How GPU Computing can Accelerate the Treatment of Neurological Disorders

Eric K Oermann, MD
Anthony B Costa, PhD
Icahn School of Medicine at Mount Sinai

Disclosures

- EKO reports no relevant financial conflict of interest
- ABC reports no relevant financial conflict of interest

How can GPU computing impact neurologic disease?

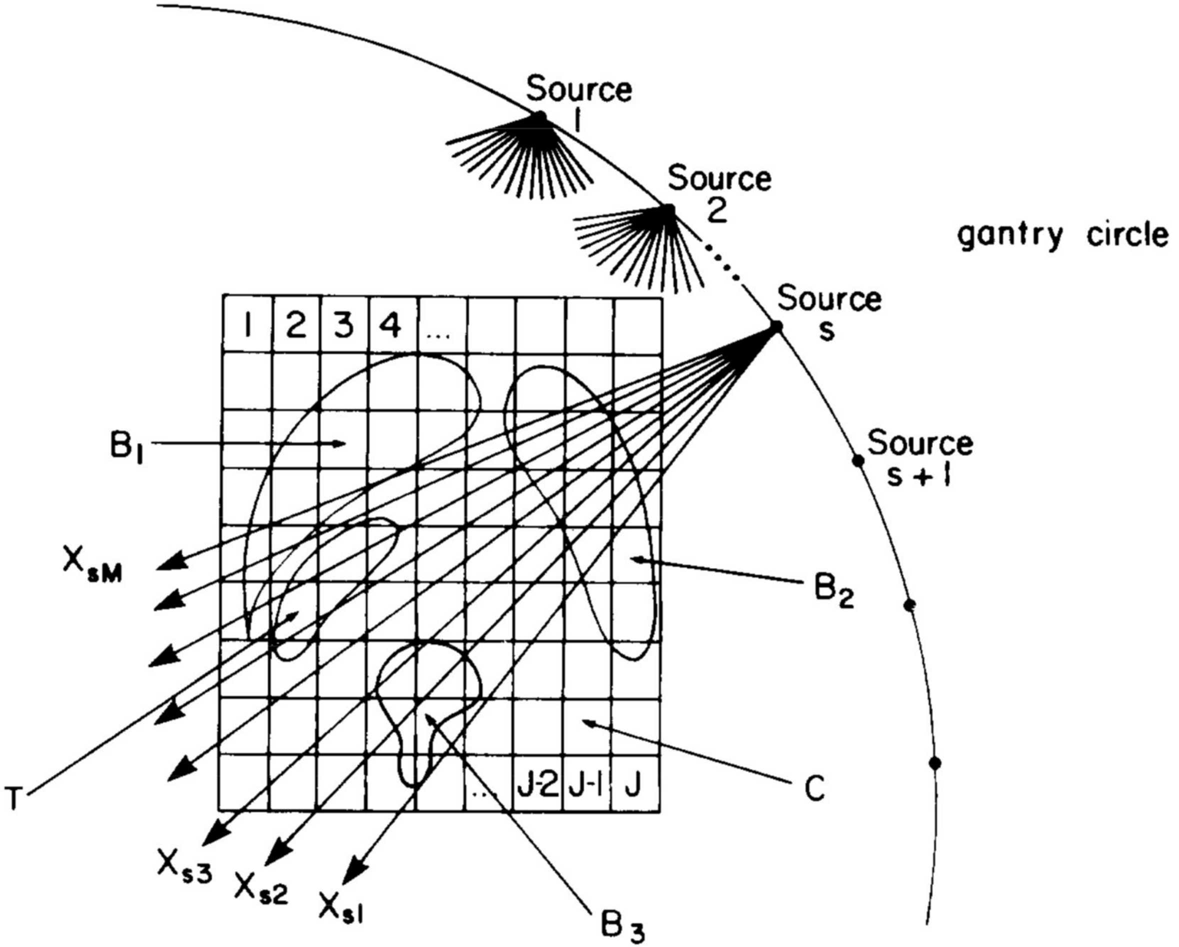
A longer story than you might think

3 Stories Enabling Neurosurgery Applications

- Computing Power → Radiation Planning
- Computing Localization → Intraoperative Applications
- Computing Density → Medical ML/DL

Basically, “what happened to enable us to build department computing resources for AI that really work?”

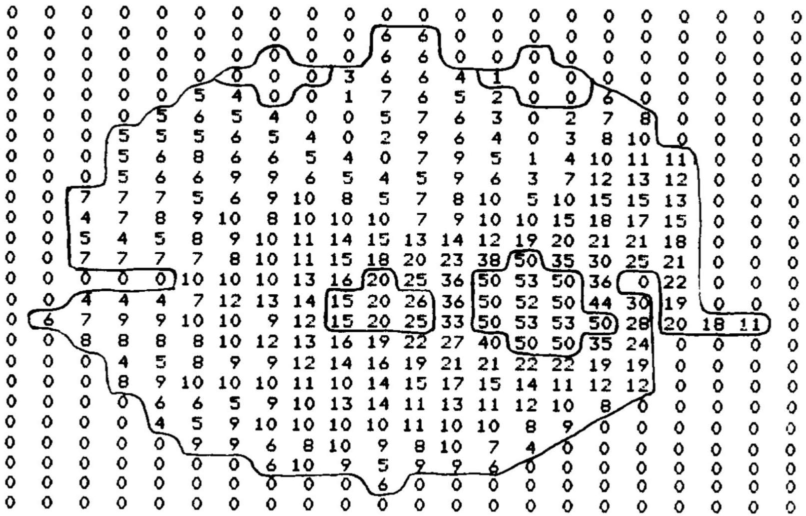
And then, what does that look like?

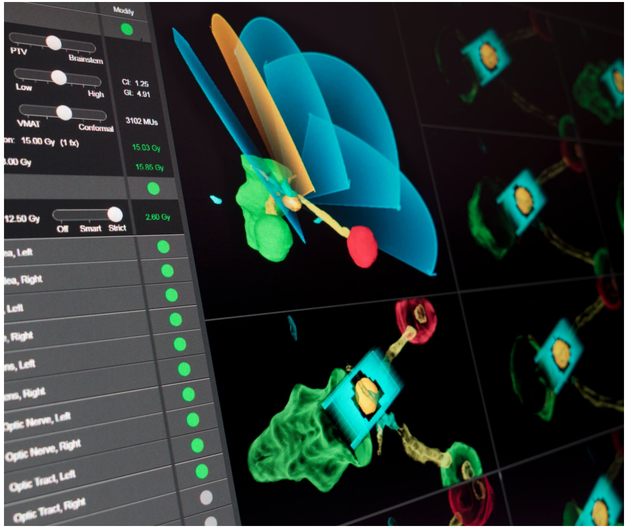


Censor, Y., Altschuler, M. D. & Powlis, W. D. Appl. Math. Comput. 25, 57–87 (1988).

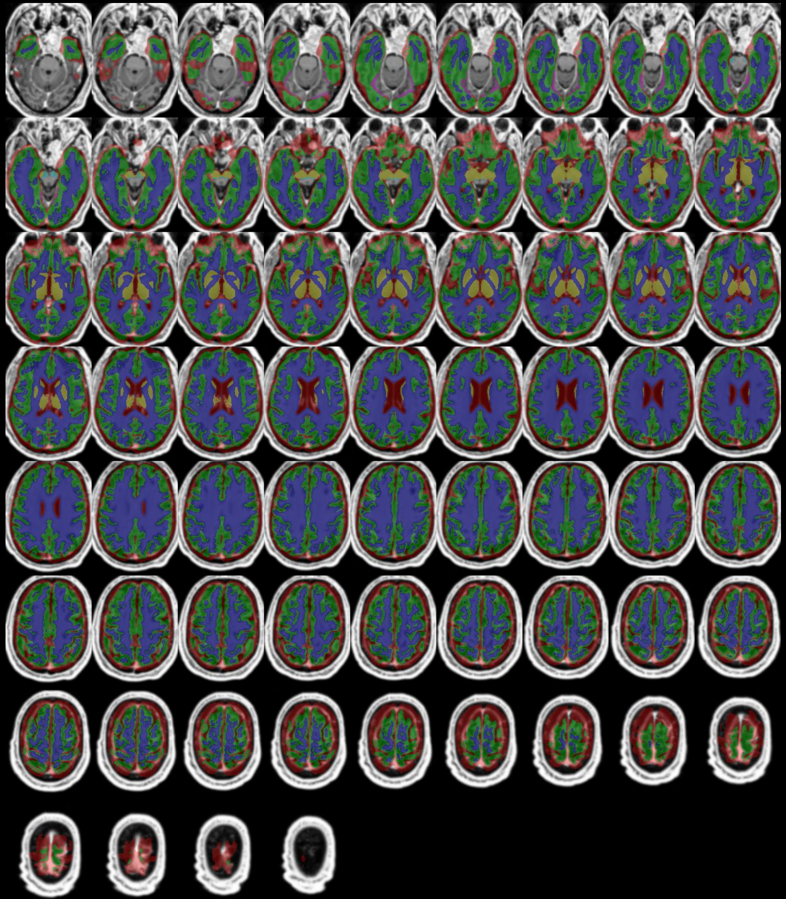
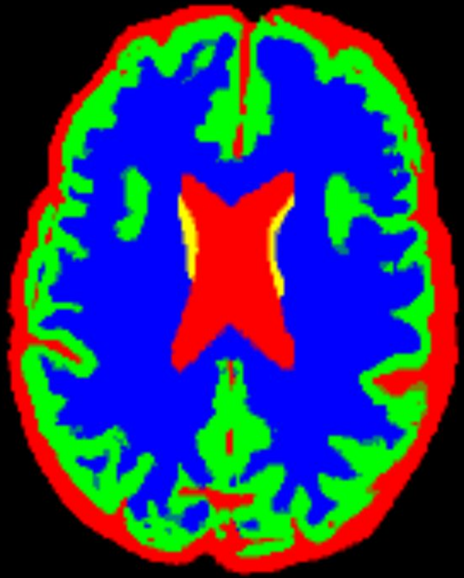
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	6	6	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	6	6	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	3	6	6	4	1	0	0	0	0	0	0	0	0
0	0	0	0	0	5	4	0	0	1	7	6	5	2	0	0	6	0	0	0	0	0
0	0	0	0	5	6	5	4	0	0	5	7	6	3	0	2	7	8	0	0	0	0
0	0	0	5	5	5	6	5	4	0	2	9	6	4	0	3	8	10	0	0	0	0
0	0	0	5	6	8	6	6	5	4	0	7	9	5	1	4	10	11	11	0	0	0
0	0	0	5	6	6	9	9	6	5	4	5	9	6	3	7	12	13	12	0	0	0
0	0	7	7	7	5	6	9	10	8	5	7	8	10	5	10	15	15	13	0	0	0
0	0	4	7	8	9	10	8	10	10	10	7	9	10	10	15	18	17	15	0	0	0
0	0	5	4	5	8	9	10	11	14	15	13	14	12	19	20	21	21	18	0	0	0
0	0	7	7	7	7	8	10	11	15	18	20	23	38	50	35	30	25	21	0	0	0
0	0	0	0	0	10	10	10	13	16	20	25	36	50	53	50	36	0	22	0	0	0
0	0	4	4	4	7	12	13	14	15	20	26	36	50	52	50	44	30	19	0	0	0
0	6	7	9	9	10	10	9	12	15	20	25	33	50	53	53	50	28	20	18	11	0
0	0	8	8	8	8	10	12	13	16	19	22	27	40	50	50	35	24	0	0	0	0
0	0	0	4	5	8	9	9	12	14	16	19	21	21	22	22	19	19	0	0	0	0
0	0	0	8	9	10	10	10	11	10	14	15	17	15	14	11	12	12	0	0	0	0
0	0	0	0	6	6	5	9	10	13	14	11	13	11	12	10	8	0	0	0	0	0
0	0	0	0	4	5	9	10	10	10	10	11	10	10	8	9	0	0	0	0	0	0
0	0	0	0	0	9	9	6	8	10	9	8	10	7	4	0	0	0	0	0	0	0
0	0	0	0	0	0	0	6	10	9	5	9	9	6	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Censor, Y., Altschuler, M. D. & Powlis, W. D. Appl. Math. Comput. 25, 57–87 (1988).

