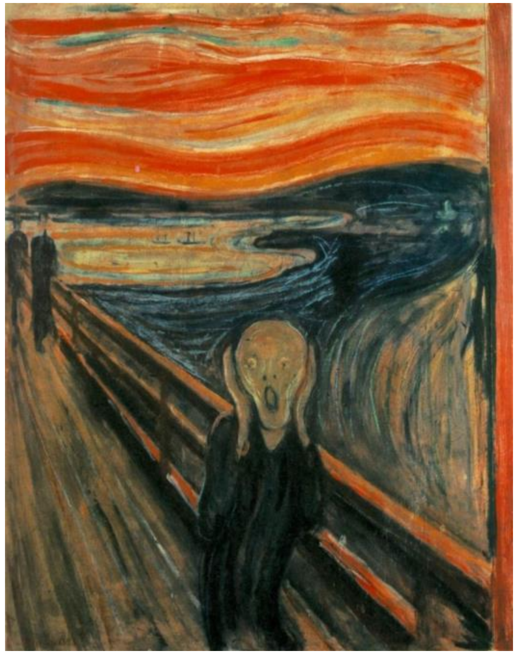
Novel Classes and Epistemic Uncertainty In Medical Imaging

Will Robinson (wgr)
Supervised by Bernie Tiddeman and Reyer Zwiggelaar

The Out-of-Distribution Problem

- Unknown examples produce unpredictable behaviour
- Error and Overconfidence
- Image Recognition networks the best-known examples



Dust Cover, 52% confidence (ResNet 152)

Tench, 36% confidence (DenseNet 161)

A. Shafaei, M. Schmidt and J. J. Little, "A Less Biased Evaluation of Out-of-distribution Sample Detectors," in *British Machine Vision Conference*, Cardiff, 2019.

The Matrix of Knowledge Problems

- Philosophical “Knowledge Problem” deals with epistemological standpoint
- “Known unknowns” etc.
- More useful to consider systemic attributes.
- Is the information valid?
- Does it remain constant within the dataset?

	Constant	**Inconstant**
Valid	Novel Class	Domain Shift
Invalid	Classed Noise	Junk Data

Medical Imaging Uncertainty

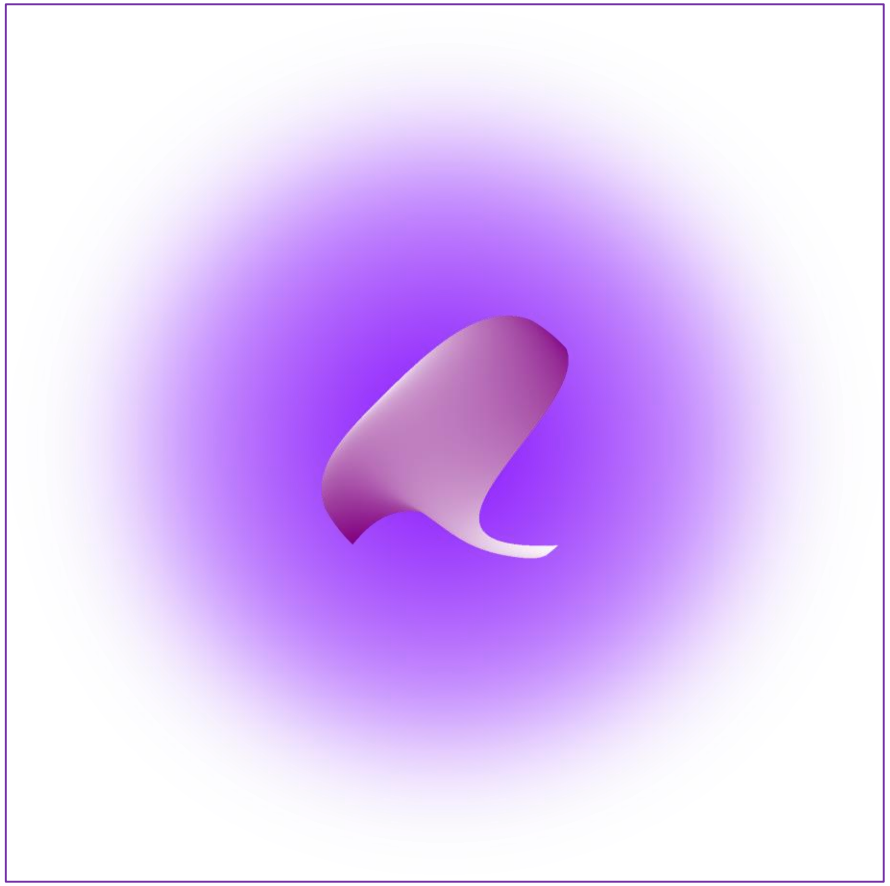
- The medical imaging literature uses a different taxonomy
- **Aleatoric** Uncertainty describes predicable statistical error caused by human variance
- **Epistemic** Uncertainty describes underlying situations distorting the distribution
- Some sources of error can be described in both ways (eg juvenile hypertrophy – underlying unspecified distorting factor, but is part of human biological variance)

