Butterfly™

# Deep Learning for Ultrasound Image Analysis

Michal Sofka, PhD

# Safe Ways to Scan the Body

**Stethoscope**

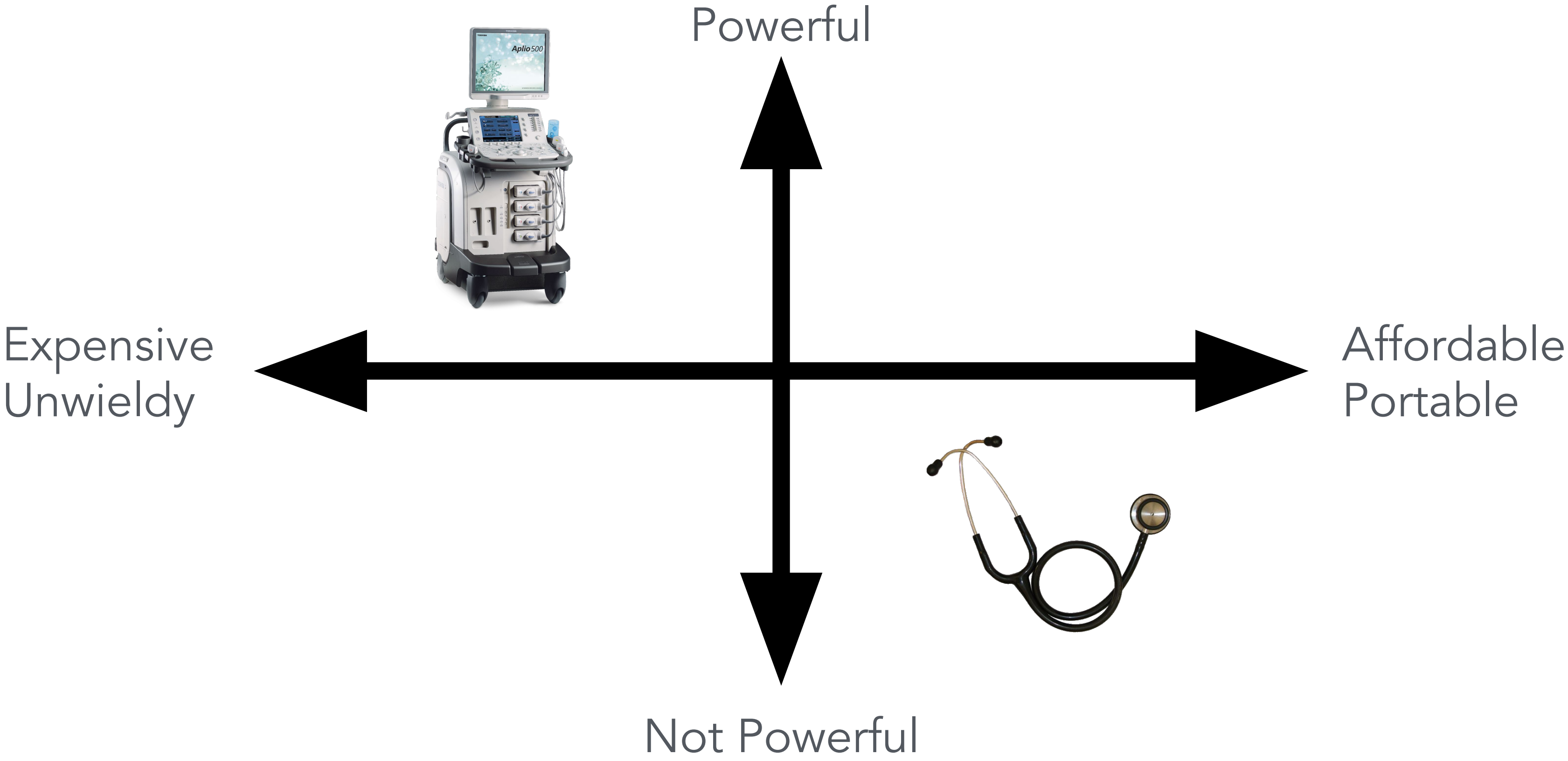
- ✓ Affordable
- ✓ Portable
- ✓ Pervasive
- ✗ Can't see inside the body
- ✗ Limited diagnostic value