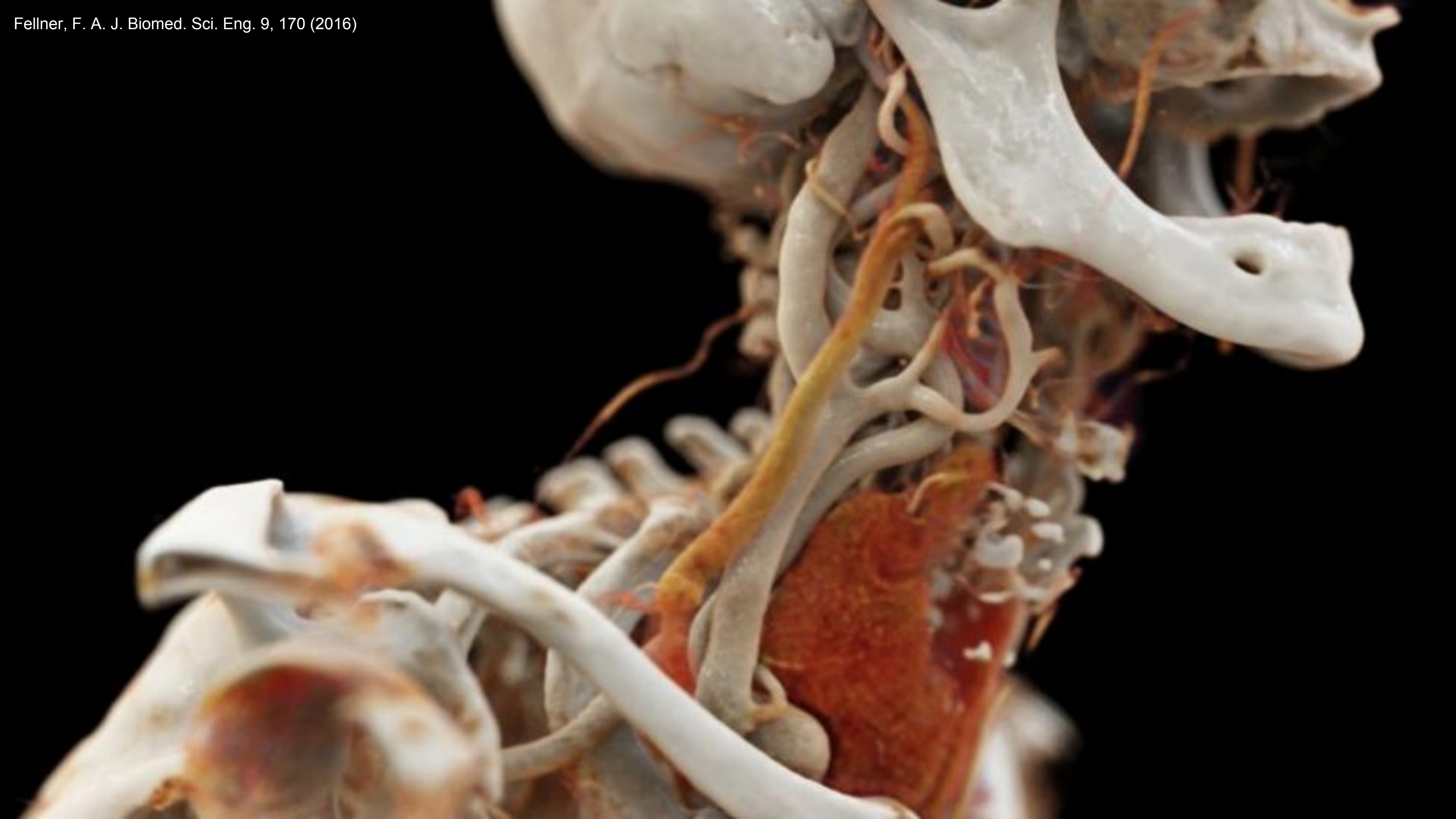




Censor, Y., Altschuler, M. D. & Powlis, W. D. Appl. Math. Comput. 25, 57–87 (1988).
https://www.brainlab.com/press-releases/brainlab-optimizes-planning-processes-algorithms-cranial-indications/

Fellner, F. A. J. Biomed. Sci. Eng. 9, 170 (2016)




VIVE
SURGICAL THEATER

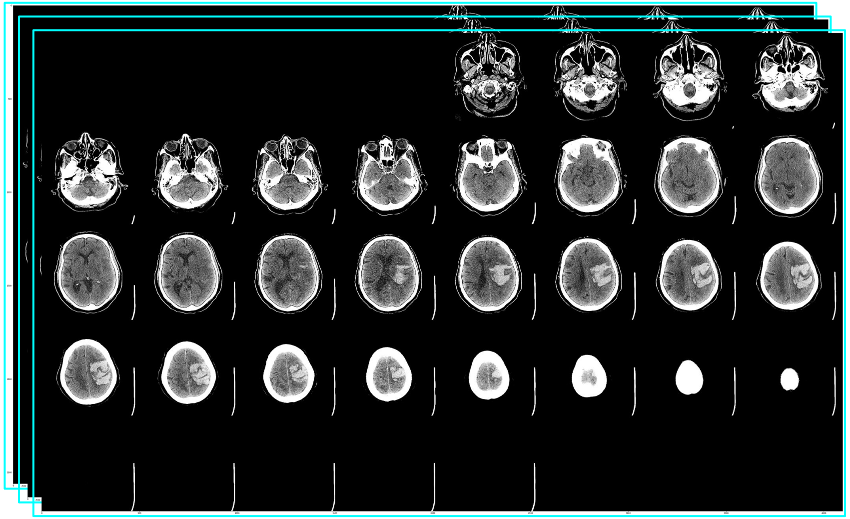
Needs of academic, medical DL

- Understand varied medical data needs
- Mixed compute/data access patterns
- Performance per dollar (financial constraints)
- Access to appropriate storage that can handle imaging down to free text
- Unified infrastructure, authentication and appropriate HIPAA privacy controls
- Support for current and future generation computing paradigms
  - E.g., Docker, Container frameworks

Medical Imaging Data IS big data

Consider 1 megapixel, 8 bit detector (# in batch, z, x, y, # channels):

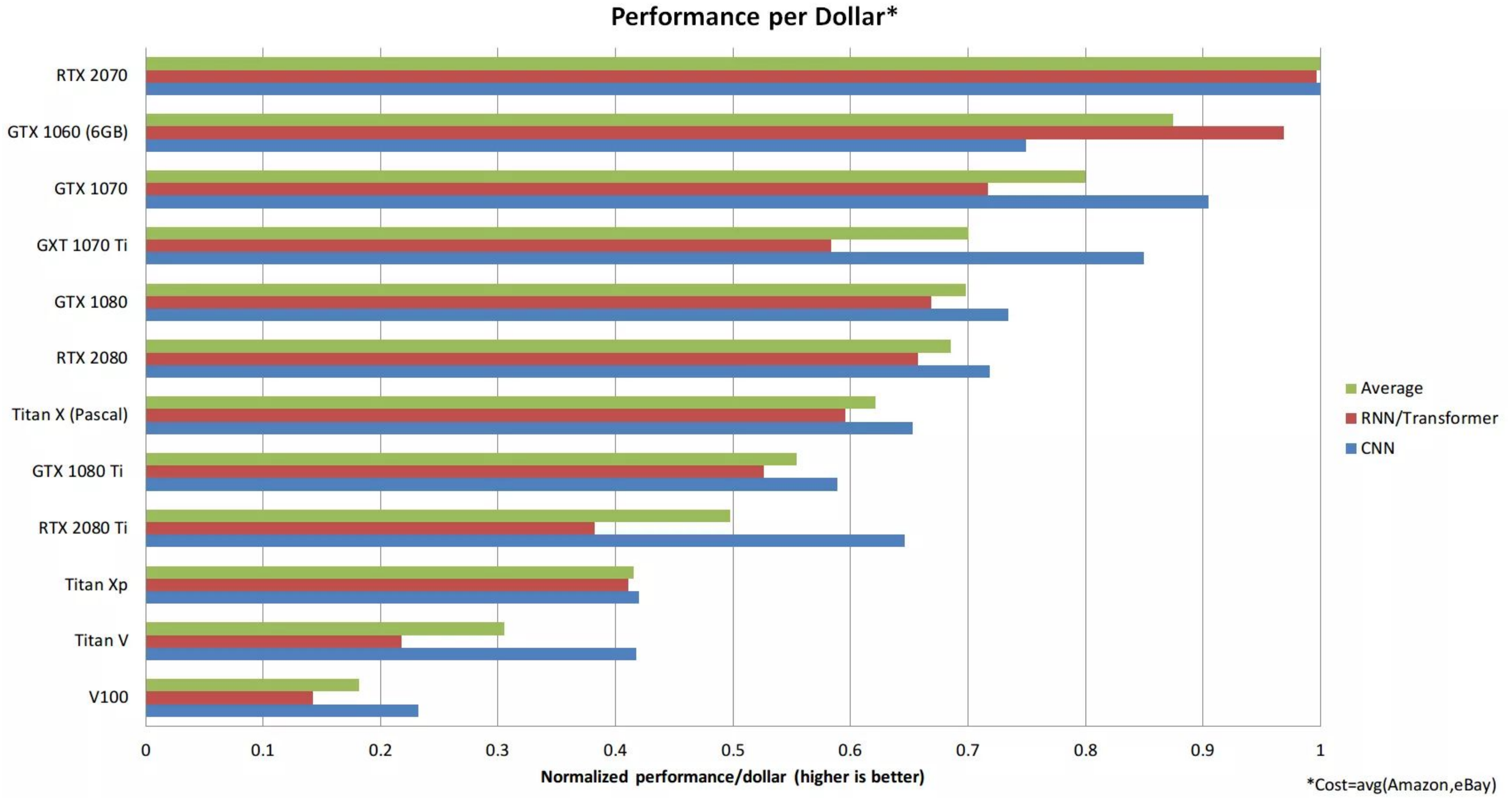
- Single slice / 2D image (1, 1, 1024, 1024, 1) = 1 Mb
- 3D image with 100 slices (1, 100, 1024, 1024, 1) = 100 Mb
- 1024 images/batch (1024, 100, 1024, 1024, 1) = **100 Gb**



