MACHINE LEARNING APPROACHES FOR MUSICAL & MEDICAL ACOUSTICS MEASUREMENTS

DR. SCOTT H. HAWLEY, @DRSCOTTHAWLEY (GITHUB, TWITTER)

BELMONT UNIVERSITY

CSM FACULTY RESEARCH SEMINAR *AND* MA/ACM, OCT 6 2017

Acknowledgement: Thanks to Drs. Robert Magruder & Thom Spence for supporting this research

"MACHINE LEARNING"

- A sub-field of Artificial Intelligence (AI)
- "3 waves of AI" (John Launchbury, DARPA):
  1. Handcrafted Knowledge: Human expertise "programmed in". Expert Systems, e.g., TurboTax. Don't learn.
  2. Statistical Learning: stat. inference, driven by optimization. "Machine Learning" (ML) for this talk. No context or 'reasoning'. Curve-fitting, clustering, (learned) tree-traversal…
  3. Contextual Adaptation: Contextual models, Perception, & Reason. (Not there yet.)
- "Deep Learning": buzzword coined by Geoffrey Hinton for multi-layer neural networks, as a 're-branding' of NN to escape longstanding stigma

MACHINE LEARNING ISN'T NEW

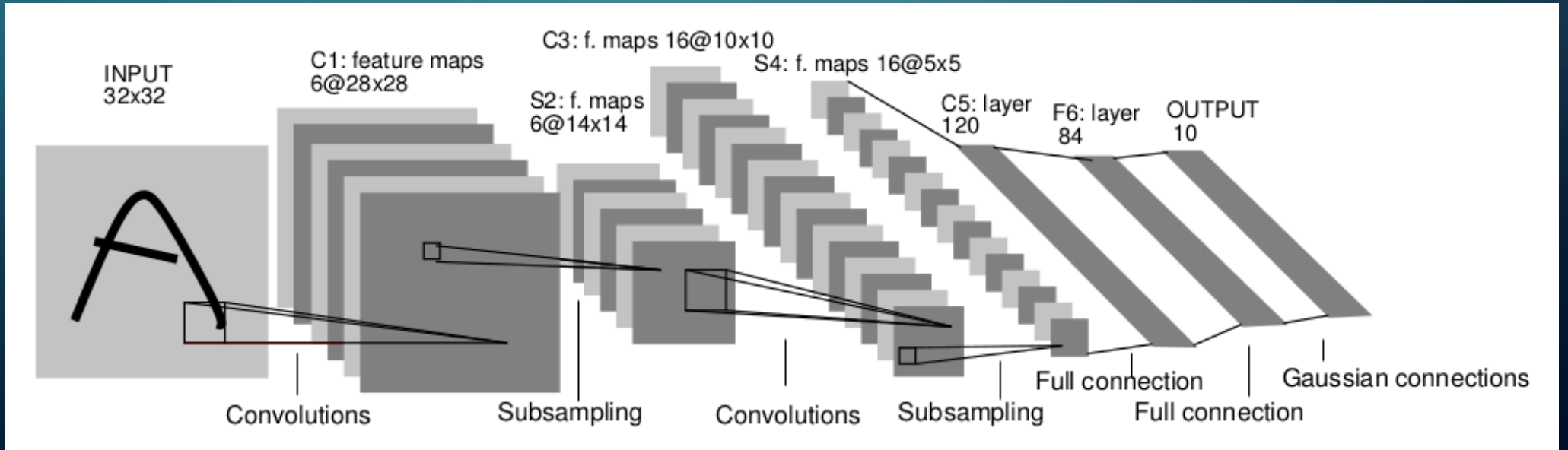
- Most methods & algorithms have been around a long time
  - Artificial Neural Networks have been studied since the 1940's (McCulloch & Pitts, 1943)
  - Logistic regression has been around since the 1800's (Verhulst, 1838)
  - Sophisticated curve-fitting, statistical analysis & predictive modelling practiced in science & industry for several decades, but weren't typically called "Machine Learning" (term coined by Arthur Samuel in 1959 but not in wide use).
  - Backpropagation algorithm for NN's by Werbos in 1974
  - …etc

WHY'S MACHINE LEARNING 'SUDDENLY' HOT?