# Safe Ways to Scan the Body

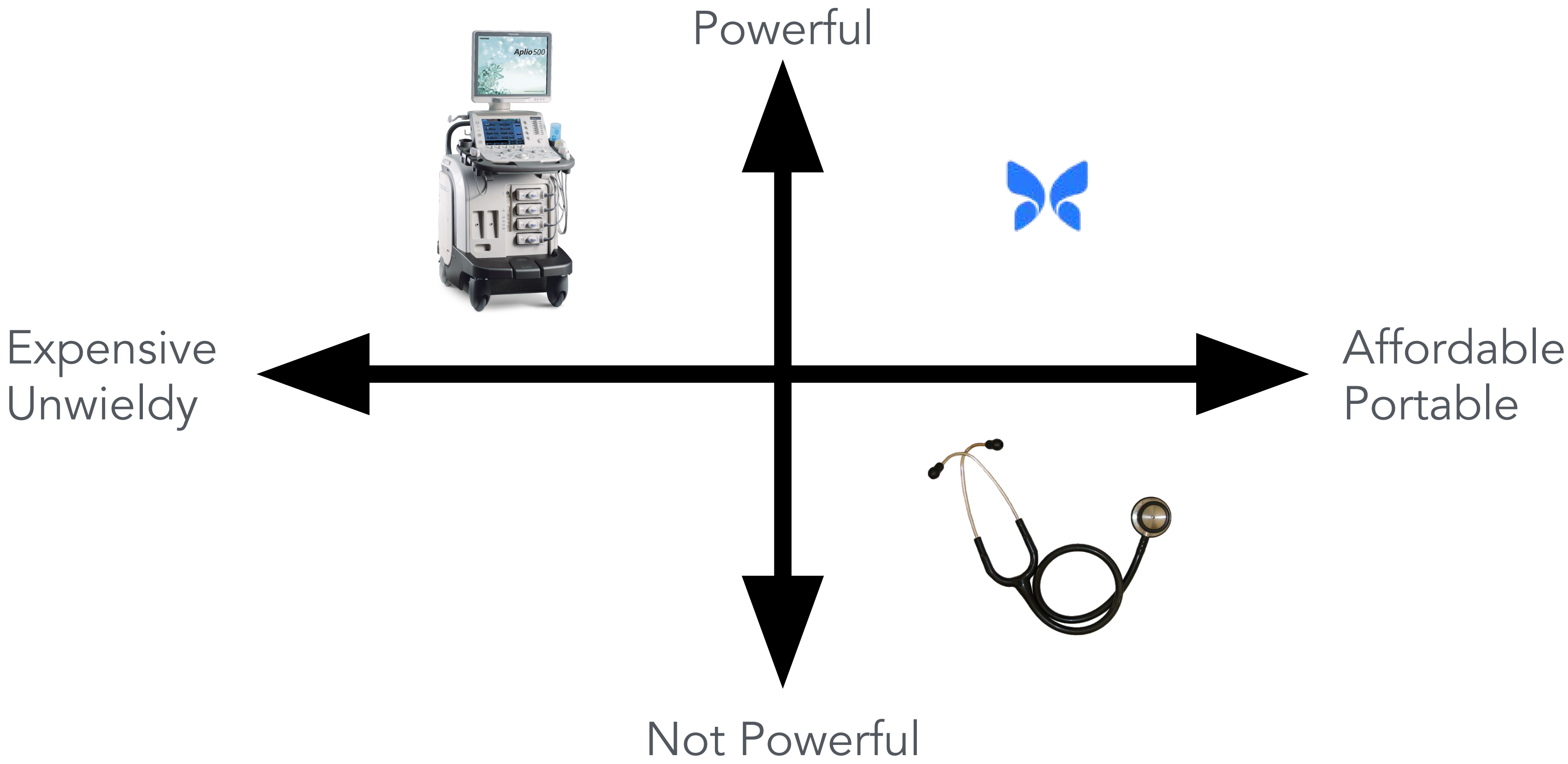
**Traditional Ultrasound**

- ✔ See inside the body in real time
- ✔ Powerful diagnostic
- ✘ Expensive
- ✘ Unwieldy
- ✘ Scarce