- Memory
- Precision
- Bandwidth

- Performance/$/Watt per application
  - 2D Imaging
  - 3D Volumetric Imaging
  - NLP, RNN, Time Series
  - Reinforcement Learning

- Comes down to:
  - What’s your data?
  - What’s your method?
  - What’s your benchmark for performance?
  - How rich are you and how much do you value your time?

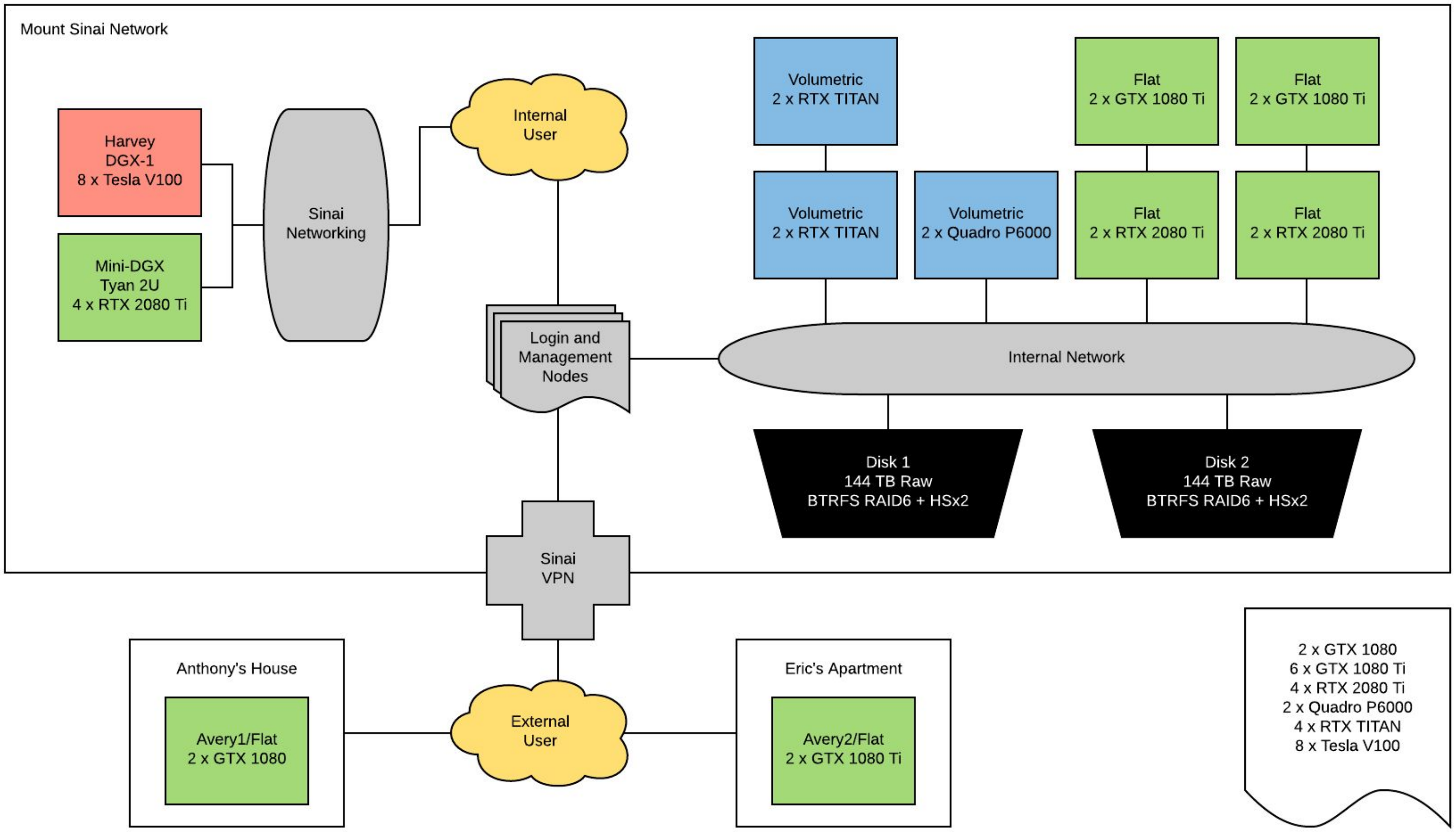


http://timdettmers.com/2018/11/05/which-gpu-for-deep-learning/

Academic medical centers tend to start with what they know and evolve

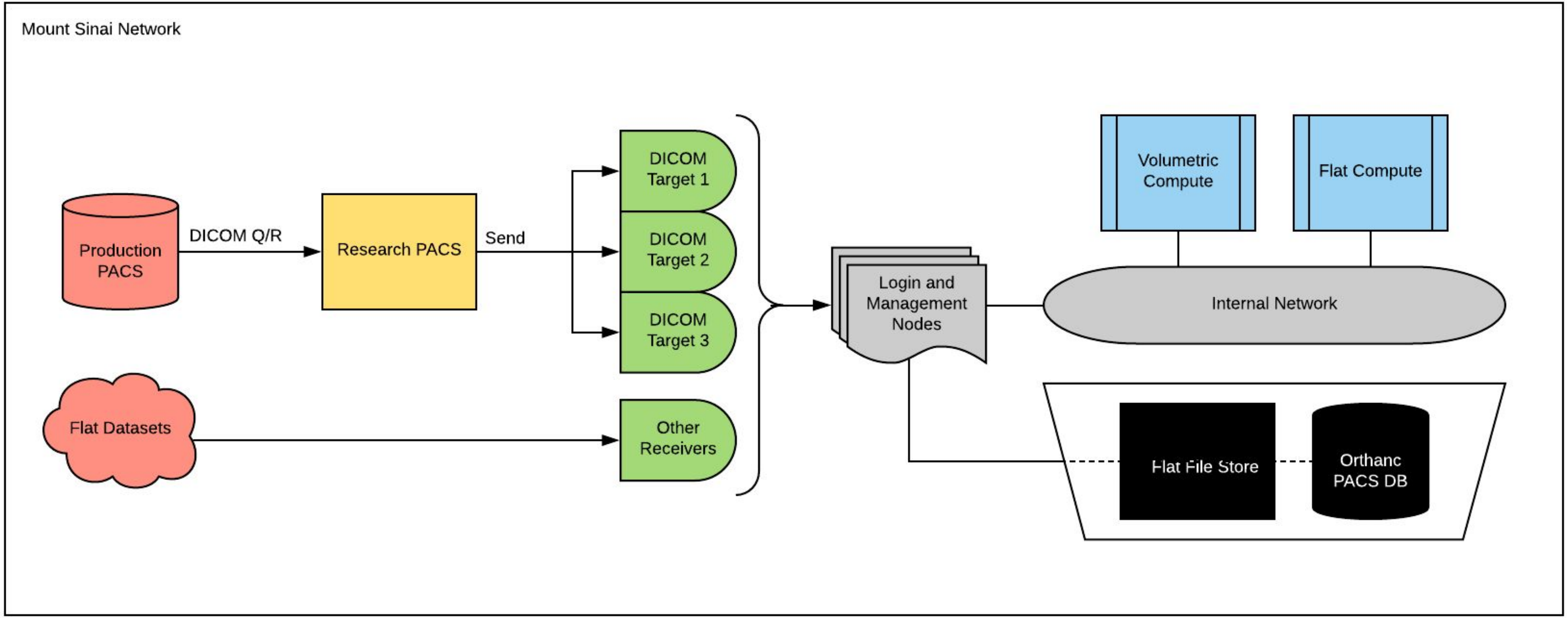
Management

- V1: Classic HPC Cluster
  - YP/NIS Authentication
  - Manual Time Sharing
  - NFS v3 XFS 20TB
- V2: Major Expansion, Not-So-Classic HPC Cluster
  - Transition to Docker/Container Frameworks
  - Manual Time Sharing
  - Manual Authentication
  - NFS v3 XFS 20TB + Local Flash/Scratch HDDs
  - Flat/Volumetric Box Allocation to Specific Projects


Mount Sinai Network
Harvey
DGX-1
8 x Tesla V100
Mini-DGX
Tyan 2U
4 x RTX 2080 Ti
Sinai
Networking
Internal
User
Volumetric
2 x RTX TITAN
Flat
2 x GTX 1080 Ti
Flat
2 x GTX 1080 Ti
Volumetric
2 x RTX TITAN
Volumetric
2 x Quadro P6000
Flat
2 x RTX 2080 Ti
Flat
2 x RTX 2080 Ti
Login and
Management
Nodes
Internal Network
Disk 1
144 TB Raw
BTRFS RAID6 + HSx2
Disk 2
144 TB Raw
BTRFS RAID6 + HSx2
Sinai
VPN
Anthony's House
Avery1/Flat
2 x GTX 1080
External
User
Eric's Apartment
Avery2/Flat
2 x GTX 1080 Ti
2 x GTX 1080
6 x GTX 1080 Ti
4 x RTX 2080 Ti
2 x Quadro P6000
4 x RTX TITAN
8 x Tesla V100

