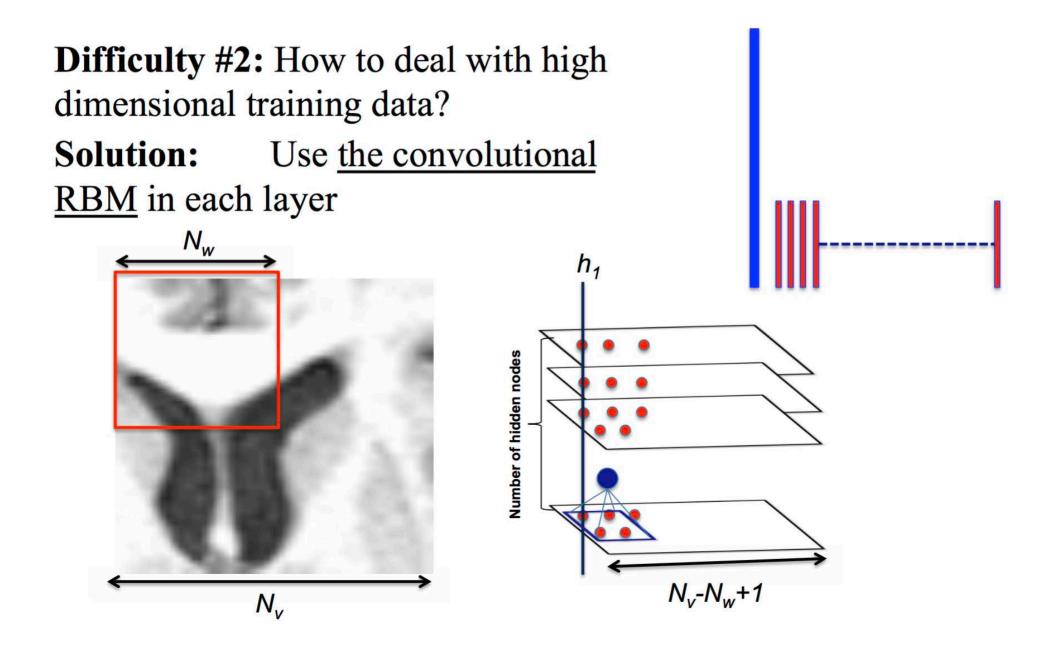
Individual

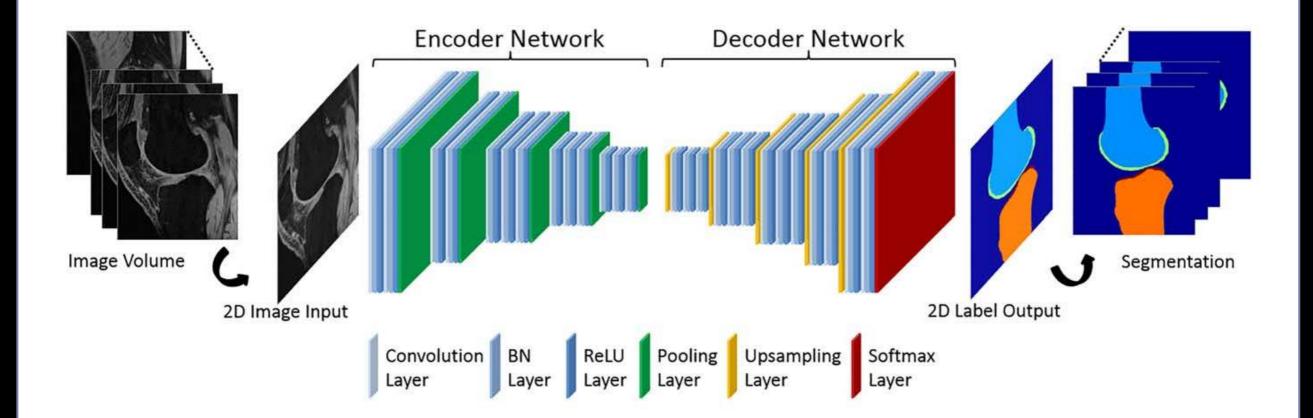
Model

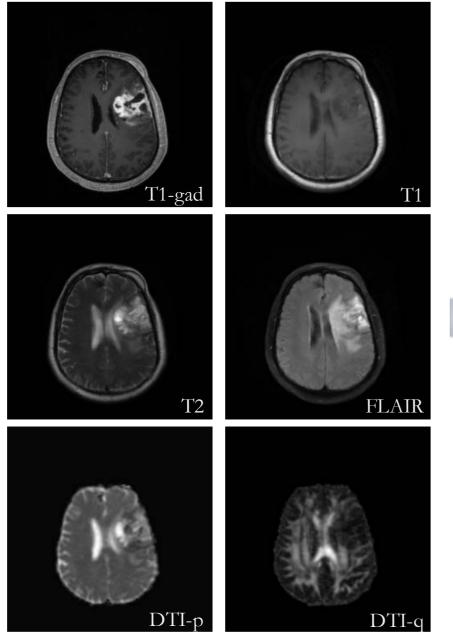
Examples: Image Registration



Examples: Image Registration