# Safe Ways to Scan the Body



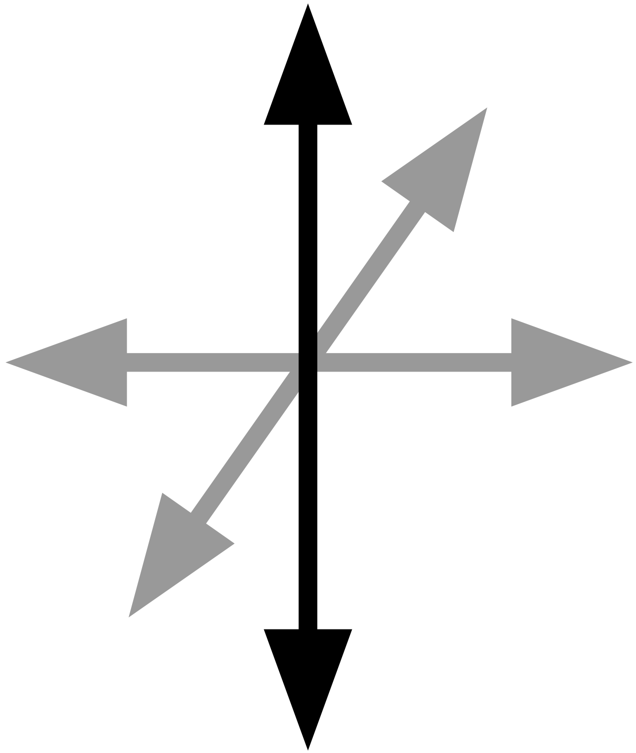
# Safe Ways to Scan the Body



# Usability

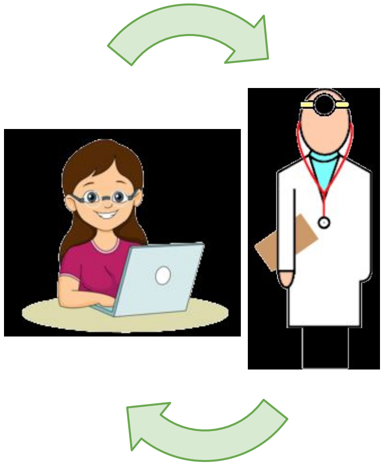
- Ultrasound requires specialized training
  - Radiologists, Sonographers
- Delay between requests and reads can be hours
- Point-of-care US

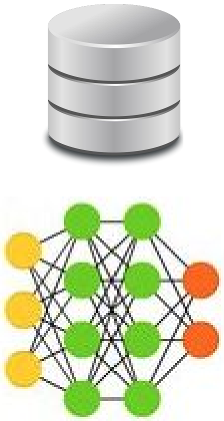
Minimal experience required



Requires years of training

# From Idea to Product





FDA
APPROVED

Great Idea → Data and Models → FDA Approval → Product Release

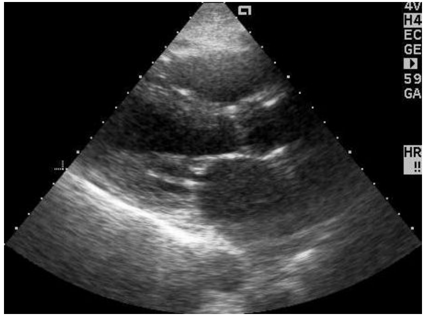
# What does it take to actually get a medical ML system to production?

# Bigger Datasets

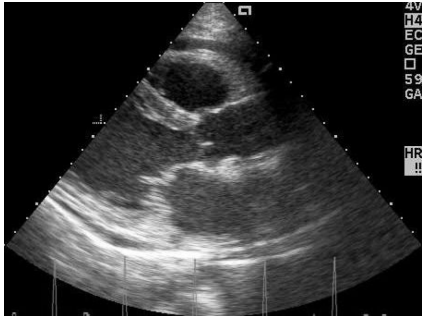
| | |
|---|---|
| MNIST Handwritten digits | 60k training + 10k testing |
| Google House Numbers | 600k training + 30k testing |
| CIFAR-10 | 50k training + 10k testing |
| PASCAL VOC | 11k training in 20 classes |
| ImageNet | 1mm Training in 1000 classes |

Most medical image/voxel datasets have fewer than 300 samples in both training and test

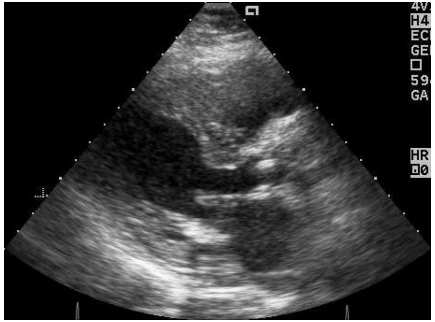
# Does your dataset distribution match the real world?