Total Compute

- “Flat” GPUs, Consumer GTX/RTX
    - Great bang for your buck, limited appropriateness for 3D volumetric work due to small amount of on-die memory (8-12GB)
    - 2 x GTX 1080 (FP32 8TF)
    - 6 x GTX 1080 Ti (FT32 10TF)
    - 2 x GTX 2080 Ti (FP32 14TF, **110TF w/ Tensor Cores**)
- “Volumetric” GPUs, Mid-Level and Enterprise
    - 3 - 10x Cost, ~double the memory
    - 2 x Quadro P6000 (FP32 12TF, 24GB OD, FP64)
    - 4 x RTX Titan (FP32 16TF, **130TF w/ Tensor Cores**, 24GB OD, RP INT4/8 + FP16/64)
    - 8 x Tesla V100 (FP32 16TF, **125TF w/ Tensor Cores,** 32GB OD, RP INT4/8 + FP16/64)
- Total Tensor flops: **5.6PF** + General Purpose FP32 @ 0.86PF


Mount Sinai Network
Production PACS
DICOM Q/R
Research PACS
Send
DICOM Target 1
DICOM Target 2
DICOM Target 3
Flat Datasets
Other Receivers
Login and Management Nodes
Volumetric Compute
Flat Compute
Internal Network
Flat File Store
Orthanc PACS DB

Management

- V3: Next-Generation Containerized Cluster
  - Towards DeepOps
  - NFS v4 288TB BTRFS RAID6 + HSs
  - LDAP Unified Authentication (2 Factor + Sinai VPN)
  - Role-Based Data Access Validation
  - ContainerOS
  - Kubernetes Docker Orchestration Framework
  - Flat/Volumetric PXE Thin Nodes
  - Managed Docker Containers for All Projects

Bitbucket

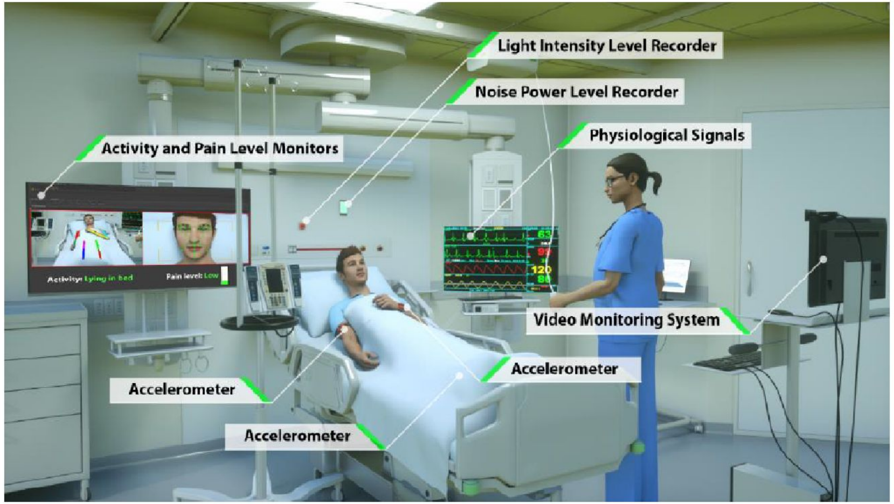
slack

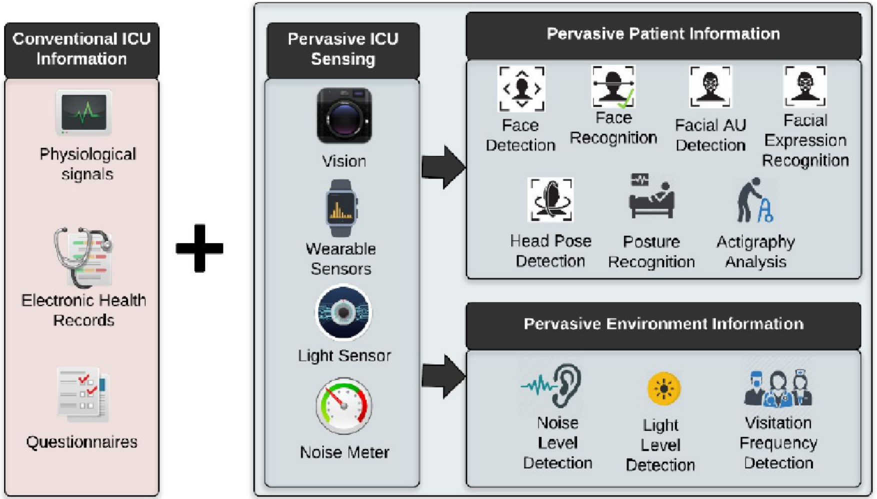
docker

How can machine learning (on GPUs) impact neurological disease?

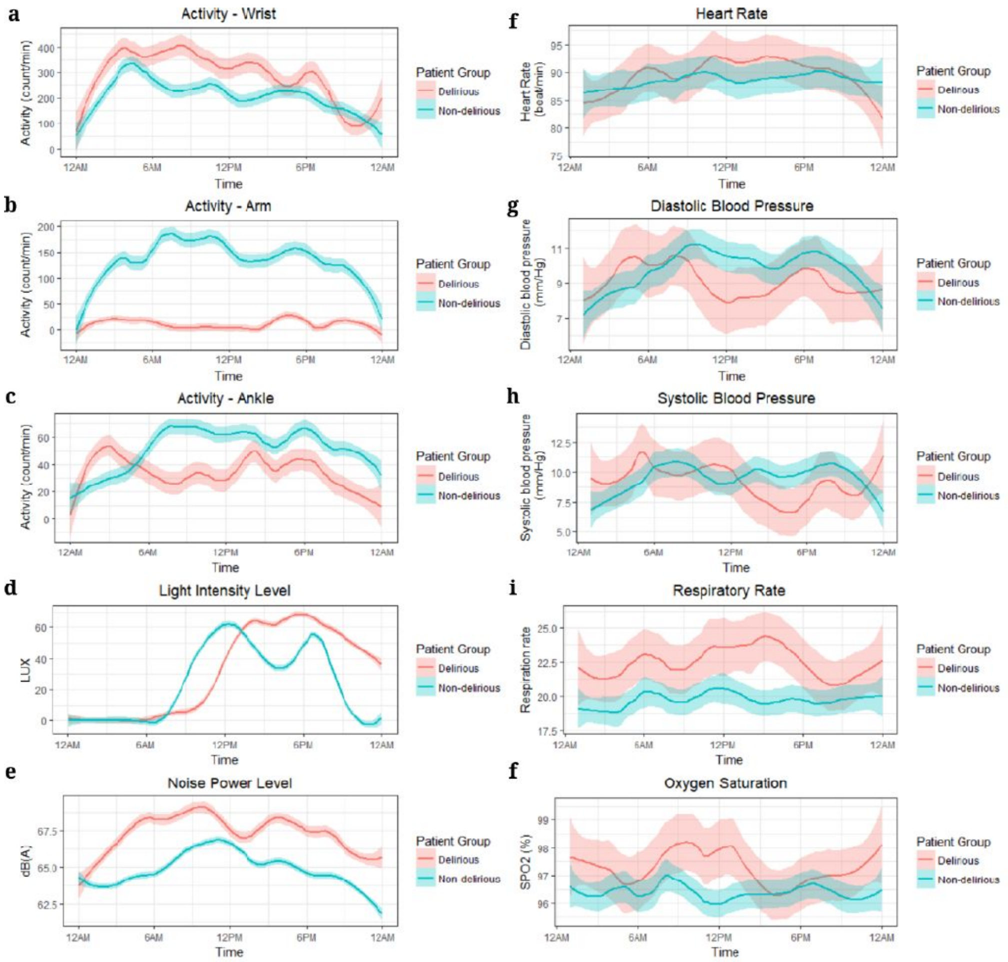
A universe of new applications

Assessments in the Neuro-ICU

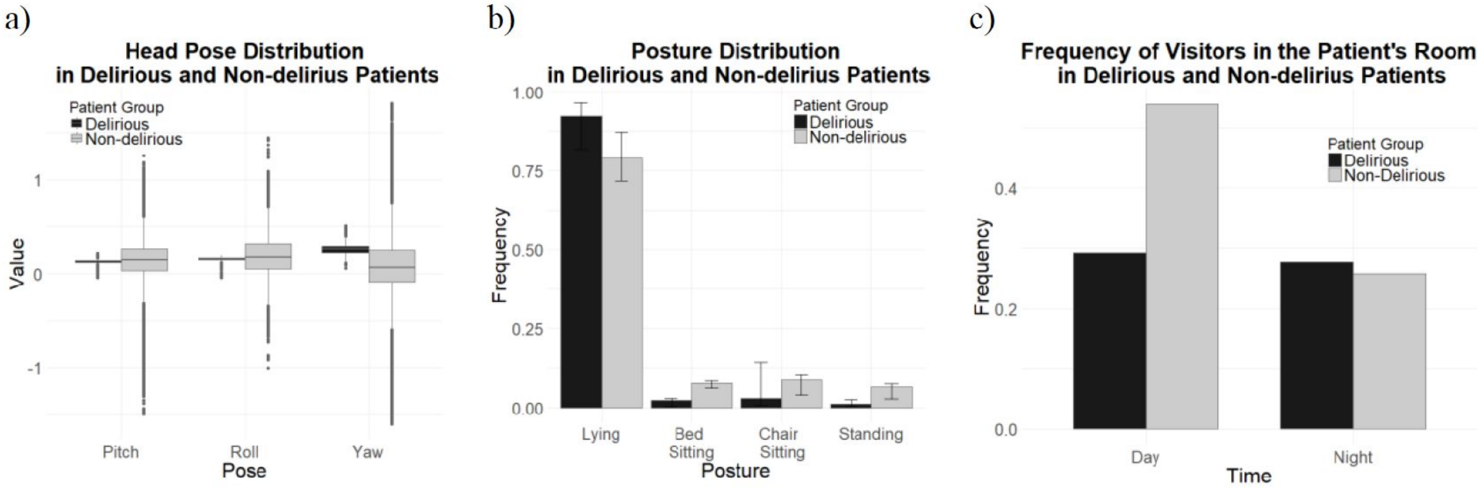




Davoudi, A. et al. The Intelligent ICU Pilot Study: Using Artificial Intelligence Technology for Autonomous Patient Monitoring. arXiv [cs.HC] (2018).