- What's new is ***scale***:
  - Yes, key improvements in algorithms have occurred, however it's also…
  - Lots of data available to train on, thanks to the **Internet**
  - Availability of fast processors: **Graphics Processing Units (GPUs)**
- What's new is ***success***:
  - To those involved in the research, it's been a long, gradual road, but to the public, it can appear as though ML and NN's have "burst" on the scene, due to success in solving complex problems:
  - Lecun, Bengio et al showed Convolutional Neural Networks could give *state of the art* performance on image classification, speech recognition, and time series prediction --- using ***the same architecture!***

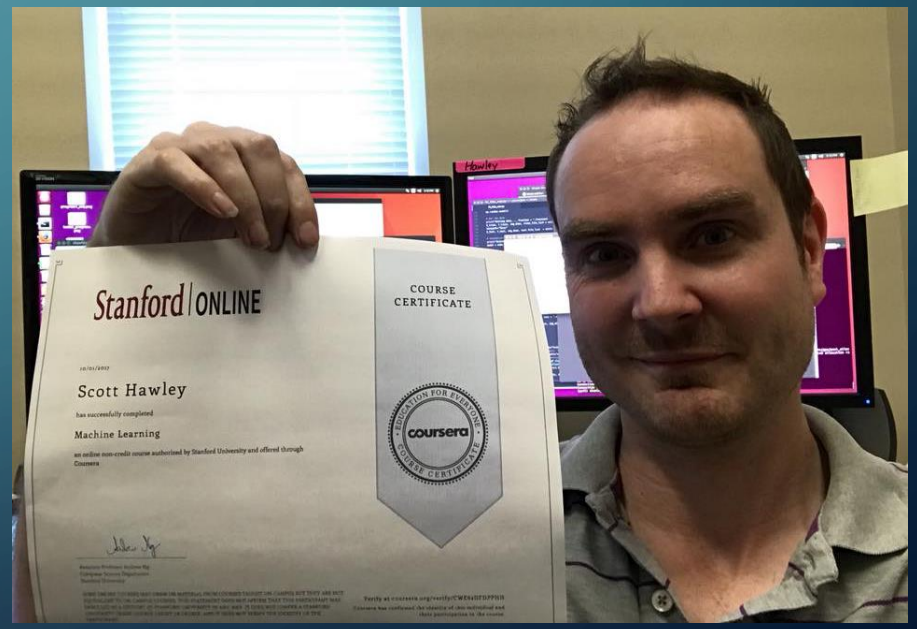
QUICKIE CONV-NET DEPICTION

- Based on visual cortex (of cats, but also humans..)
- Multiple ‘feature extractors’ with limited visual fields that 'sweep’ across images
- Successively smaller patches, increasing hierarchy
- Great for classifying images, including spectrograms of audio files!
- Can be tricky to train



FOR MORE AI / ML BACKGROUND…

- Take Dr. Hooper's A.I. class!
- Online tutorials, e.g. deeplearningandneuralnetworks.com, machinelearningmastery.com, and just Google what you want to know
- Andrew Ng's free ML course on Coursera.org:



Oct 2, 2017

MEDICAL ACOUSTICS: HEART/LUNG DIAGNOSIS

- Two-time finalists for AIGrant.org (Spring, Fall 2017) YouTube video pitch…

- Uses my CNN classifier code: Panotti

"Infrastructure for Community Construction and Moderation of Media Datasets for Human and A.I. Training"

Scott Hawley, Tamara Baird, Frank Baird

Grassroots approach to audio dataset creation

Moving forward: In talks w/ Belmont Nursing, IRB…

MUSICAL ACOUSTICS: STEELPAN OSCILLATIONS

- Met Dr. Andrew Morrison at ASA (Hawaii) & ISMA (Montreal). He runs… https://www.zooniverse.org/projects/achmorrison/steelpan-vibrations
- Crowdsourcing humans to annotate high-speed laser-interferograms of Caribbean steelpan drums. YouTube Video: http://bit.ly/2yqjxU5