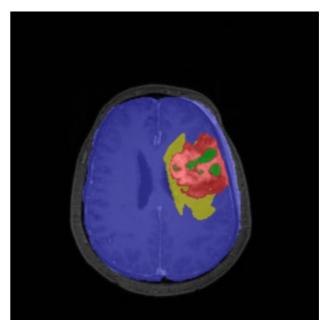




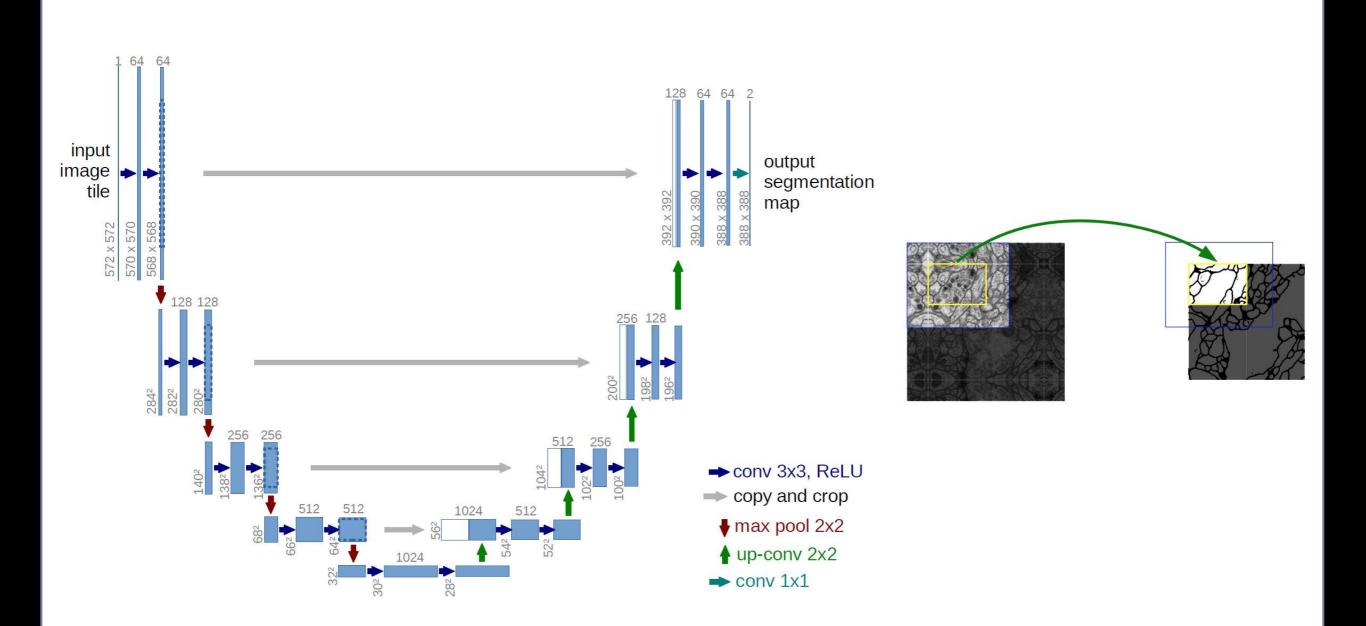


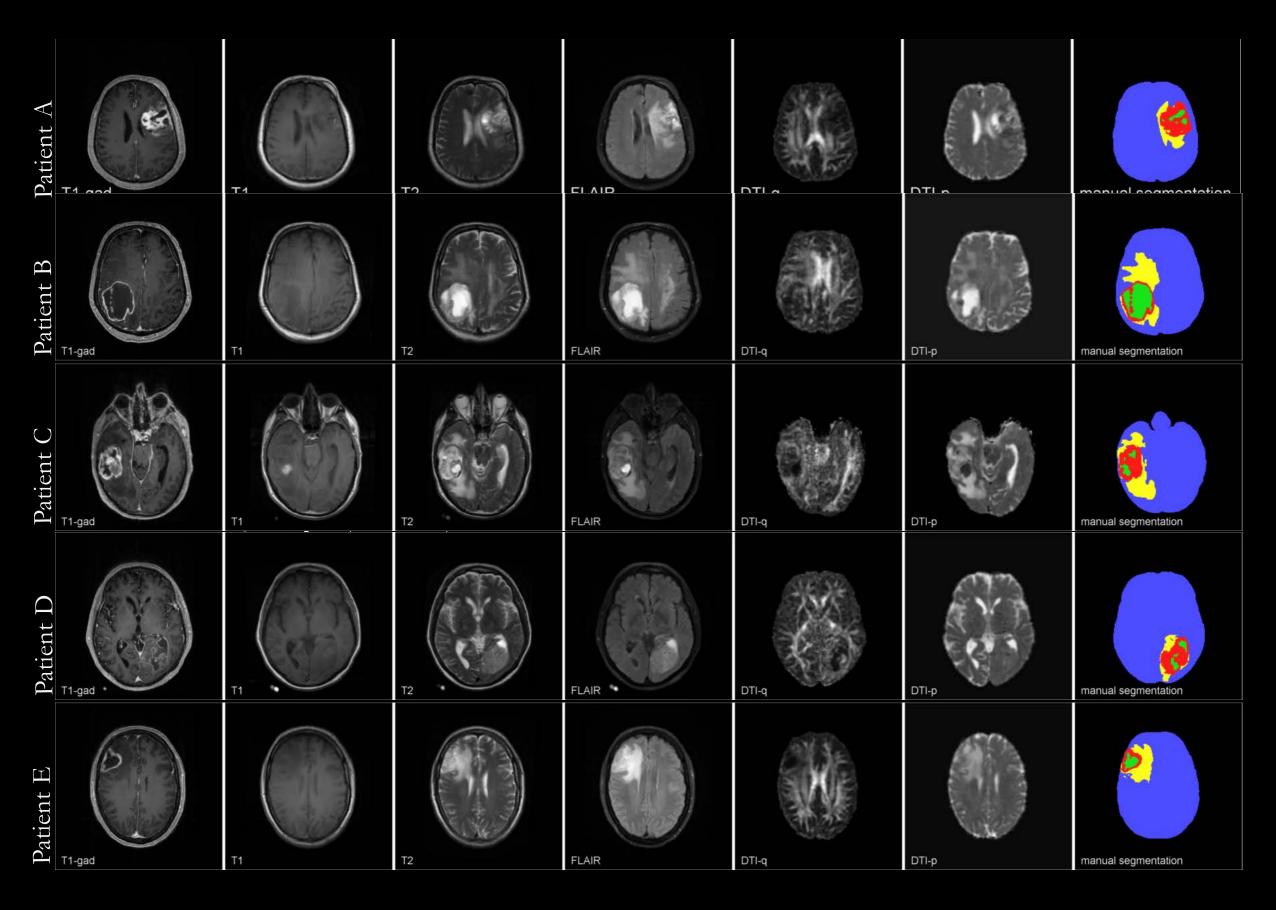
Multi-channel 3D MRI input data

Segmentation of tumorous tissues:

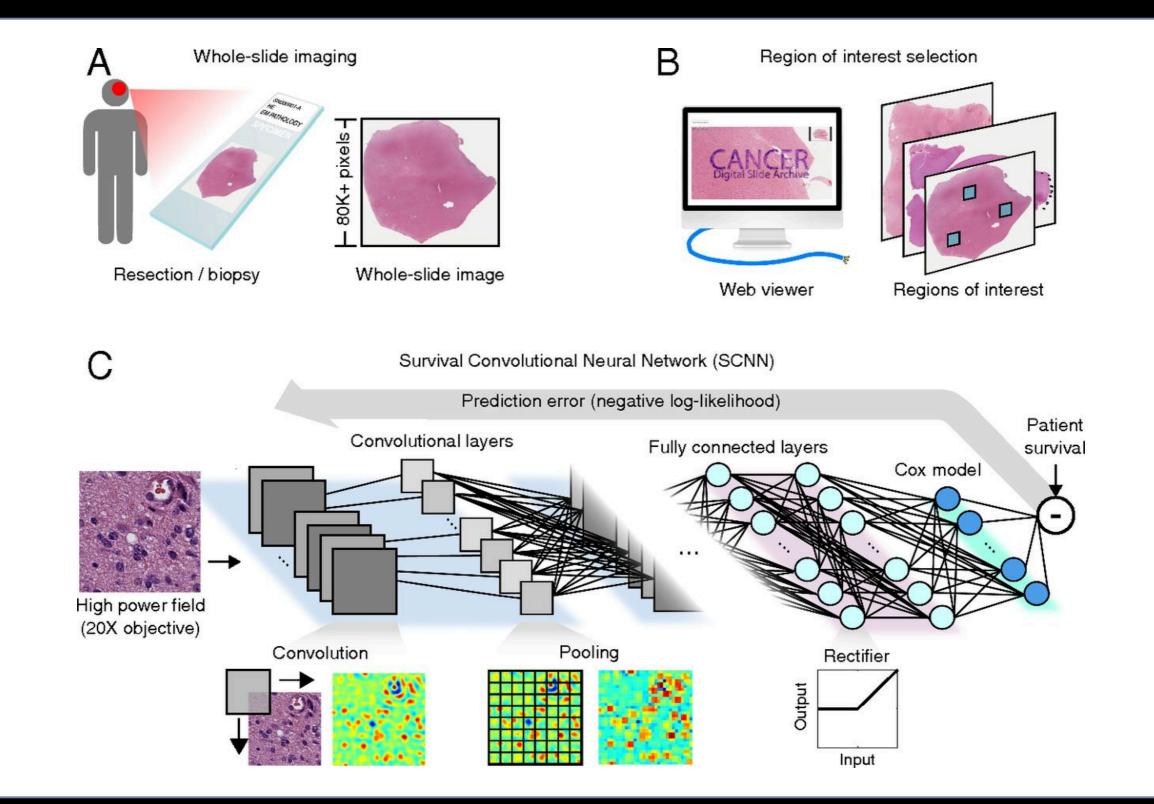


- ---- Active cells
- ---- Necrotic core
- ---- Edema
- ---- Background

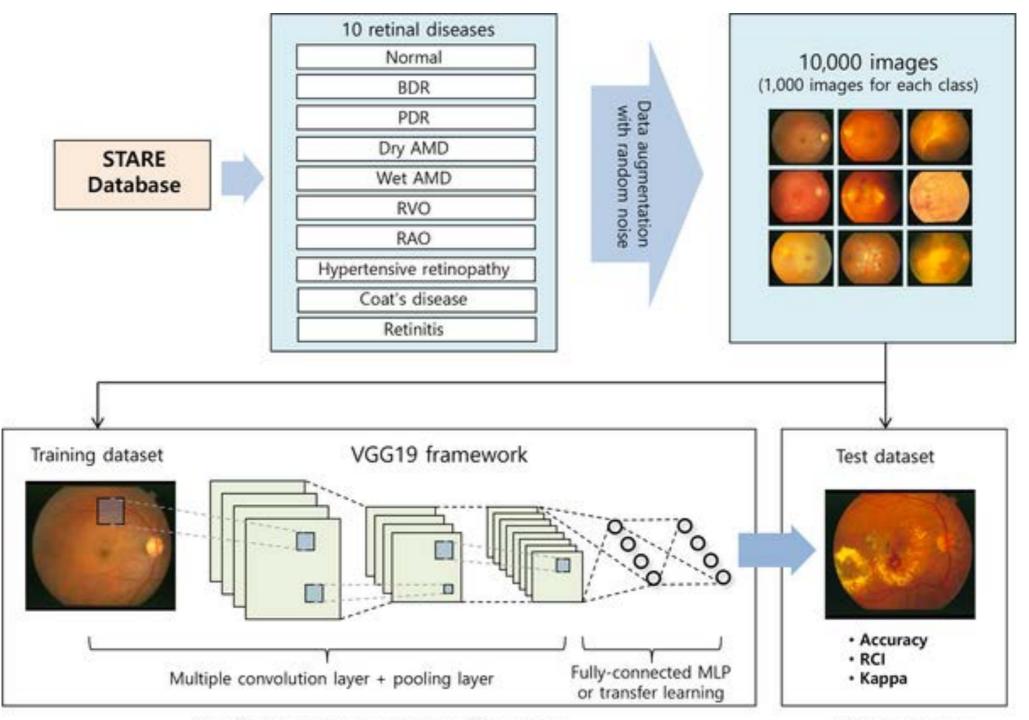




Examples: Predicting Survival from Histopathology



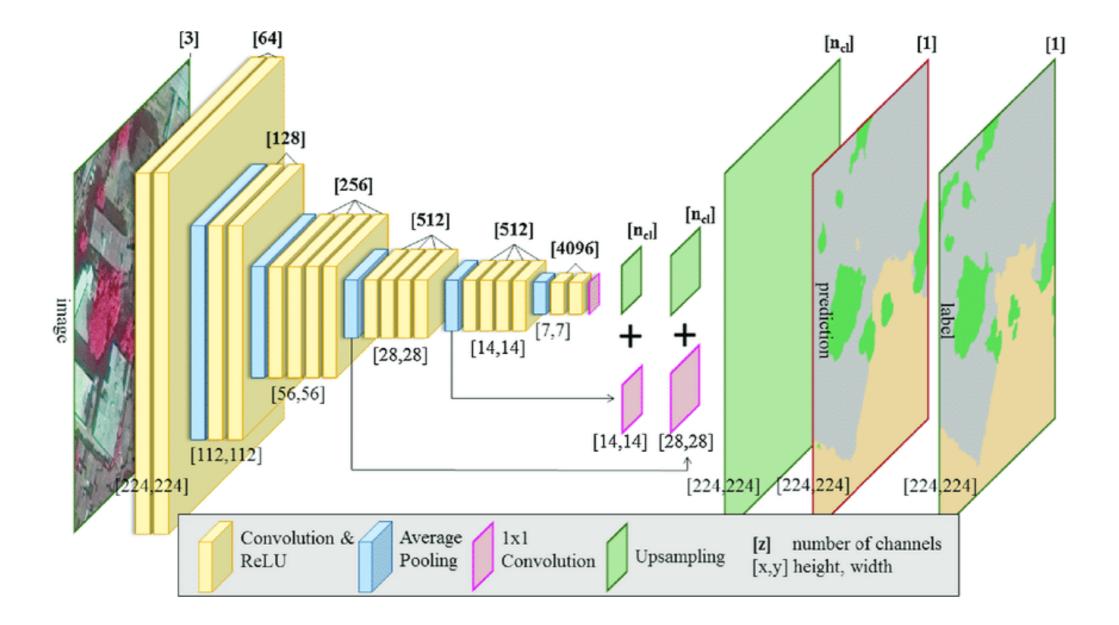
Examples: Classifying Retinal Disease



Multi-category deep learning training phase

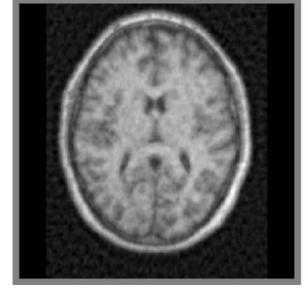
Validation phase