Davoudi, A. et al. The Intelligent ICU Pilot Study: Using Artificial Intelligence Technology for Autonomous Patient Monitoring. arXiv [cs.HC] (2018).



	Non-delirious (N=15)	Delirious (N=5)	p-value
Mean of activity for whole day, median (IQR)	53.9 (19.5, 161.6)	332 (251, 457.6)	**0.01**
Mean of activity for daytime	69.7 (16.7, 198.1)	347.4 (318.9, 384.8)	**0.03**
Standard deviation of activity for daytime	246.8 (99.3, 472.5)	640.6 (487.5, 697.2)	**0.05**
Mean of activity for nighttime	46.4 (22.8, 94.7)	332.3 (310.7, 541.8)	**0.008**
Standard deviation of activity for nighttime	192.5 (130.5, 313.3)	664.1 (469.1, 930.9)	**<0.01**
Activity of 10-hour window with highest sum of activity (M10)	60137.3 (15029.5, 176498.3)	282918.1 (254729.3, 457448)	**<0.01**
Time of M10	318 (157, 574)	275 (47, 627)	1
Time of M10 (hour)	6 (3, 9)	5 (1, 11)	1
Activity of 5-hour window with lowest sum of activity (L5)	3916.7 (1195.7, 10236.2)	44163.2 (1949.3, 54779.3)	0.35
Time of L5	298 (176, 959)	1067 (212, 1119)	0.36
Time of L5 (hour)	5 (3.5, 16.5)	18 (4, 19)	0.36
Relative amplitude	0.9 (0.7, 0.9)	0.9 (0.8, 1)	0.61
Standard deviation of activity for whole day	199.9 (116.3, 456.3)	558.4 (523.1, 826.4)	**0.02**
RMSSD	223.4 (137.3, 469.7)	538.7 (487.4, 730.2)	**0.04**
RMSSD/SD	1.1 (1.0, 1.2)	0.9 (0.9, 1)	**<0.01**
Number of immobile minutes during the day	564 (416, 654)	345 (200, 384)	**0.02**
Number of immobile minutes during the night	602 (580, 650)	344 (314, 374)	**0.01**

Davoudi, A. et al. The Intelligent ICU Pilot Study: Using Artificial Intelligence Technology for Autonomous Patient Monitoring. arXiv [cs.HC] (2018).

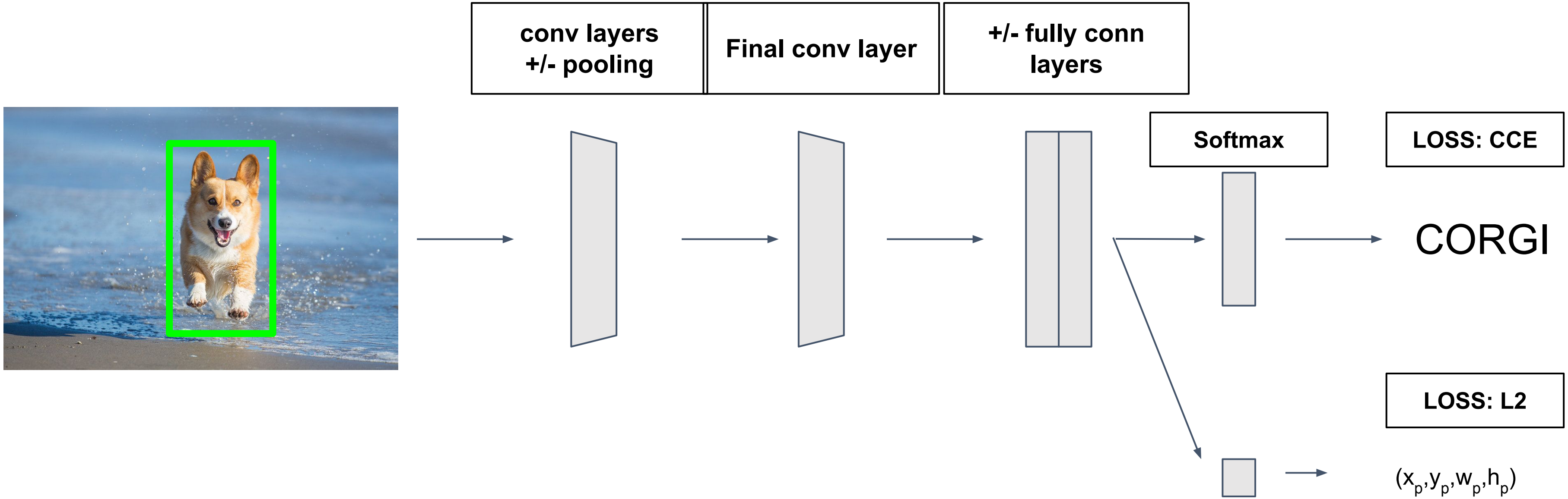
Convolutional Neural Network Approaches to Brain Imaging

Classification and Localization

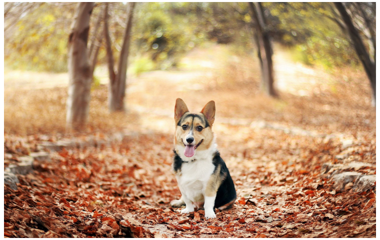
- **Input**: N classes + BBox (x,y,w,h)
- **Output**: Class K where K is in N + (xp,yp,wp,hp)
- **Performance Metrics**: Accuracy + Jaccard similarity (or Dice)

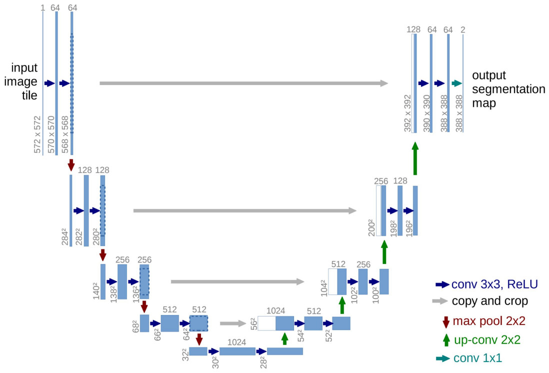
$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}.$$