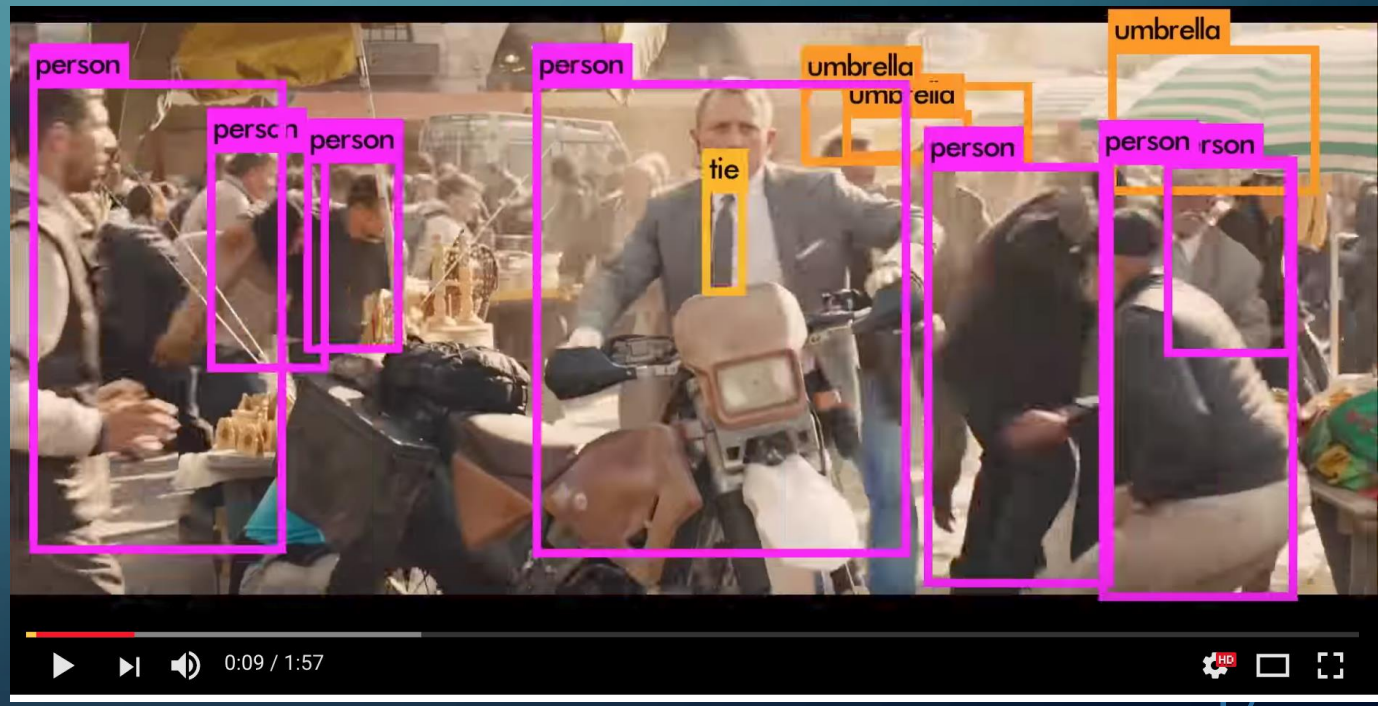
The task: -- Try it yourself & help out!

- How many antinodes in each image?
- Draw ellipse around each antinode
- Count the fringes/rings in each AN

- My big mouth: "If you amass enough labeled data, one could use it to train a ML algorithm to do that"

ML PROBLEM IS CALLED "OBJECT DETECTION"

- Beyond mere image classification, we *segment* an image get 'bounding boxes' where different objects are. Example: YOLOv2 video
- Strategy:
  - Take a pre-trained image classifier, and feed it different 'patches' of the image
  - Patches where classifier outputs high probability matches = where your objects are
- Doing it *fast* is tricky