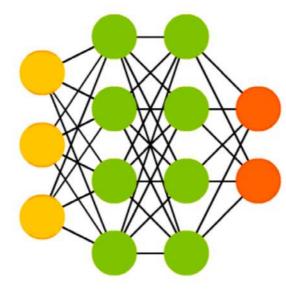
Examples: Classifying Retinal Disease



Examples: Denoising/Dealiasing Images

Reconstruction Network





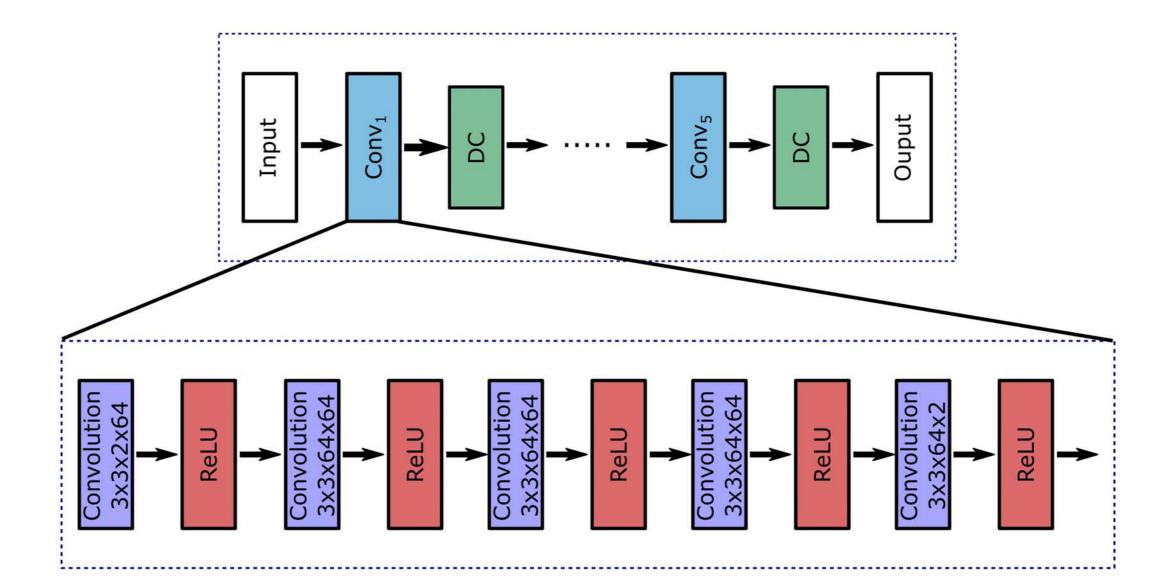


Undersampled

Recovered

$$\hat{x} = \min_{x} \left\| F_{u}x - y_{u} \right\|_{2} + \left\| C(x_{u}) - x \right\|^{2}$$
Data
Consistency
consistency
with network

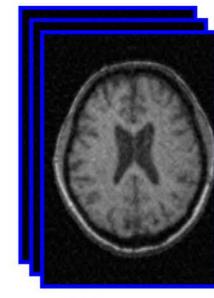
Model-based Deep Learning: Cascaded CNNs