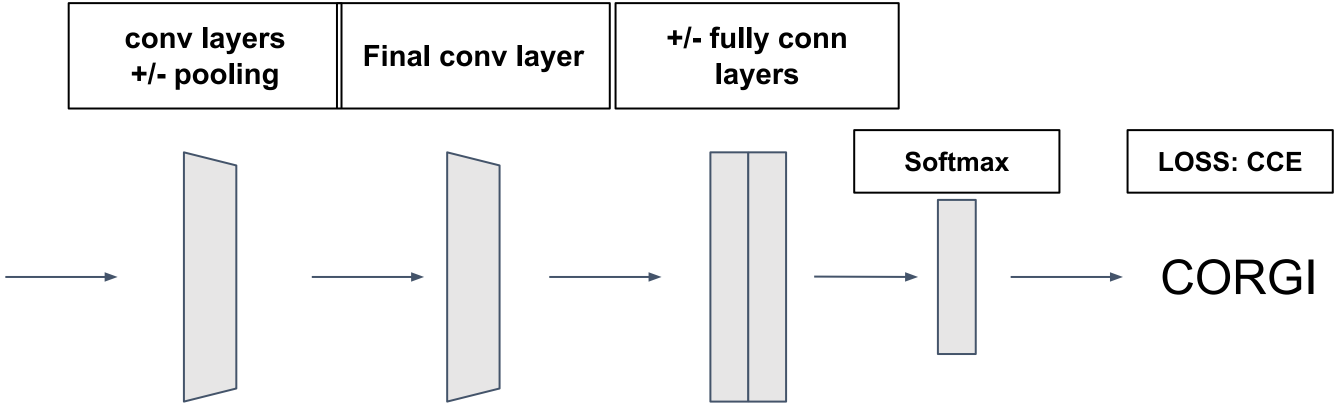
$$QS = \frac{2|X \cap Y|}{|X| + |Y|}$$



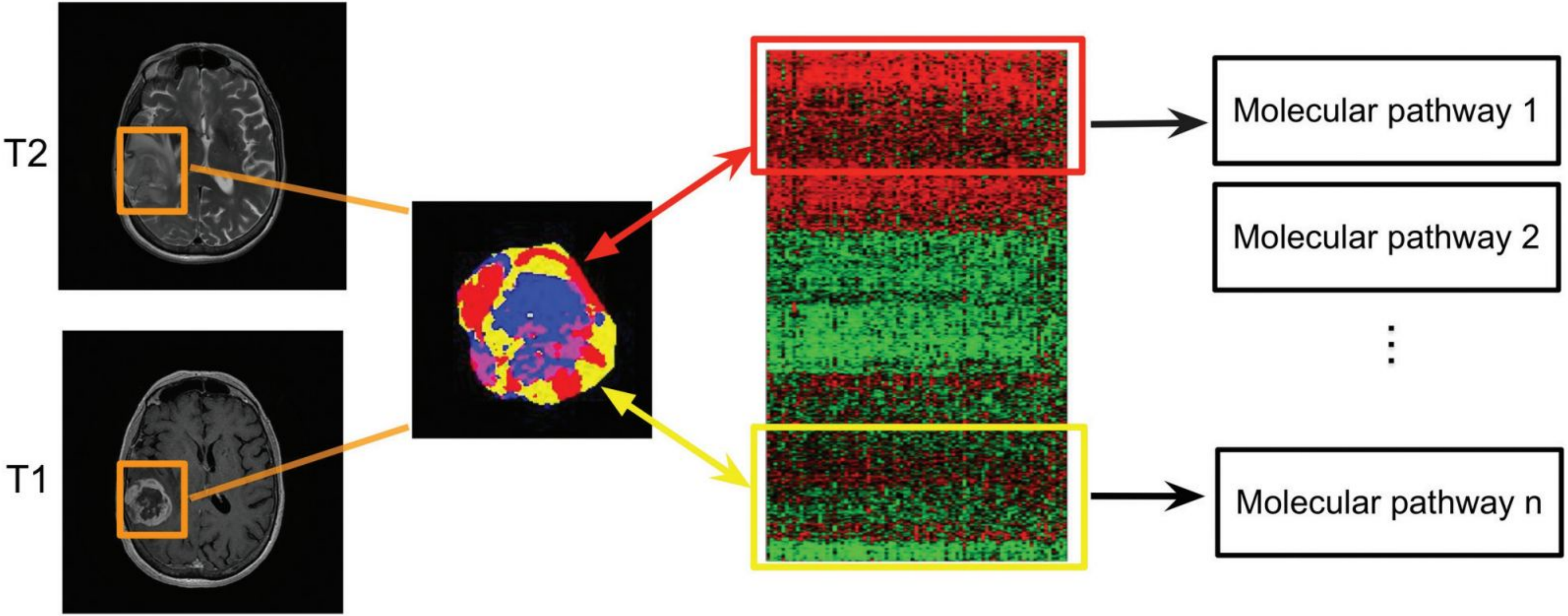
Segmentation and Classification





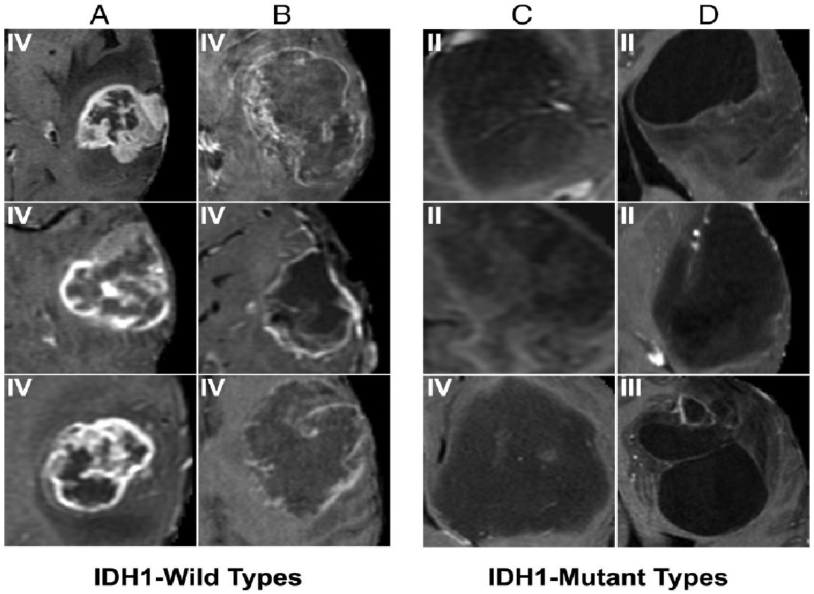


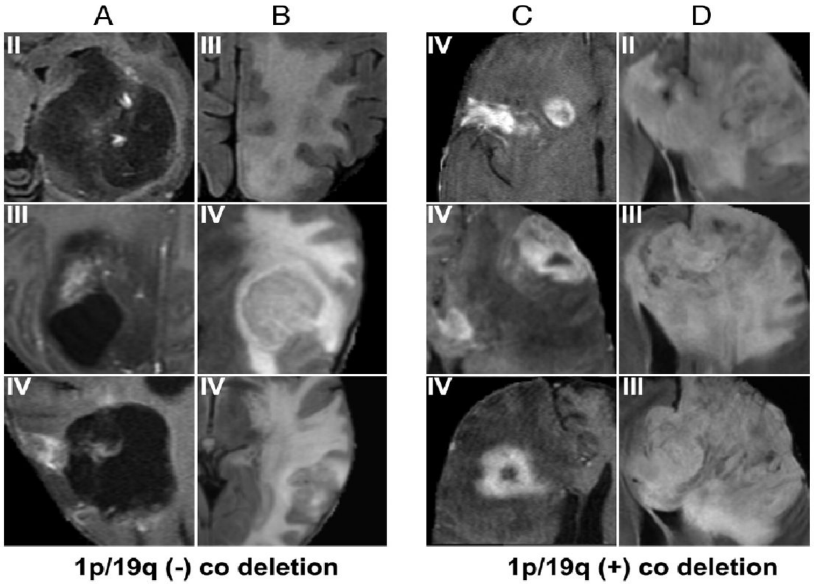
Brain Biopsies



Zhou, M. et al. Radiomics in Brain Tumor: Image Assessment, Quantitative Feature Descriptors and Machine-learning Approaches. AJNR Am. J. Neuroradiol. 39, 208 (2018).

Brain Biopsies





Chang, P. et al. Deep-Learning Convolutional Neural Networks Accurately Classify Genetic Mutations in Gliomas. AJNR Am. J. Neuroradiol. (2018). doi:10.3174/ajnr.A5667

Weak Supervision


TRUST NO ONE

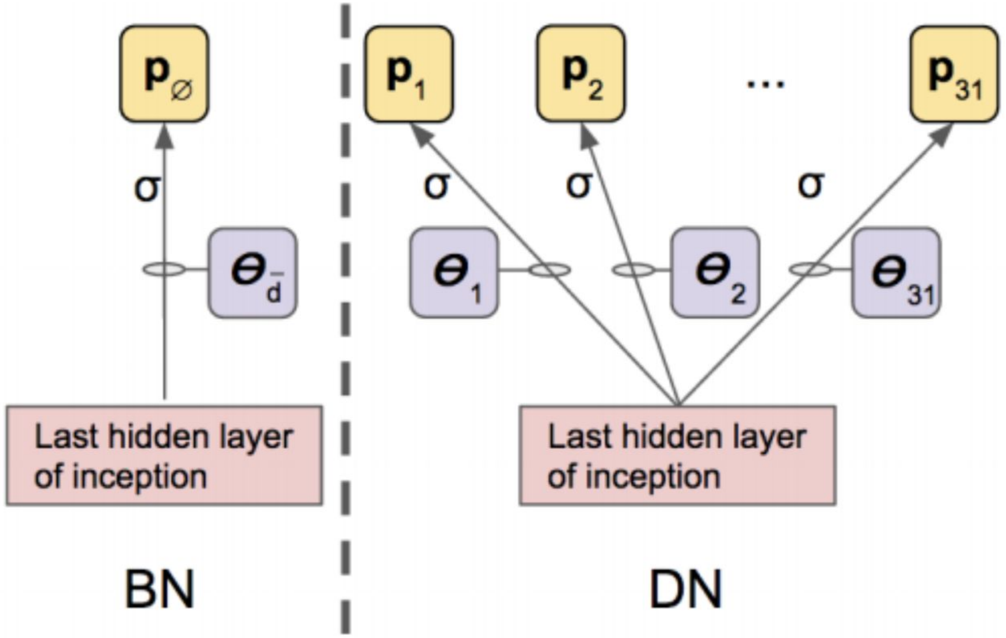
Two Kinds of Labels

Gold Standard Labels
Ground Truth

Silver Standard Labels
Noisy Labels

Are Medical GT Labels Fool's Gold?

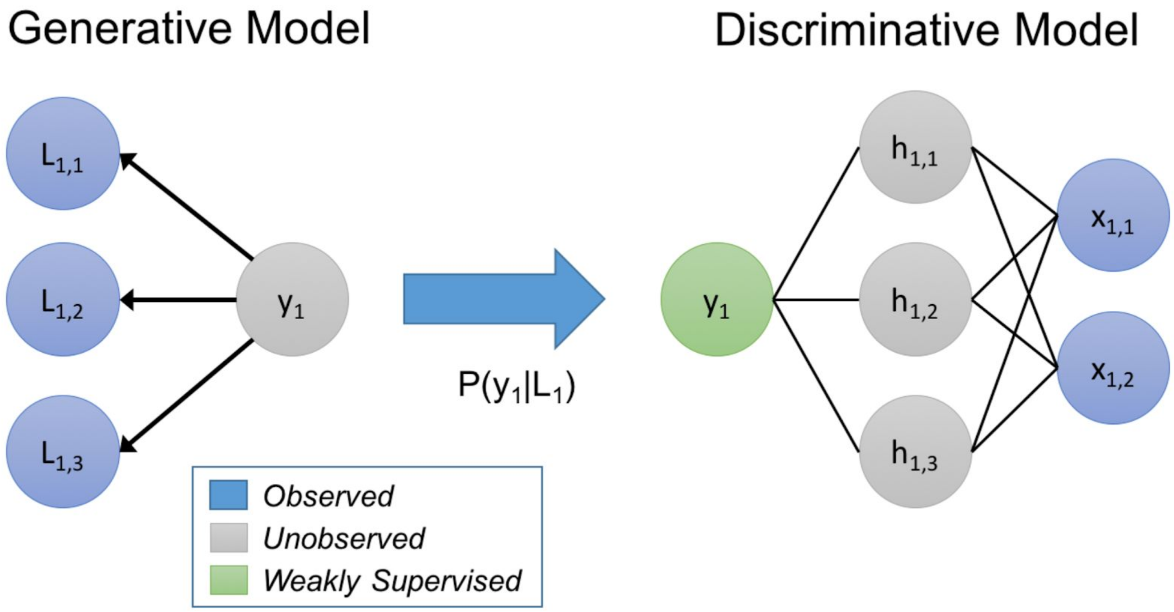
- Medical labels can be challenging with low IRR
  - Google Retinopathy dataset = 55.4%
  - IRR and **70.1% agreement between each expert and her/himself** at a later time point!
- Can average labels using EM.
- However, *average of modeled raters* may outperform *model of average raters*.
- Guan et al. 2017 had 1.97% decrease in test loss



Guan et al. 2017 - Who Said What - Modeling Individual Labelers Improves Classification
Whitehill et al. 2009 - Whose Vote Should Count More - Optimal Integration of Labels from Labelers of Unknown Expertise

Weak Supervision with Generated Silver Labels

Solution? Accept noise in our label set.