EXCEPT… THIS PROBLEM'S A BIT DIFFERENT

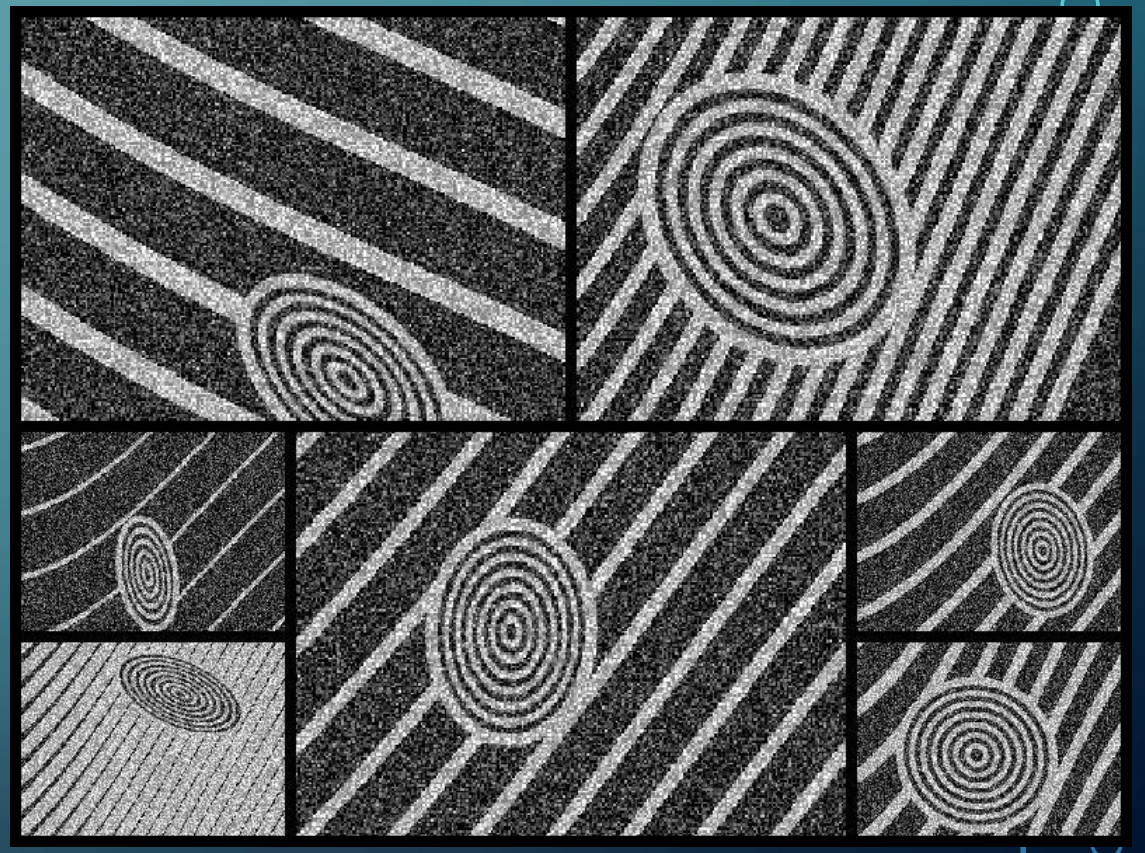
- We want not just 'serviceable' bounding boxes, but *precise* centroids of steelpan oscillation antinodes – this is a *regression* problem.
- Counting rings could be regarded as either classification or regression
- Not classifying typical categories – not people vs. cars, rather 'science' images
- Training one's own Convolutional Neural Network (CNN) is tricky

(ONE) SOLUTION

- Use a CNN 'pre-trained' on ImageNet, but re-train using steelpan images
- Try regression, predicting:
  - To begin: Center X, Center Y, # Rings (Ring-counting could also be classification)
  - Later: full ellipses: semimajor & semiminor axes, angle
- Start with 'fake data': Just to get started, rather than use Morrison's 'real' images, I generated fake ones via Python & OpenCV…

FAKE TRAINING DATA, PRE-TRAINED NETWORK

- Made 50,000 images w/ elliptical rings, one per image
- Added noise &‘wavy’ backgrounds
- Wrote code in Python via Keras lib.
- Tried pretrained CNNs “InceptionV3”, “MobileNet”, and “Inception-ResNetV2”
- Ran on fast computers…



COMPUTERS!

- Chem & Physics has two machines for machine learning

Fall 2016:

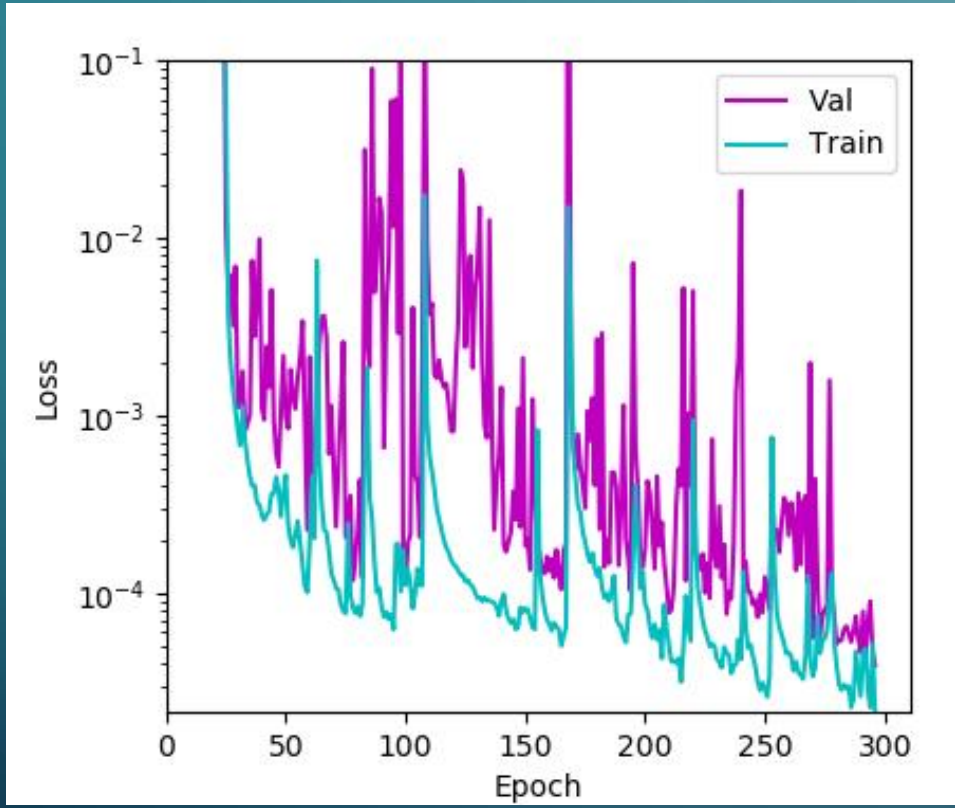
- GTX 1080 GPU (2560 CUDA cores, 8GB VRAM)
- 32 GB RAM
- Liquid-Cooled i7-6700 4-core 3.4 GHz
- Samsung 950 Pro SSD
- Used for ML & VR (HTC Vive)