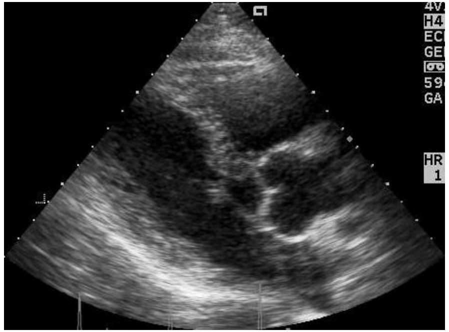
Normal 43%



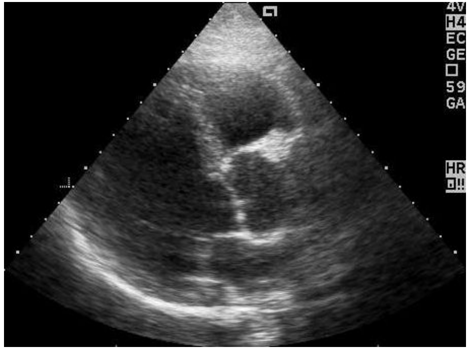
Dilated 13%



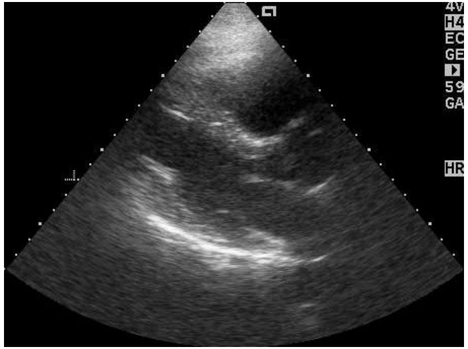
LV hypertrophy 9%