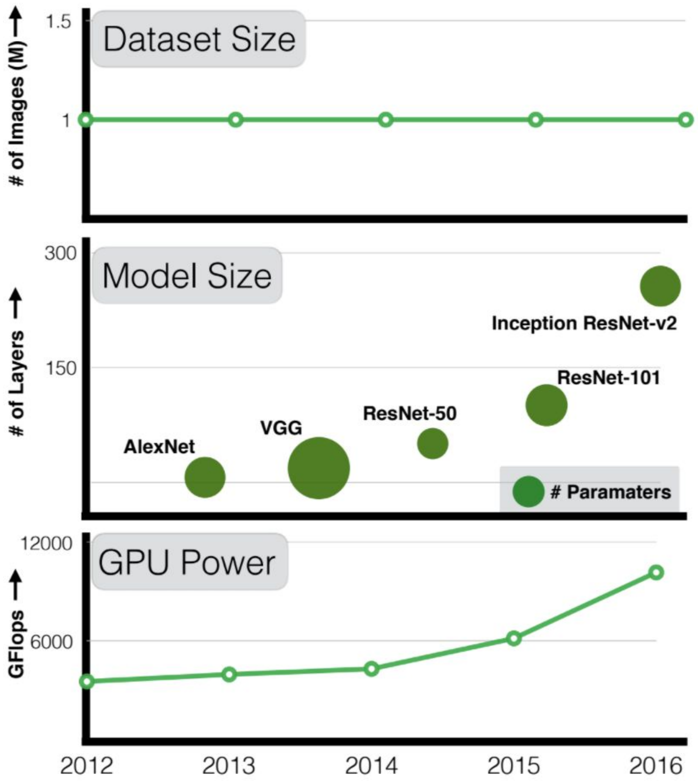


Alex Ratner, Stephen Bach and Chris Ré - Snorkel Blog

The Unreasonable Effectiveness of Big Data with Silver Labels

But does this work? Consider the following trends in computer vision with ImageNet….

What if we had a dataset 300x ImageNet's size with noisy labels?



C Sun, et al. Revisiting Unreasonable Effectiveness of Data in Deep Learning Era - arXiv 2017

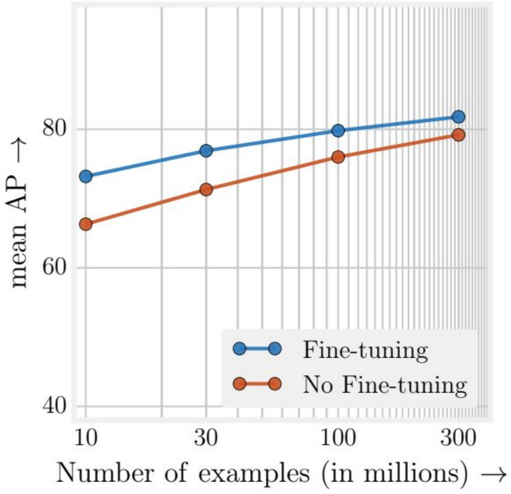
The Unreasonable Effectiveness of Big Data

Effect of pre-training ResNet-101 on JFT-300M's silver labels

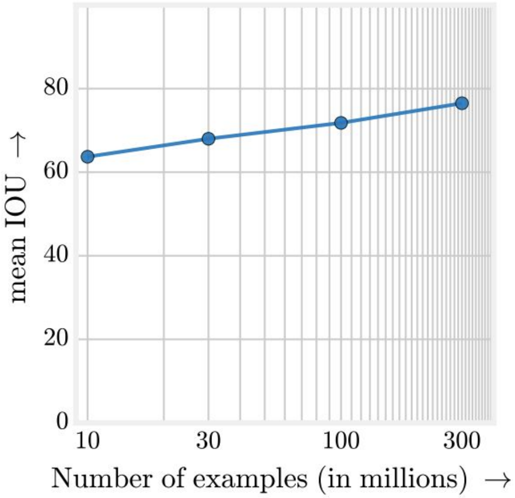
Initialization	Top-1 Acc.	Top-5 Acc.
MSRA checkpoint [16]	76.4	92.9
Random initialization	77.5	93.9
Fine-tune from JFT-300M	**79.2**	**94.7**

Classification on ImageNet 'val' set

Object detection on PASCAL-VOC Test set



Semantic segmentation on PASCAL-VOC Test set



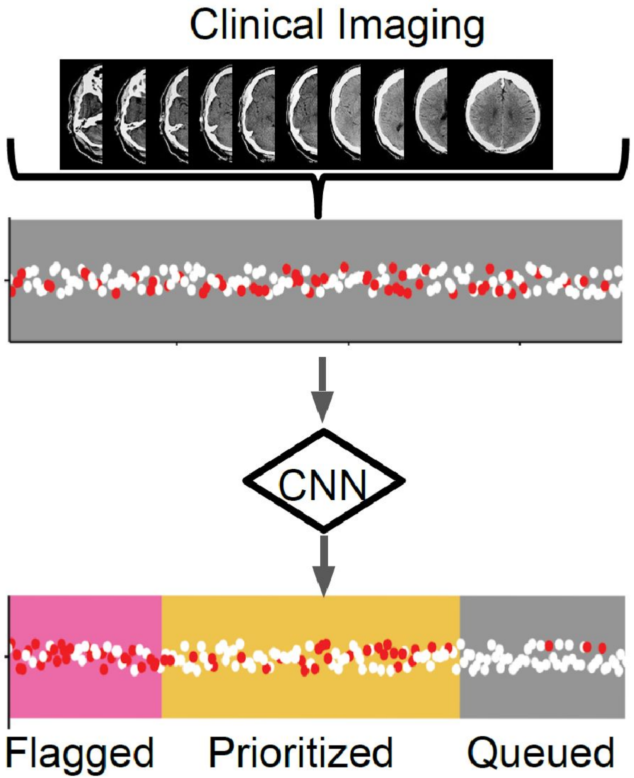
initialization	mIOU
ImageNet	73.6
300M	75.3
ImageNet+300M	**76.5**

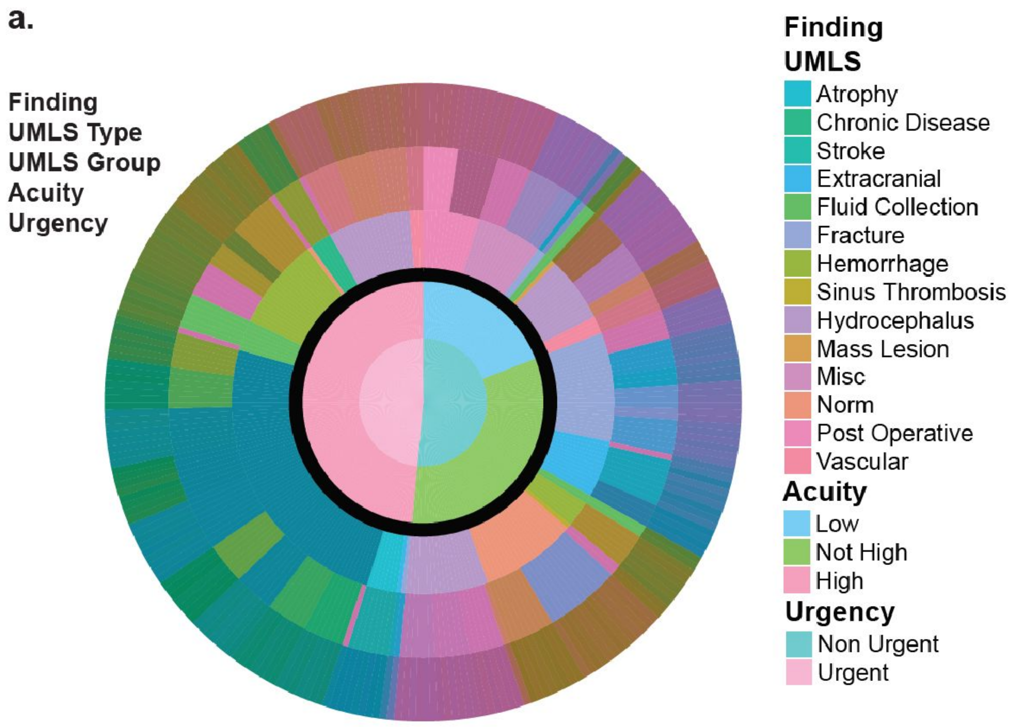
C Sun, et al. Revisiting Unreasonable Effectiveness of Data in Deep Learning Era - arXiv 2017

Application to Acute Neurologic Events

Titano, J. J. et al. Automated deep-neural-network surveillance of cranial images for acute neurologic events. Nat. Med. (2018). doi:10.1038/s41591-018-0147-y

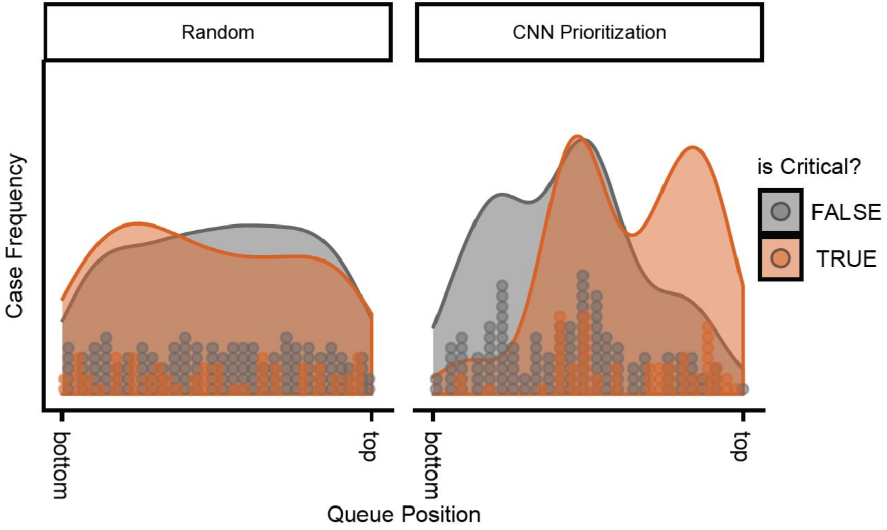
Faster Interpretation of Imaging

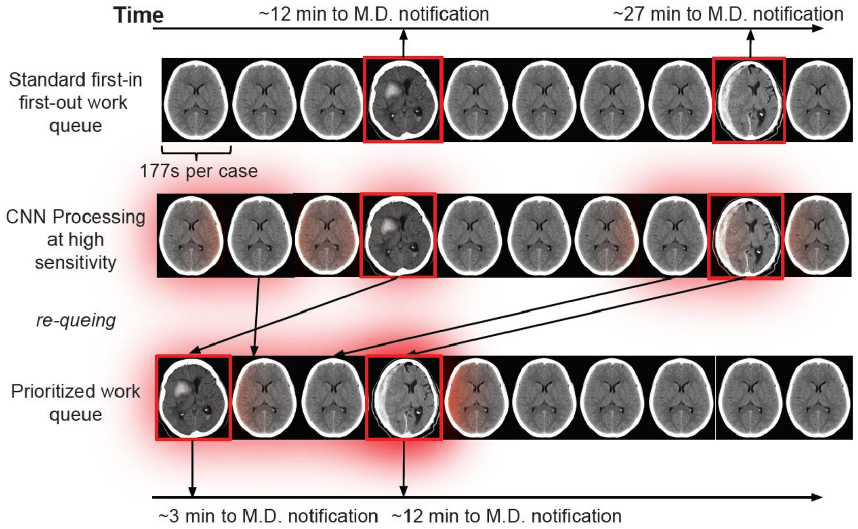




Titano, J. J. et al. Automated deep-neural-network surveillance of cranial images for acute neurologic events. Nat. Med. (2018). doi:10.1038/s41591-018-0147-y