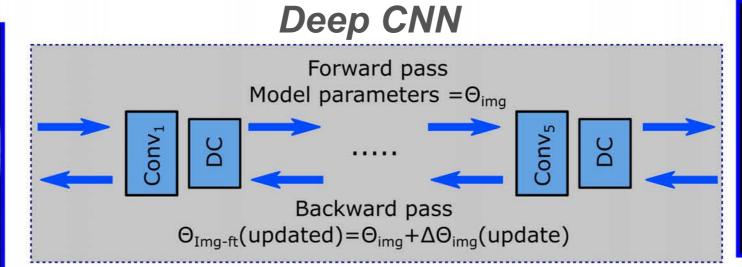


Examples: Denoising/Dealiasing images

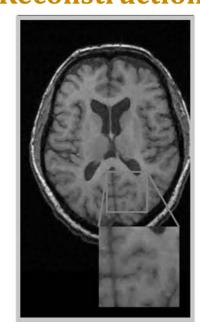
Undersampled

Fully-sampled



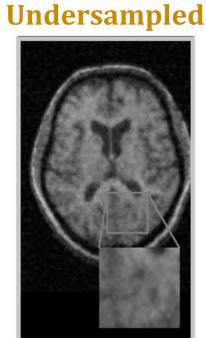


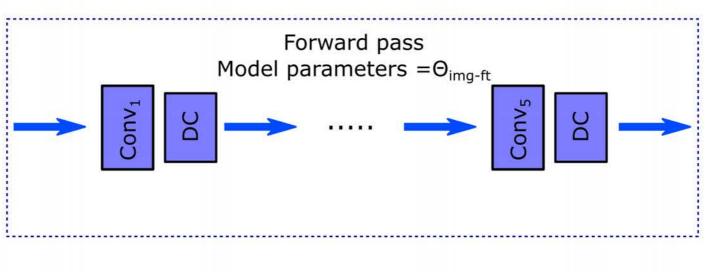




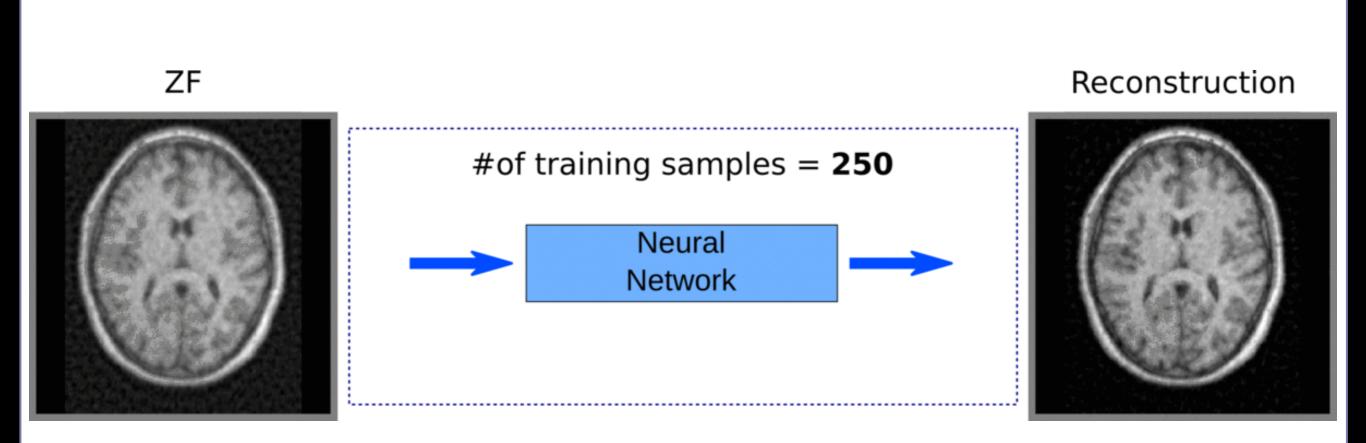
Testing

Training

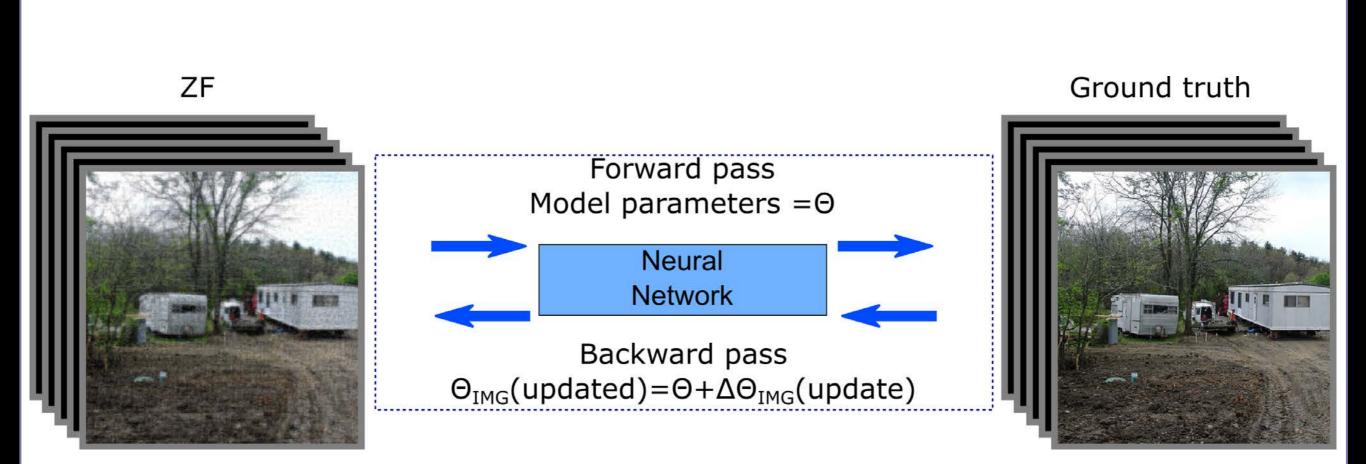




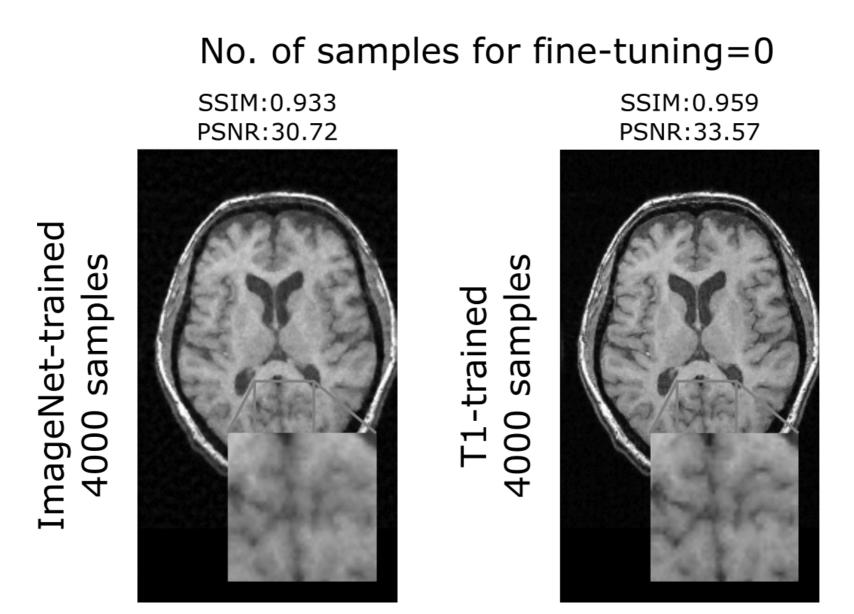
Medical Data Are Scarce



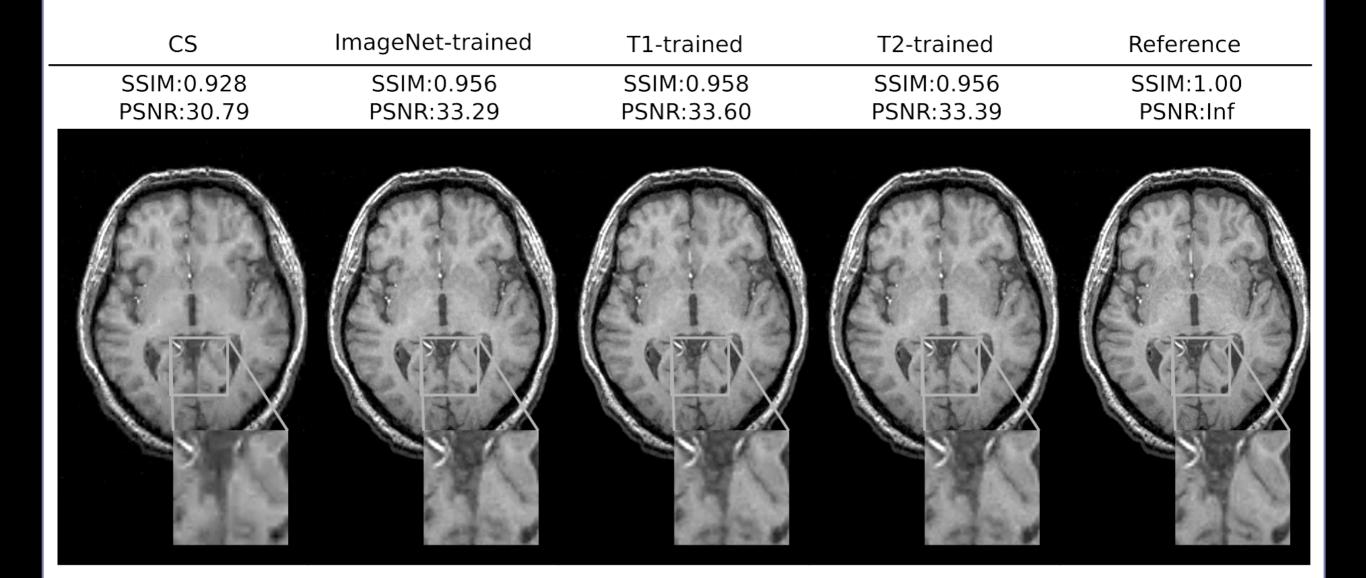