Sigmoid Septum 8%

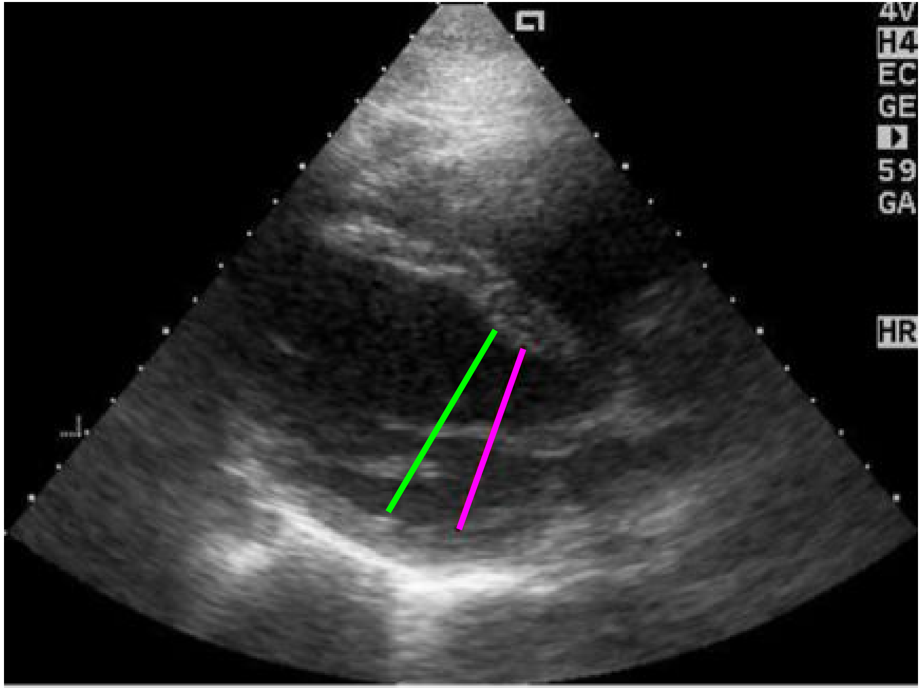


Banana Shaped 6%

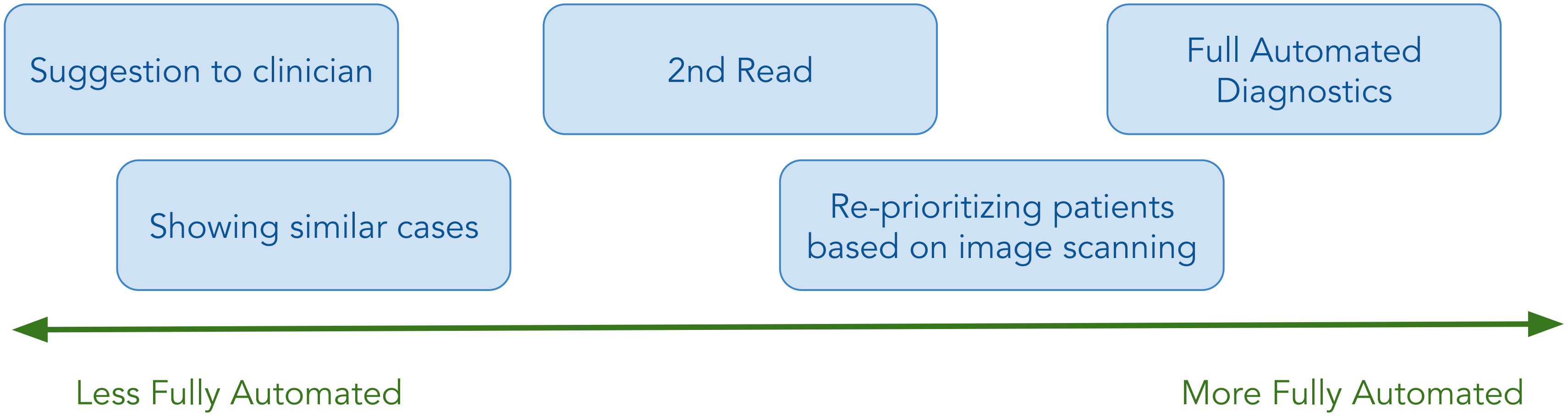


Cavity Obliteration 5%