The Topology of Error

A different way of looking at the problem is by envisaging the complete set of possible samples as a hyperspace. We can then model several volumes within that hyperspace.

Aleatoric Domain

The original training distribution forms an irregular volume. Statistical variation will mean that there will be a “halo” of valid samples around that volume. This is the aleatoric domain.



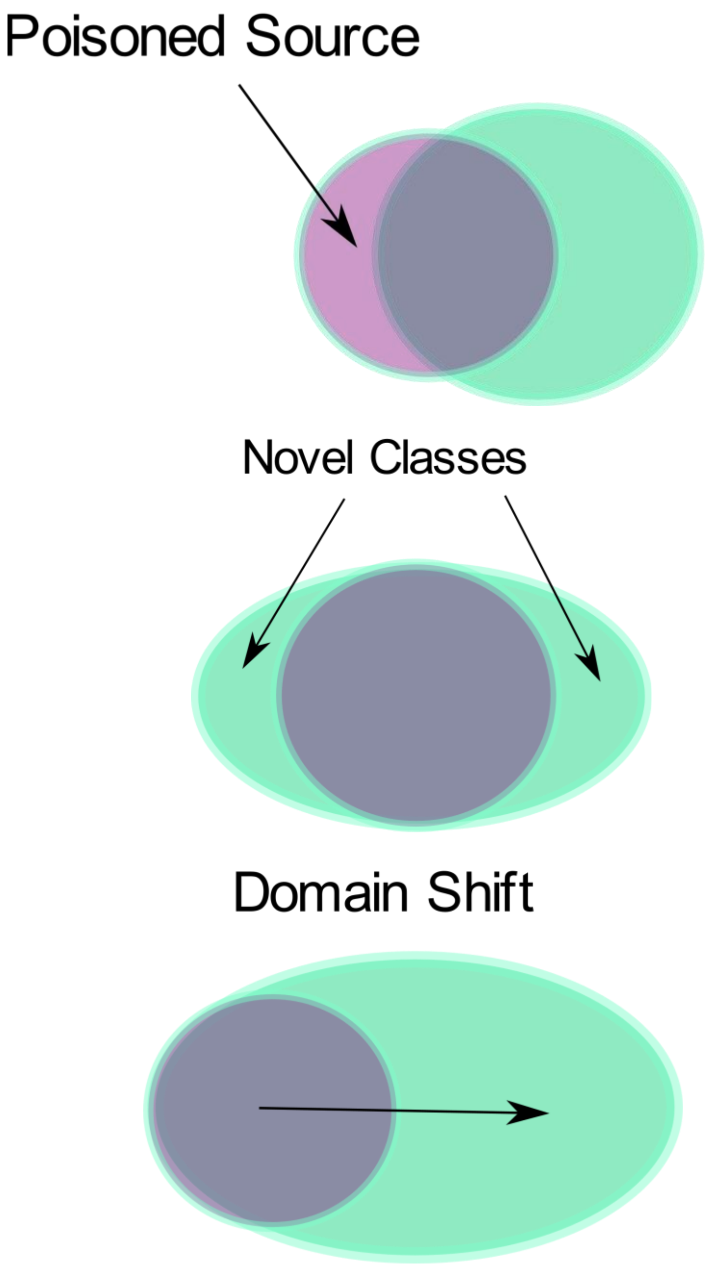
Bounding Aleatoric Uncertainty

- PAC-Bayes is a major field of interest for bounding ML models.
- Stochastic neural networks can be used to sustain the entire embedding space as a set of likelihood ratios instead of sample points (analogous to Variational Autoencoders).
- All bounding models have a habit of finding vacuous bounds (likelihood higher than 1.0), especially on high-variance datasets.
- Recent work has successfully found non-vacuous bounds on stochastic convolutional models for medical imaging (Sicilia et al 2021)