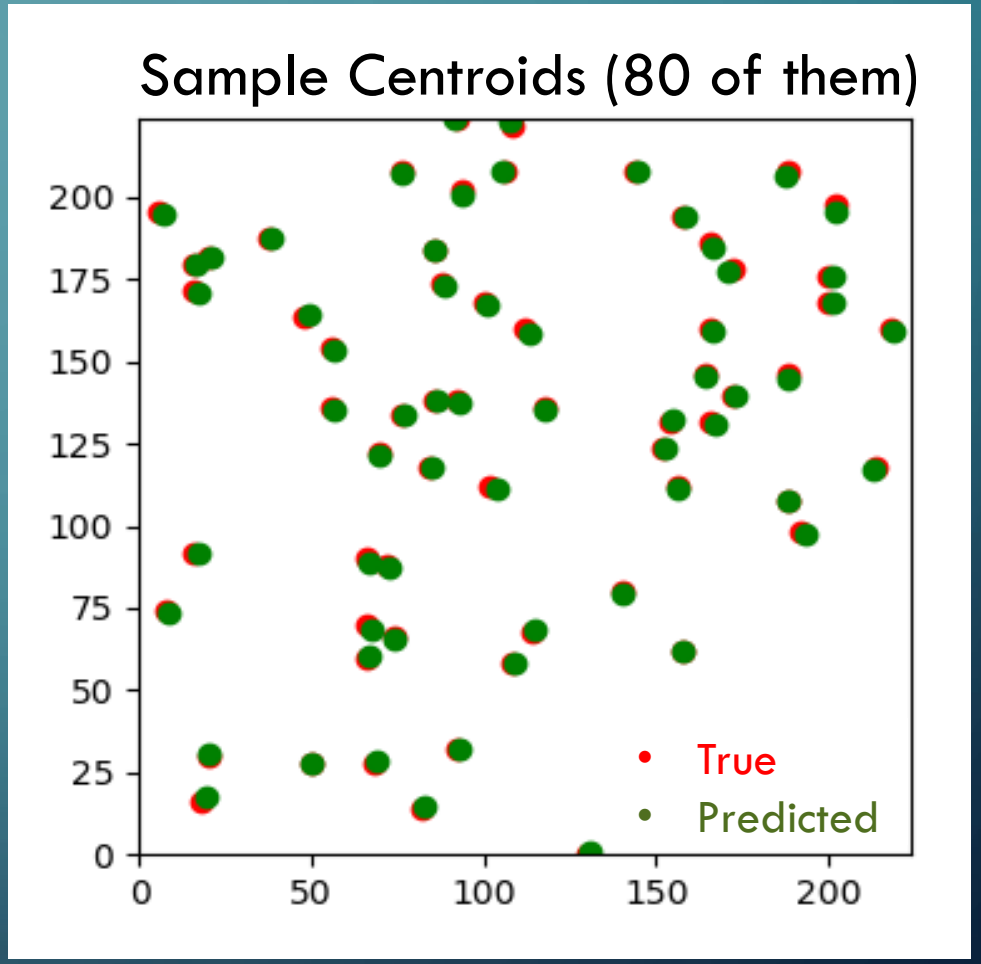
April 2017:

- 2 Titan X GPUs (3072 CUDA cores each, 12GB VRAM)
- 64 GB RAM
- Liquid-Cooled Xeon 6-core 3.6 GHz
- Samsung 950 Pro SSD
- Linux only ;-)

~30 Teraflops total between two machines!