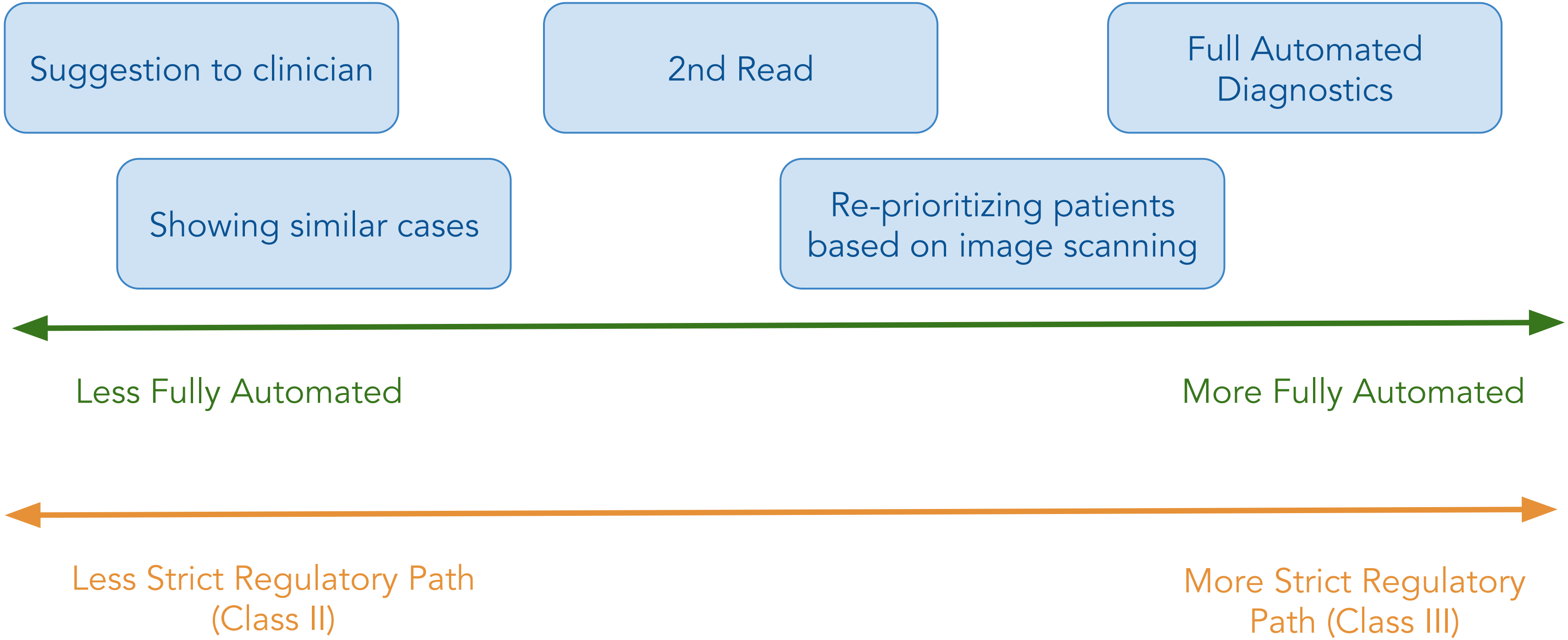
# Annotator variability



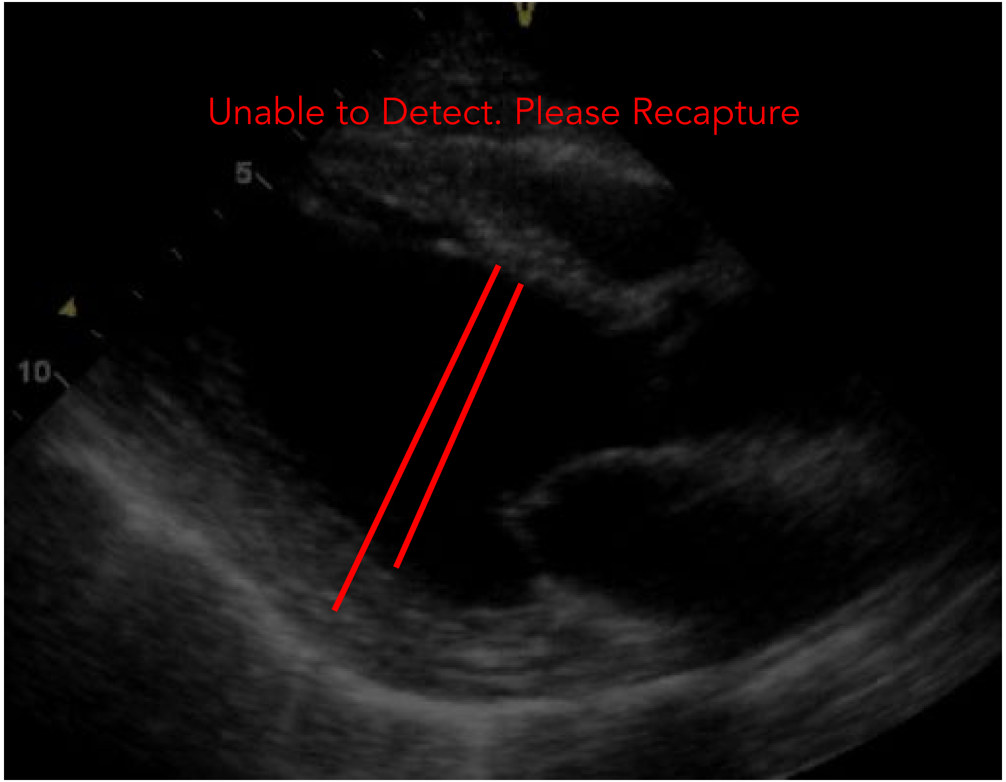
# You don’t need to be fully automatic to be clinically useful