- Since these bounds are derived from the training data, they are inherently unable to bound the validity volume.

Epistemic Errors

We can presume a volume of all possible *valid samples.* This is the epistemic domain, whose bounds are defined by domain validity rather than statistical variance.

These forms of error extend from differences between the training volume and the epistemic domain. Each type of error can only be bounded with underlying domain knowledge.



Measures of Epistemic Uncertainty

While we cannot universally bound epistemic uncertainty, we can use domain knowledge to measure specific elements of the validity volume.

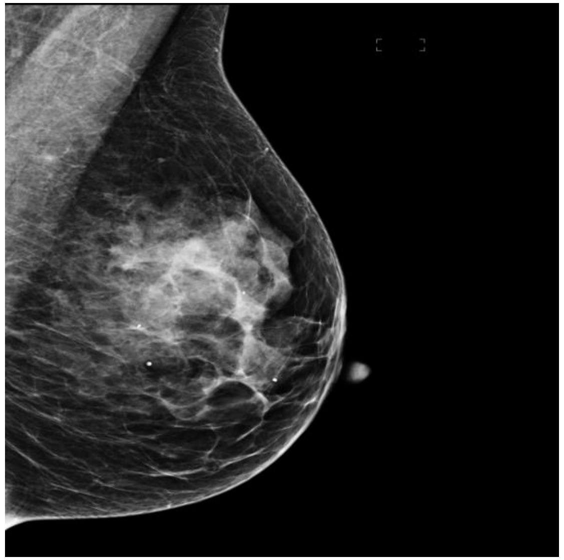
Examples:

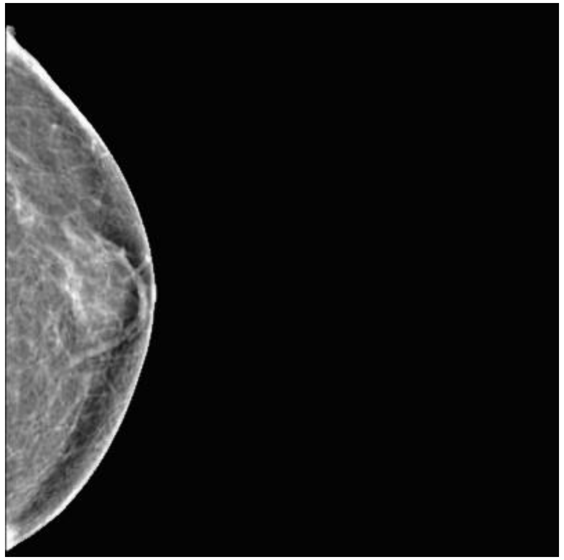
- Dropped Class (Novel Class)
- Contextual Interpolation (Domain Shift)
- Adversarial Retraining (Distributional Invalidity)

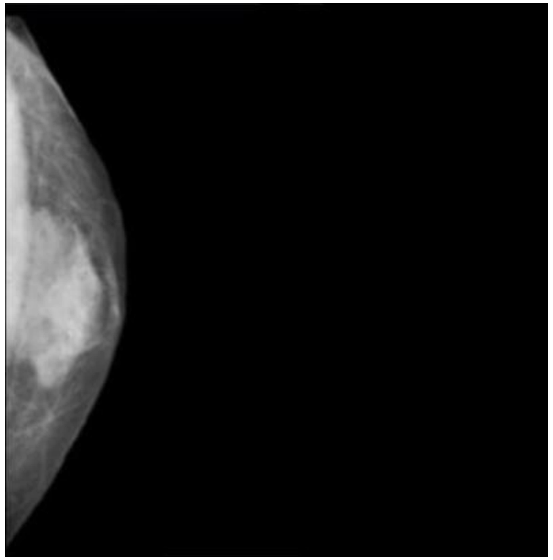
Note that these are harder to do with medical imaging than classical image sets, as medical error is both complex and infrequent.