RESULTS SO FAR…





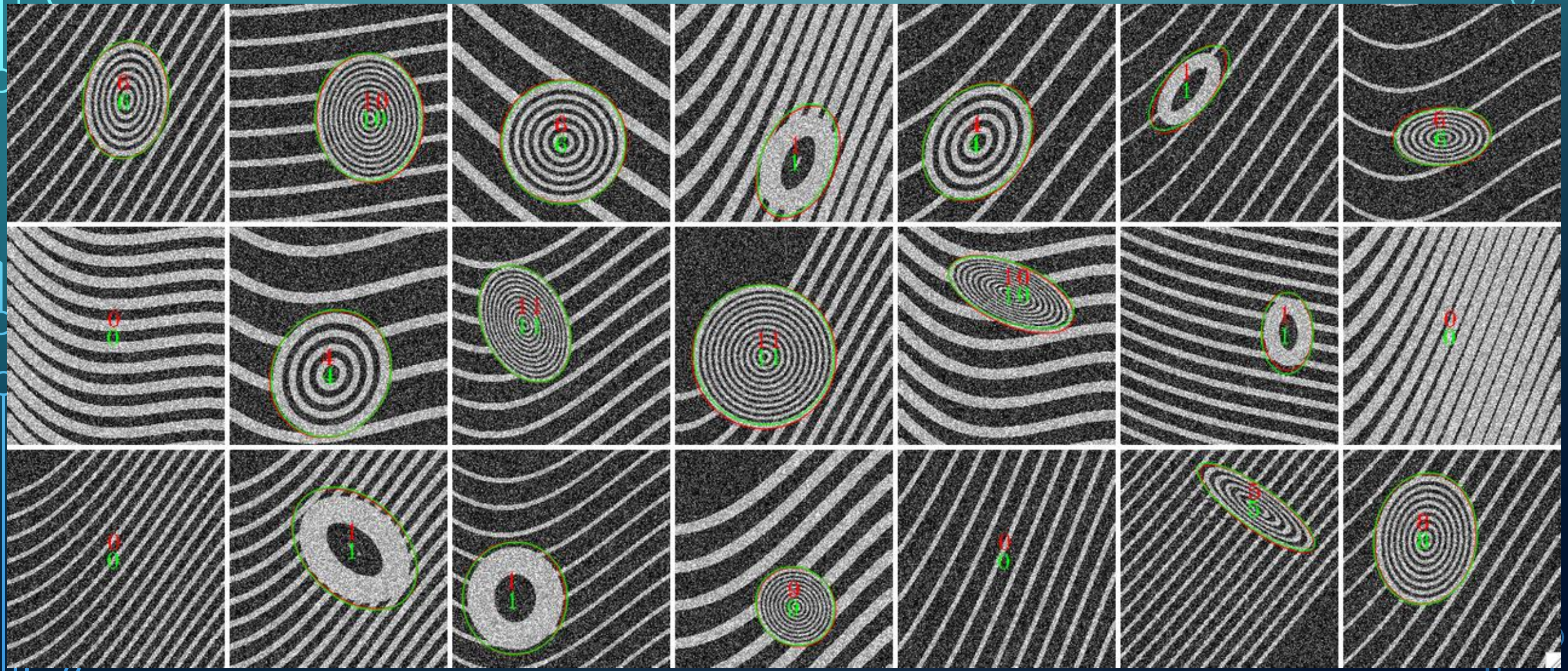
- Accuracy levels off for > 30,000 images. Validation & Test metrics are comparable ☺
- Results depend on which CNN is used: MobileNet is *fast*, IRv2 more *accurate* but slow
- Best Ring-counting accuracy: 99.8%, Mean pixel error of centroid locations: < 1 Pixel!

RESULTS CONT'D

Red = True

Green = Predicted

#s = Count # of rings in antinode (0 - 11)



WHAT'S NEXT FOR…"SPNet"?

Get it? ESPI = "SP" = SteelPan?

- So far the code is limited: One antinode per image
- Multiple antinodes per image:
  - Implement localization algorithm (YOLO or SSD)
  - Might need to 'reframe' fringe-counting task from regression to classification
- Switch from fake to real data
  - Morrison's got 11,000 labelled images.

ALSO NEXT! SAME CODE: MEDICAL IMAGES!

Sept 27, 2017
News Release:

This is Object Detection!