# You don't need to be fully automatic to be clinically useful

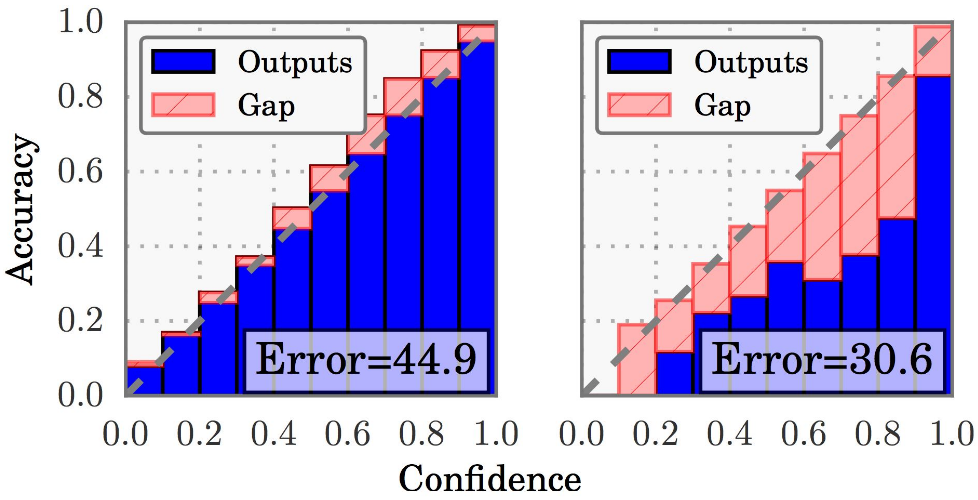


# Quality Control



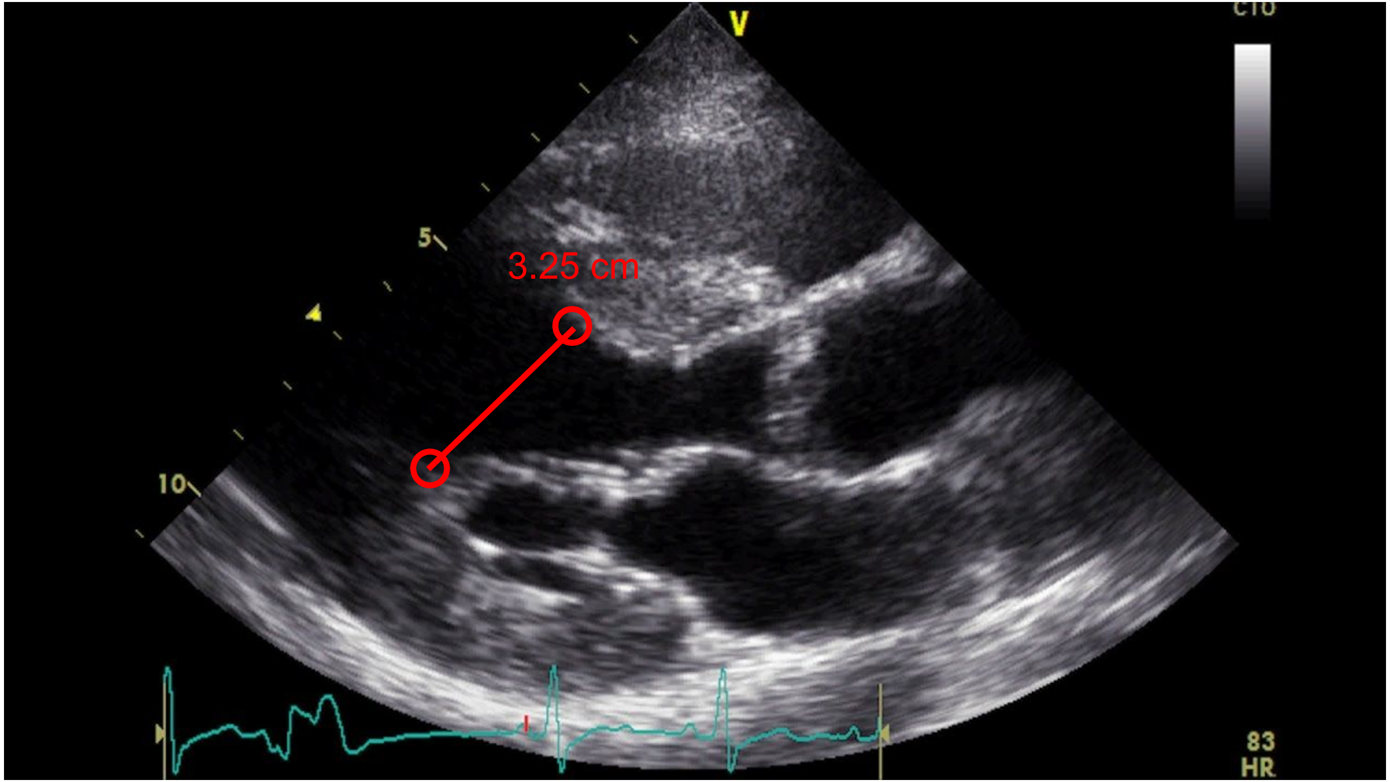
# Model Confidence

- Important to know when a model is not confident.
  - Most DL models are poorly calibrated [1].
  - If a model isn't confident, need to turn over control to a human.

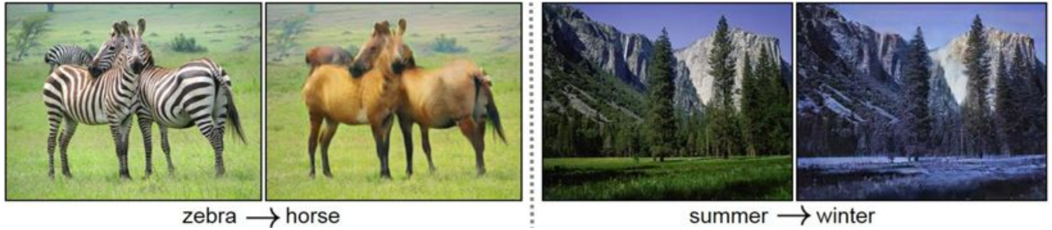



[1] *On Calibration of Modern Neural Networks*, Chuan Guo, Geoff Pleiss, Yu Sun, Kilian Q. Weinberger

