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

- Expensive
- Unwieldy
- Affordable
- Portable
- Powerful
- Not Powerful
Safe Ways to Scan the Body

- Expensive
- Unwieldy
- Powerful
- Affordable
- Portable
- Not Powerful
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

Great Idea → Data and Models → FDA Approval → Product Release
What does it take to actually get a medical ML system to production?
### Bigger Datasets

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>MNIST Handwritten digits</td>
<td>60k training + 10k testing</td>
</tr>
<tr>
<td>Google House Numbers</td>
<td>600k training + 30k testing</td>
</tr>
<tr>
<td>CIFAR-10</td>
<td>50k training + 10k testing</td>
</tr>
<tr>
<td>PASCAL VOC</td>
<td>11k training in 20 classes</td>
</tr>
<tr>
<td>ImageNet</td>
<td>1mm Training in 1000 classes</td>
</tr>
</tbody>
</table>

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

- Suggestion to clinician
- Showing similar cases
- 2nd Read
- Re-prioritizing patients based on image scanning
- Full Automated Diagnostics
You don’t need to be fully automatic to be clinically useful

- Suggestion to clinician
  - Showing similar cases
- 2nd Read
  - Re-prioritizing patients based on image scanning
- Full Automated Diagnostics

Less Fully Automated

Less Strict Regulatory Path (Class II)

More Fully Automated

More Strict Regulatory Path (Class III)
Unable to Detect. Please Recapture
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.

Provenance and Correctability

3.25 cm
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

<table>
<thead>
<tr>
<th>Train</th>
<th>Test</th>
<th>Accuracy</th>
<th>Mean Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>External Device</td>
<td>Butterfly</td>
<td>49.6%</td>
<td>62.0%</td>
</tr>
<tr>
<td>Fake Butterfly</td>
<td>Butterfly</td>
<td>95.2%</td>
<td>96.7%</td>
</tr>
<tr>
<td>Butterfly</td>
<td>Butterfly</td>
<td>97.2%</td>
<td>98.1%</td>
</tr>
</tbody>
</table>

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
Guilford, CT

New York, NY

msofka@4catalyzer.com