NIH National Institutes of Health
Turning Discovery Into Health

Search

Health Information | Grants & Funding | News & Events | Research & Training | In

Home » News & Events » News Releases

NEWS RELEASES

Media Advisory Wednesday, September 27, 2017

NIH Clinical Center provides one of the largest publicly available chest x-ray datasets to scientific community

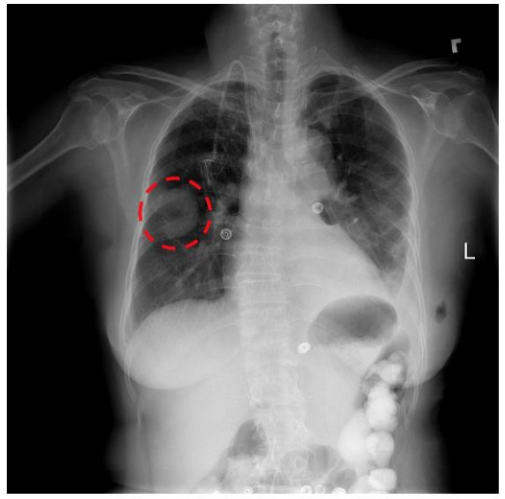
The dataset of scans is from more than 30,000 patients, including many with advanced lung disease.

AddThis

What

The NIH Clinical Center recently released over 100,000 anonymized chest x-ray images and their corresponding data to the scientific community. The release will allow researchers across the country and around the world to freely access the datasets and increase their ability to teach computers how to detect and diagnose disease. Ultimately, this artificial intelligence mechanism can lead to clinicians making better diagnostic decisions for patients.

NIH compiled the dataset of scans from more than 30,000 patients, including many with advanced lung disease. Patients at the NIH Clinical Center, the nation's largest hospital devoted entirely to clinical research, are partners in