# Provenance and Correctability



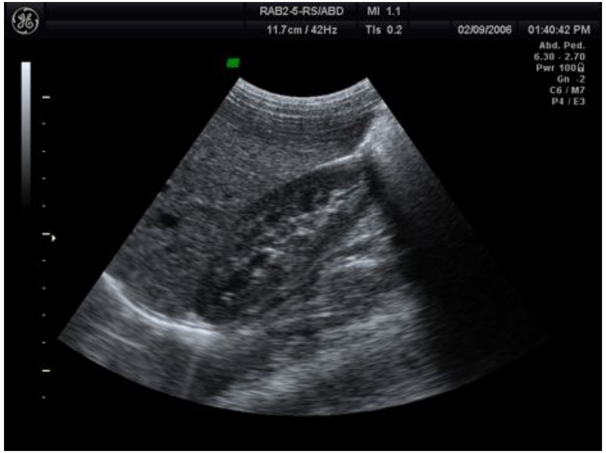
# Domain Adaptation



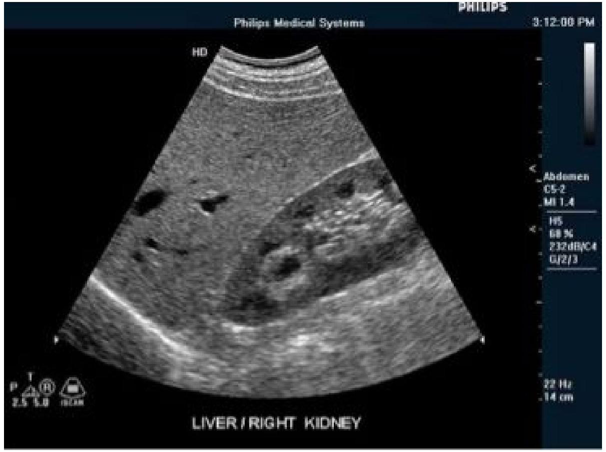


Zhu et al, ICCV 2017

Liu et al., NIPS 2017




?




GE → Philips

# Domain Adaptation

- Pixel-level domain adaptation across imaging devices
- Can we ensure that we don't introduce unwanted artifacts

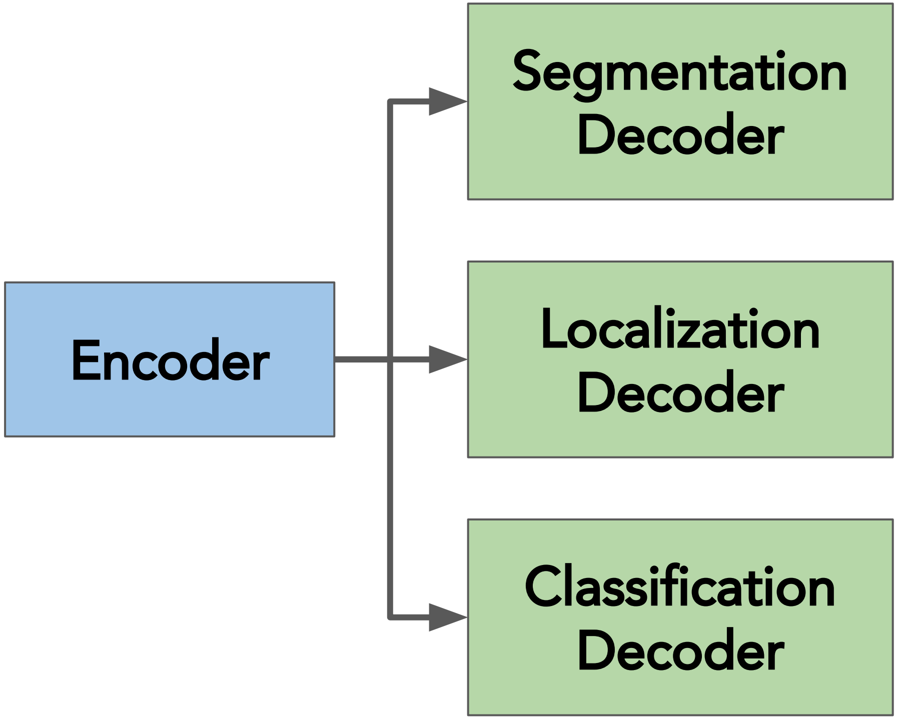
| Train | Test | Accuracy | Mean Class |
|---|---|---|---|
| External Device | Butterfly | 49.6% | 62.0% |
| Fake Butterfly | Butterfly | 95.2% | 96.7% |
| Butterfly | Butterfly | 97.2% | 98.1% |


Harry Yang, USC
PhD Candidate

*Domain Adaptation via Dataset Mimicry*, Harry Yang, Nathan Silberman, *In Progress*

# Multi-Task Models

- How can we leverage data across tasks?
  - ImageNet-style pretraining?
  - Avoid O(N) data scaling
- Deploy smaller models with shared layers
  - Wider models?



# Butterfly Network

msofka@4catalyzer.com