Transfer Learning



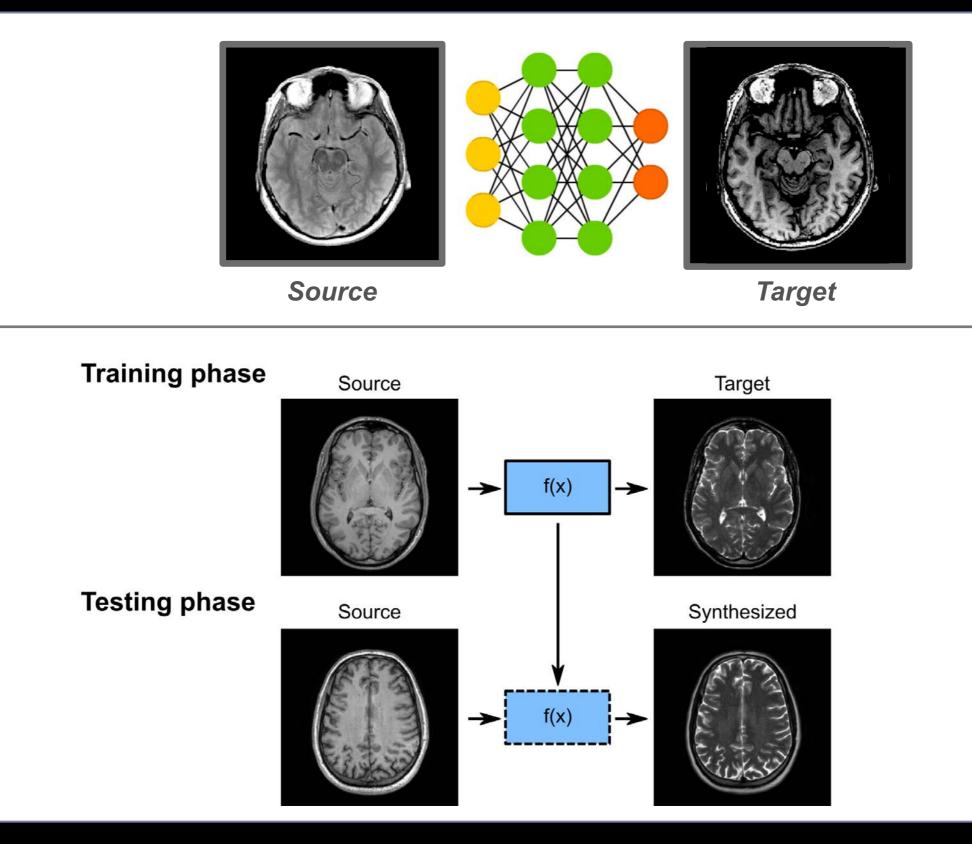
Fine Tuning



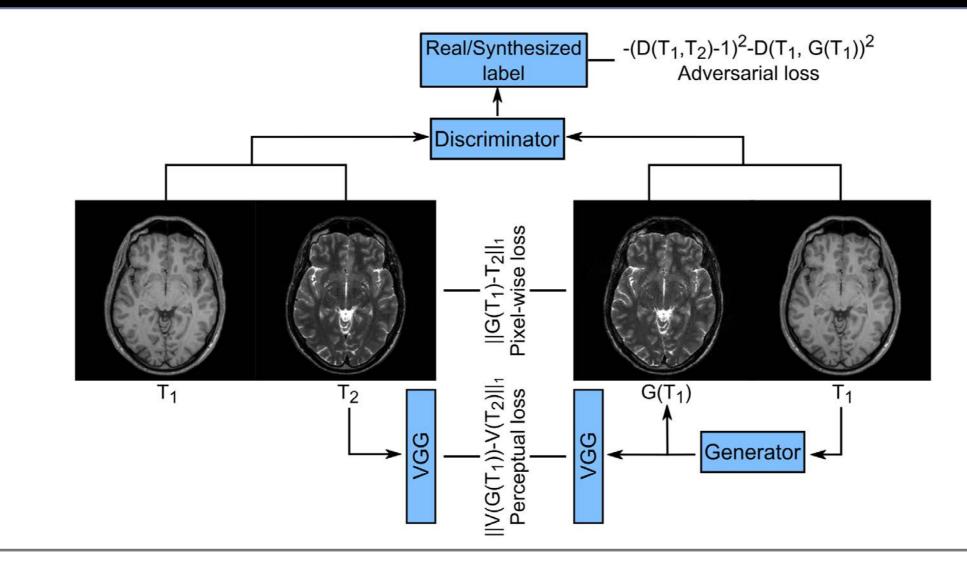
Examples: Denoising/Dealiasing images



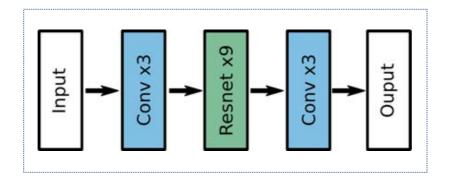
Examples: Synthesizing Missing Images



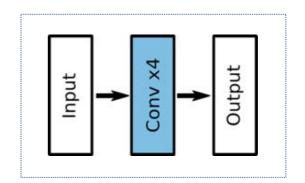
Network Architecture



Generator

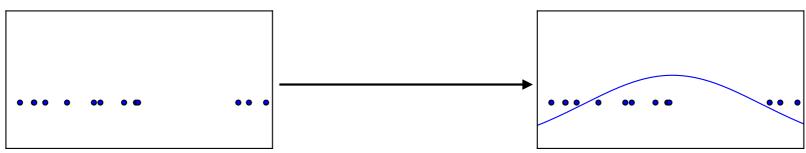


Discriminator

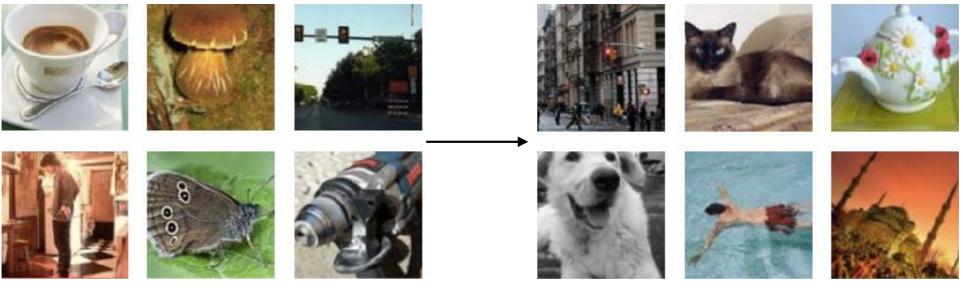


What is a Generative Model?