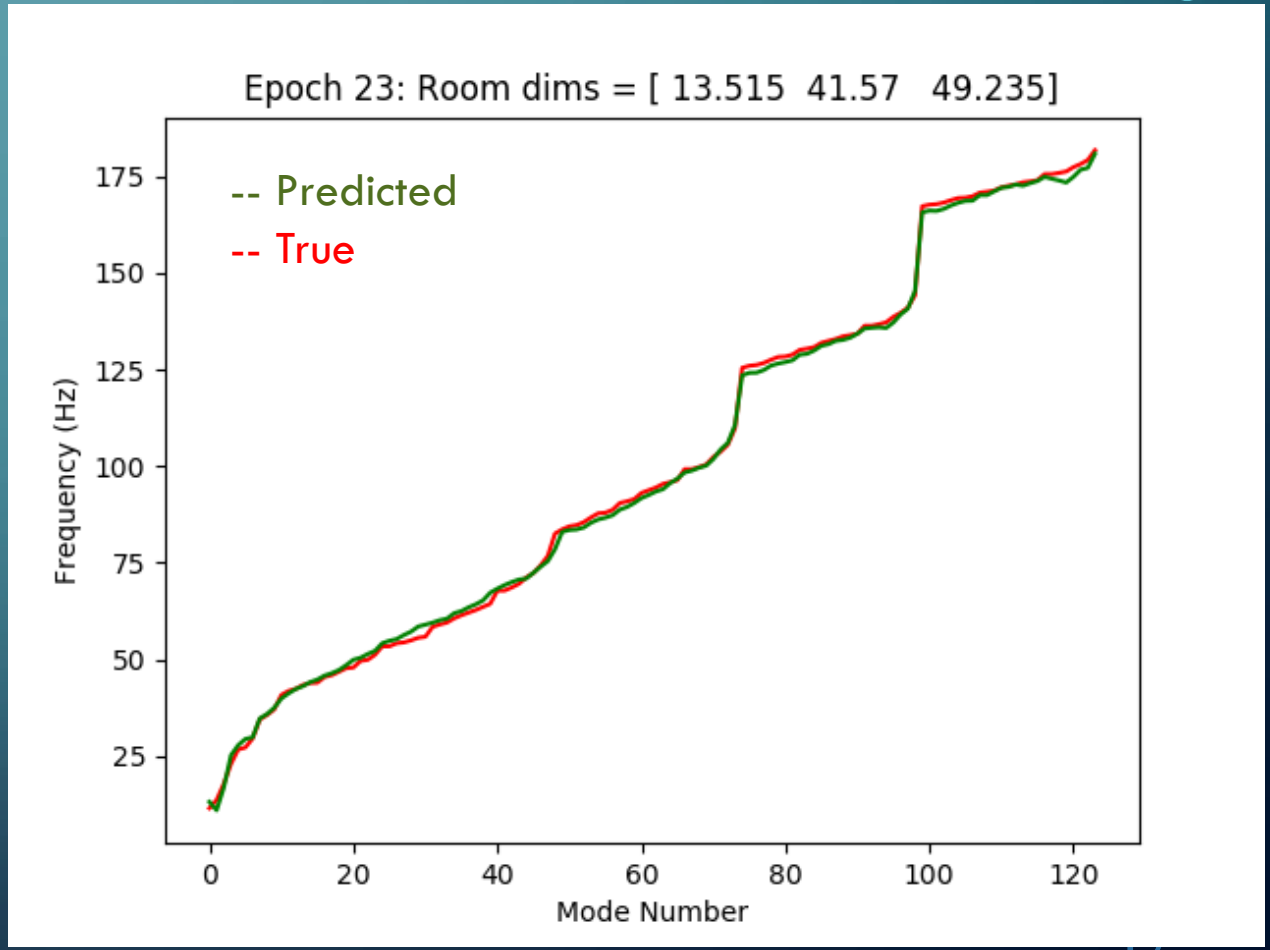
MORE ML-ACOUSTICS: LEARNING ROOM SHAPES

- (Full post & code: https://drscotthawley.github.io/Learning-Room-Shapes)
- During Karlheinz Brandenburg's visit, he said that learning room shapes from the sound of the room is still an open question.
- So"for fun" decided to try: NN system (MLP), learns to use room mode freq's to predict (boxy) room dimensions, *or vice versa*, i.e. it learns the "Rayleigh Equation" for 3D standing waves:

$$f_{nx,ny,nz} = \frac{v_s}{2}\sqrt{\left(\frac{n_x}{L}\right)^2 + \left(\frac{n_y}{W}\right)^2 + \left(\frac{n_z}{H}\right)^2}, \{n_x, n_y, n_z = 0, 1, 2, 3, ...\}$$

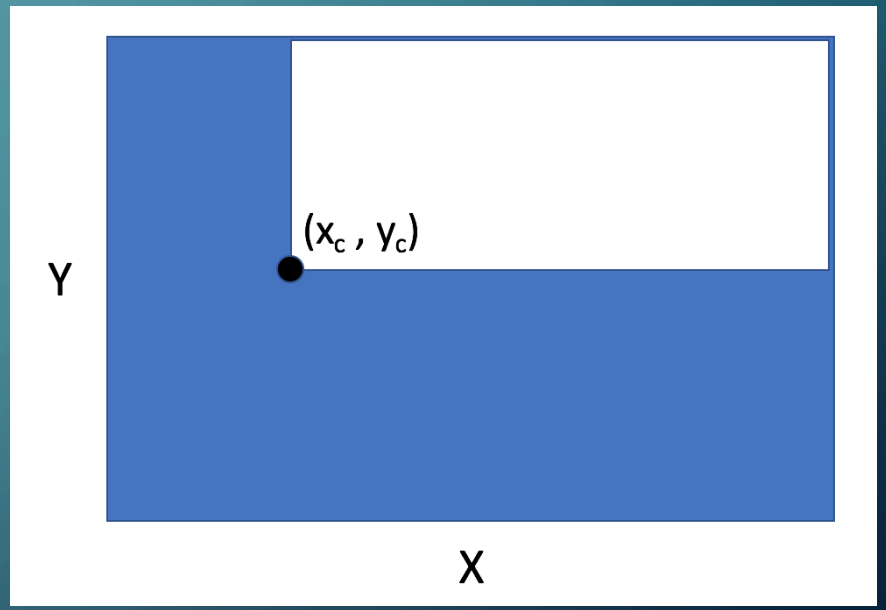
RESULTS

- Code learns to map 'both ways':
  - [f_i's] ⇒ [L, W, H]
  - [L, W, H] ⇒ [f_i's] (Sample at right:)
- If you train it against *sorted* freq's, code learns to *output* sorted freq's!
- 'Helping' the nonlinearity (e.g. via [$1/L^2$, $1/W^2$, $1/H^2$] doesn't improve results (??)



UP NEXT FOR ROOM-SHAPE-GUESSING

- L-Shaped Rooms!
  - NO EQUATION for non-rectangular rooms, need a simulator that does Boundary Element calculations
  - Possible collab via internet, e.g. **Andy Rundquist** (Hamline U., St. Paul)
  - Belmont students??



HOW CAN YOU GET INVOLVED?

- Dr. Hawley has *underutilized* GPU cycles and multiple *fascinating* project ideas! Talk to him!! ☺

  scott.hawley@belmont.edu, @drscotthawley on GitHub & Twitter

- Students & faculty get $200 credit on Amazon Web Services!

  ("But how can I even get started?" LearnPython.org)