Novel Class Problems

- Cross-domain class of problem
- One or more unseen classes
- Valid / Constant
- Linear Example: Hexadecimal MNIST
- Orthogonal Example: Gynecomastia/Juvenile Hypertrophy







Sources: OPTIMAM, Public Domain Collection

Novel Class Ensemble

Concept

- Classes naturally share much in common.
- NNs find an *exclusionary* minimum
- Therefore it is hard to train a non-linear function around a subset of existing classes which *includes* potential classes
- Distinguishing one class from the others would produce a set of representations of distinctions
- The idea – does this set of distinctions in summation produce a representation of the elements of a “generic class”?

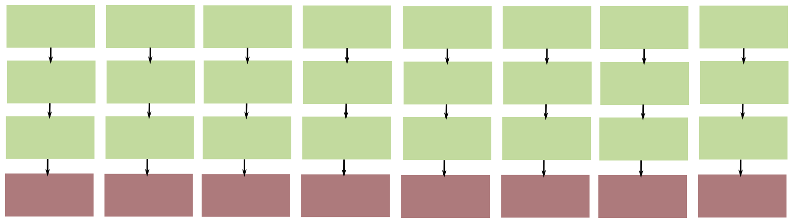
Additional questions

- Is there an effect from the relative inter-class similarity?
- Underfitting? Overfitting?

Novel Class Ensemble

Architecture

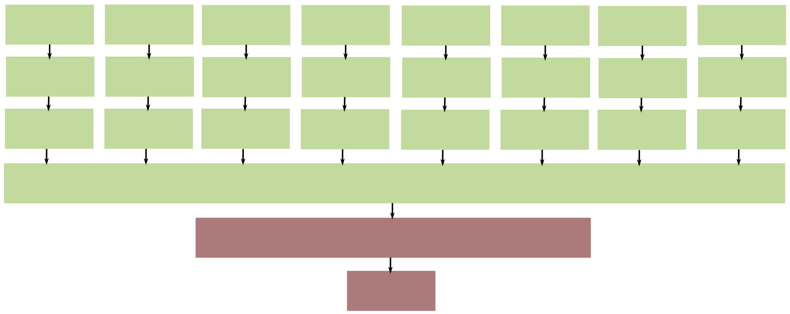
Step 1: Ensemble of n-2 simple convolutional subnetworks pretrained as binary classifiers for one class.



Step 2: Combine locked ensemble filter outputs together into a larger network

Step 3: Train combined network to recognise one dropped class from noise/junk

Step 4: Test combined network on remaining dropped class and noise/junk



Novel Class Ensemble Results

- Single-class accuracy (baseline 0%/33%):
  MNIST 35%
  CIFAR 16%
- Note that in both cases the ensemble can distinguish junk data with very high accuracy (>95%).
- Mean SSIM between all classes:
  MNIST: 0.38670
  CIFAR: 0.18803

Medical Imaging Application

- High structural similarity (especially in mammography)
- However, many medical classes are not discrete, but instead progressive (eg BIRADS)
- Dropped class not usable for binary classification or segmentation
- Skin cancer provides a convenient discrete class problem
- ISIC 2018 and 2019 provide convenient examples

Current Challenges

- Network is often unstable
- ISIC sets produce degenerate inter-class distinctions
- Unbalanced datasets
- Overfitting in combined network

Summary

- Out-of-distribution (OOD) problems
- Taxonomies of Error
- Aleatoric vs Epistemic Uncertainty
- Novel Class Problems
- Ensemble of Binary Subnetworks
- Early positive results
- Lots of work to do!