• Density estimation



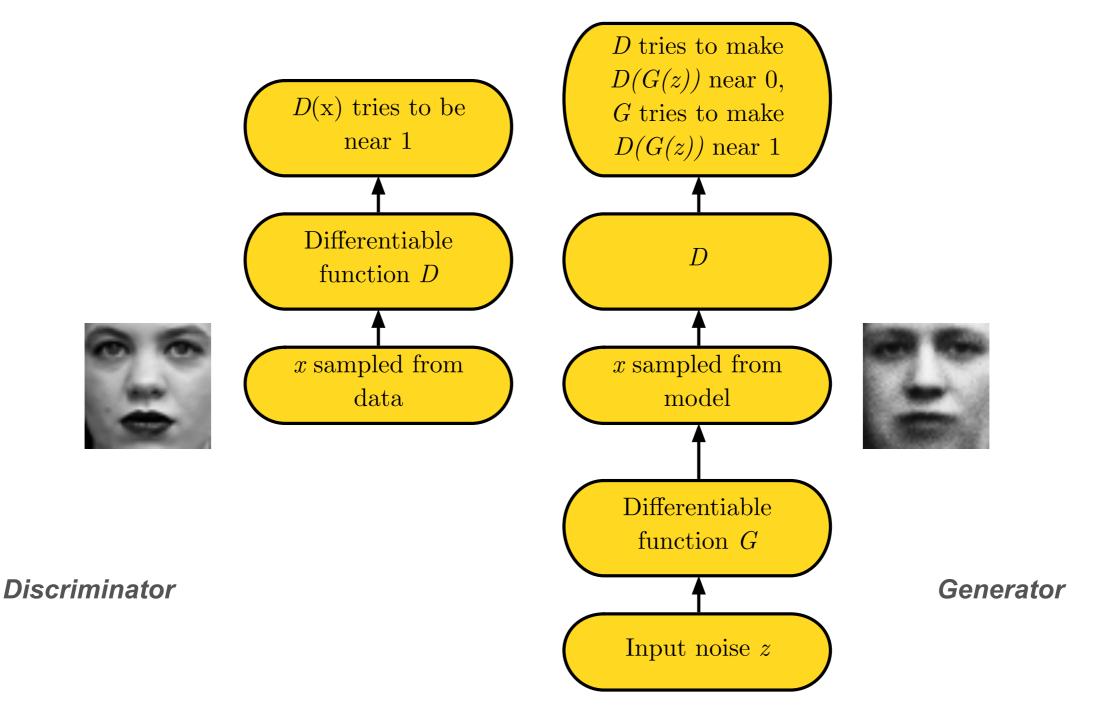
• Sample generation



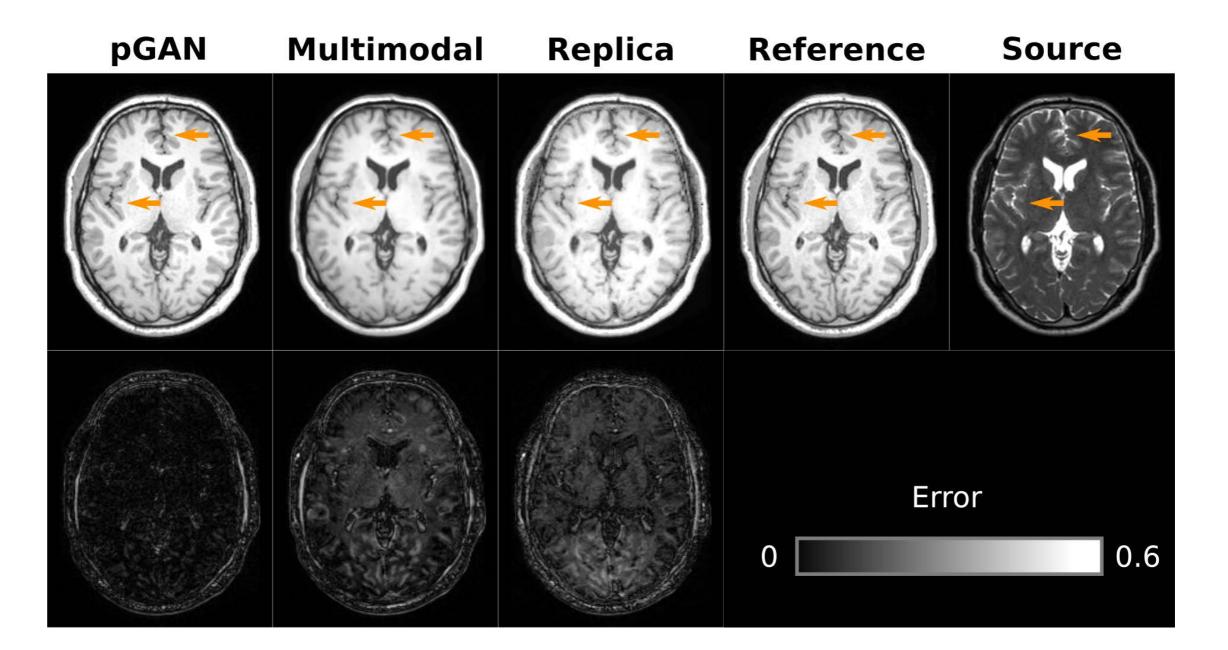
Training examples

Model samples

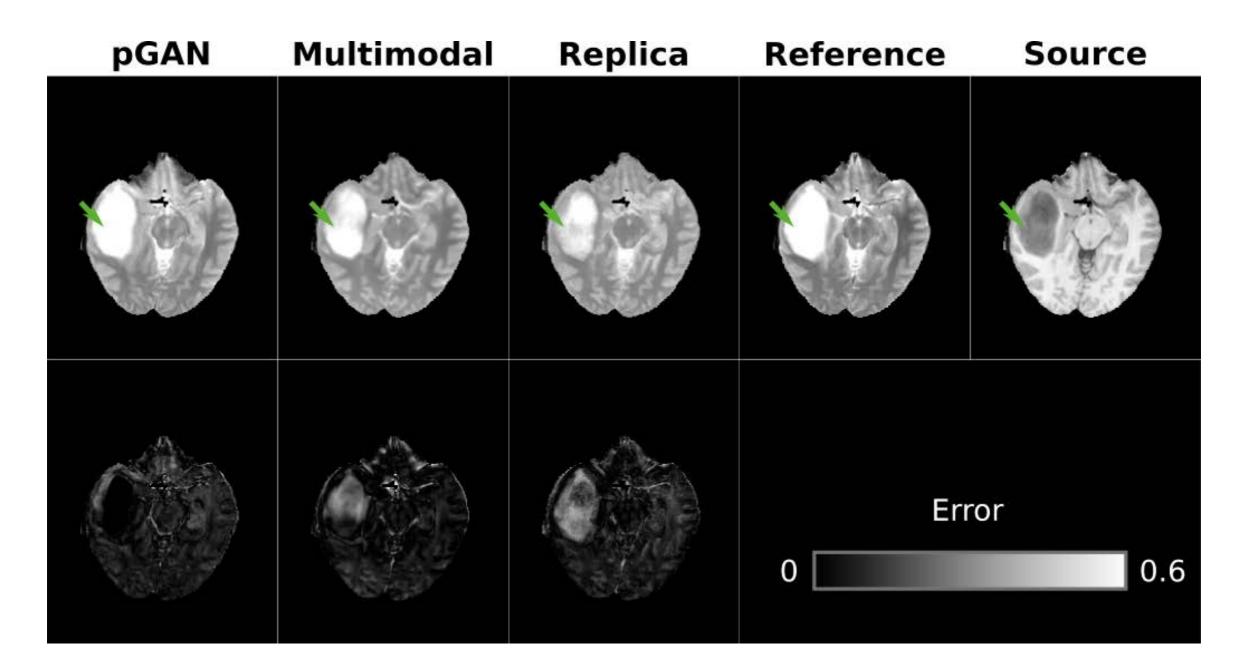
Generative Adversarial Network (GAN)



Examples: Synthesizing Missing Images



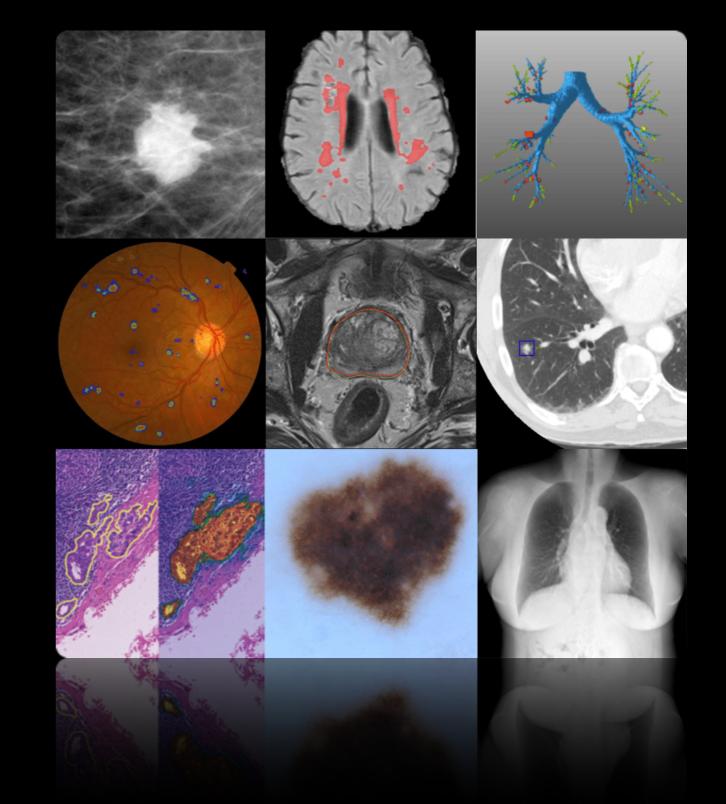
Examples: Synthesizing Missing Images



Future Outlook:

DL Tasks:

- Image Reconstruction
- Image Synthesis
- Data Acquisition
- Image Segmentation
- Tumor/Lesion Detection
- Disease Grading
- Monitoring
- Treatment Planning



Acknowledgment

Slide Credits

- Book: Principles of Neural Science
- Book: Medical Imaging Signals and Systems
- Lecture: Pascal Vincent
- Lecture: David Sontag
- Lecture: Ulas Bagci
- Lecture: Emine U Saritas
- Paper: Jarret, CVPR 2009
- Paper: Rueckert, arXiv 2019
- Paper: Waldrop, PNAS 2019
- Paper: Krizhevsky, NIPS 2012
- Paper: Jiang, SVN 2017
- Site: asimovinstitute.org
- Site: vinodsblog.com
- Site: doi.org/10.1016/j.media.2017.07.005