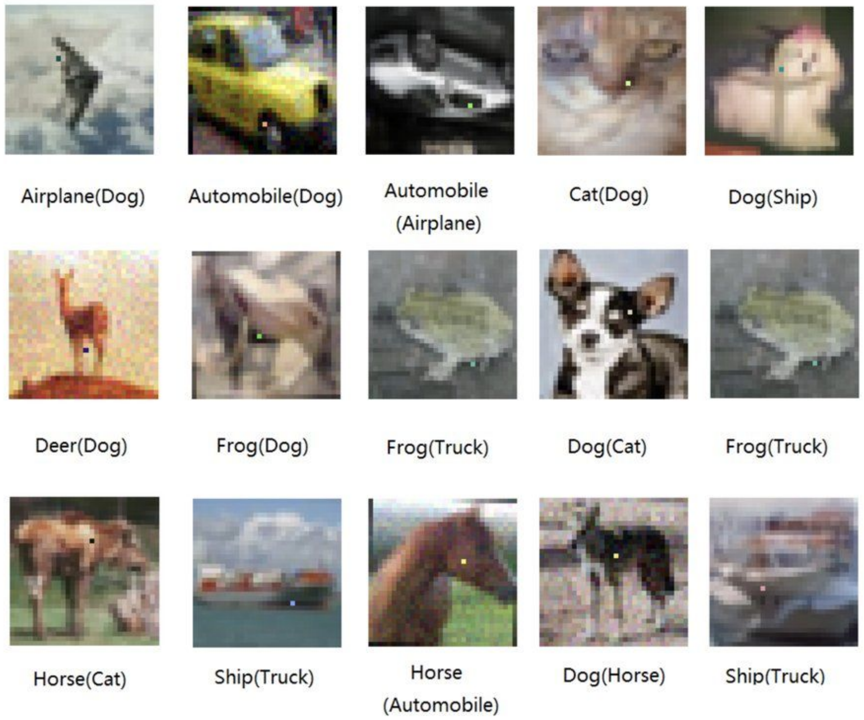
Faster Interpretation of Imaging

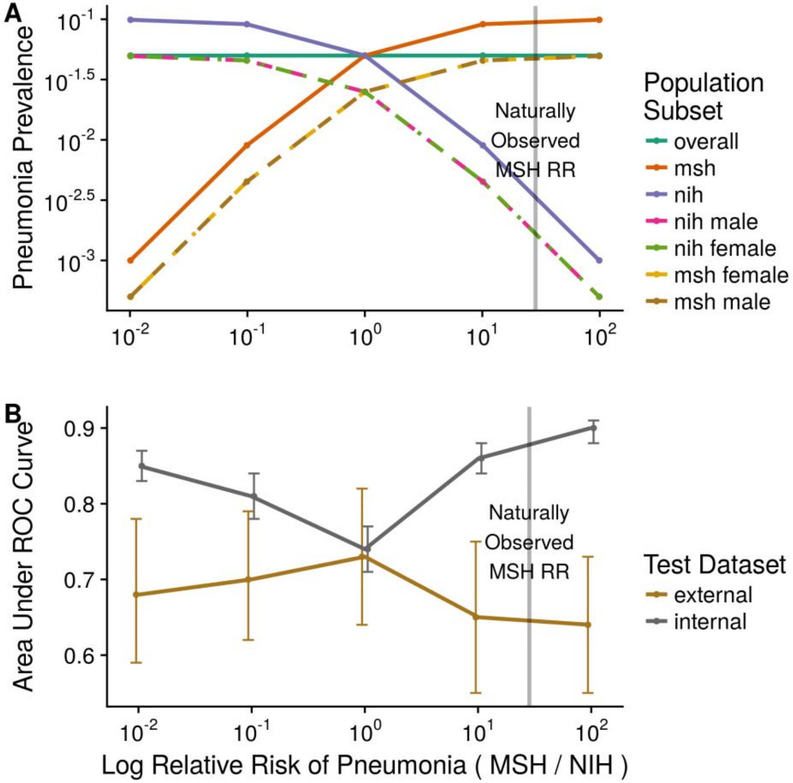




Titano, J. J. et al. Automated deep-neural-network surveillance of cranial images for acute neurologic events. Nat. Med. (2018). doi:10.1038/s41591-018-0147-y

Disclaimer #1: Generalization of deep models is not guaranteed





Zhang, C., Bengio, S., Hardt, M., Recht, B. & Vinyals, O. Understanding deep learning requires rethinking generalization. arXiv [cs.LG] (2016).

Disclaimer #2: Weak Classifiers are Easily Distracted

ResNet-50

('bucket', 0.43788964),
('tub', 0.13390972),
('caldron', 0.11801116)

SSD-300

Average Precision (AP) @[IoU=0.50:0.95 | area= all | maxDets=100] = 0.900
Average Precision (AP) @[IoU=0.50 | area= all | maxDets=100] = 1.000
Average Precision (AP) @[IoU=0.75 | area= all | maxDets=100] = 1.000

Disclaimer #2: Weak Classifiers are Easily Distracted

ResNet-50

	Ground true with people	Ground true without people
Prediction with people	1540	767
Prediction without people	277	2416

Classification Accuracy: 0.7912

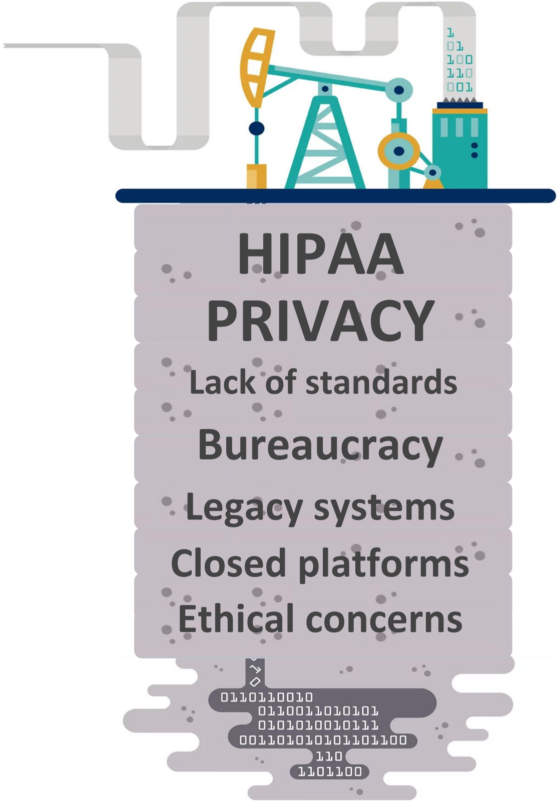
SSD-300

	Ground true with people	Ground true without people
Prediction with people	2236	31
Prediction without people	457	2276

Detection Accuracy: 0.9024

AISINAI NCTS

Disclaimer #3: Data is Everything



Disclaimer #4: Medical Data Paid for in Human Lives

We are going to need more training data...

MSHS MEDICAL A.I. CONSORTIUM

AISINAI

Neurological Surgery

Eric Karl Oermann, MD
Instructor, Department of Neurosurgery

Anthony Costa, PhD
Assistant Professor, Department of Neurosurgery
Director, Sinai BioDesign

Joshua B Bederson, MD
Chairman, Department of Neurosurgery

Holly Oemke, BA
Program Manager, Sinai Biodesign

Margaret Pain, MD
Houseofficer, Department of Neurosurgery

Raj Shrivastava, MD
Associate Professor, Department of Neurosurgery

John Caridi, MD
Assistant Professor, Department of Neurosurgery

Neha Dangayach, MD
Assistant Professor, Department of Neurosurgery
Research co-director for ICCM

Radiology

Joseph Titano, MD
Fellow, Department of Radiology

Javin Schefflein, MD
Houseofficer, Department of Radiology

Burton Drayer, MD
Chairman, Department of Radiology

Brett Marinelli, MD
Houseofficer, Department of Radiology

Nathaniel Swinburne, MD
Houseofficer, Department of Radiology

Andres Su, MD
Houseofficer, Department of Radiology

Michael Cai, MD
Houseofficer, Department of Radiology

Orthopedics

Samuel Cho, MD
Associate Professor, Department of Orthopedics and Neurosurgery

Jun Kim, MD
Houseofficer, Department of Orthopedic Surgery

Komal Srivastava, BA
Grant Specialist, Department of Orthopedic Surgery

Radiation Oncology

Sonam Sharma, MD
Assistant Professor, Radiation Oncology

ISMMS

Fred Kwon, MSE
MD/PhD student

Martin Kang, BS
Medical Student

Deepak Kaji, BS
MD/PhD student

Varun Arvind, BS
MD/PhD student

COLLABORATORS:

Stanford MEDICINE

Alice Fan, MD
Assistant Professor of Oncology

Viola Chen, MD
Fellow, Department of Oncology

Icahn School of Medicine at Mount Sinai — *Translational and Molecular Imaging Institute*

Zahi Fayad, PhD
Director MSHS TMIII
David Mendelson, MD
Director of Informatics

MERCK

Joseph Lehar, PhD
Director of Computational Biology

Hammerlab:
Alex Rubinsteyn, PhD
Postdoc, GGS

intel

Peter Tang, PhD
Senior Fellow

Google

Marcus Badgeley, MEng
PhD student, Google / Verily
Medical student, ISMMS