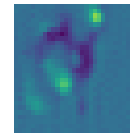
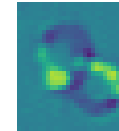
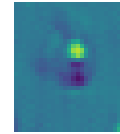




**Moedal: ML 8<sup>th</sup> + 14<sup>th</sup> feb**

[l.millward@qmul.ac.uk](mailto:l.millward@qmul.ac.uk)

# New Labeled data



	Old	New
Total Pits	615	1393
Canidates	52	259
Top, surf	90	199
Bottom, sur	363	811
Misc	110	80
Maybe		44

- Heuristic  
~ 500 examples per class  
for smooth training
- Now within range
- Can repeat old studies
- Prior problem hitting  
statistical fluctuations  
Caused large systematic  
problems w learning

Loosened Candidate threshold  
New 'Maybe' category

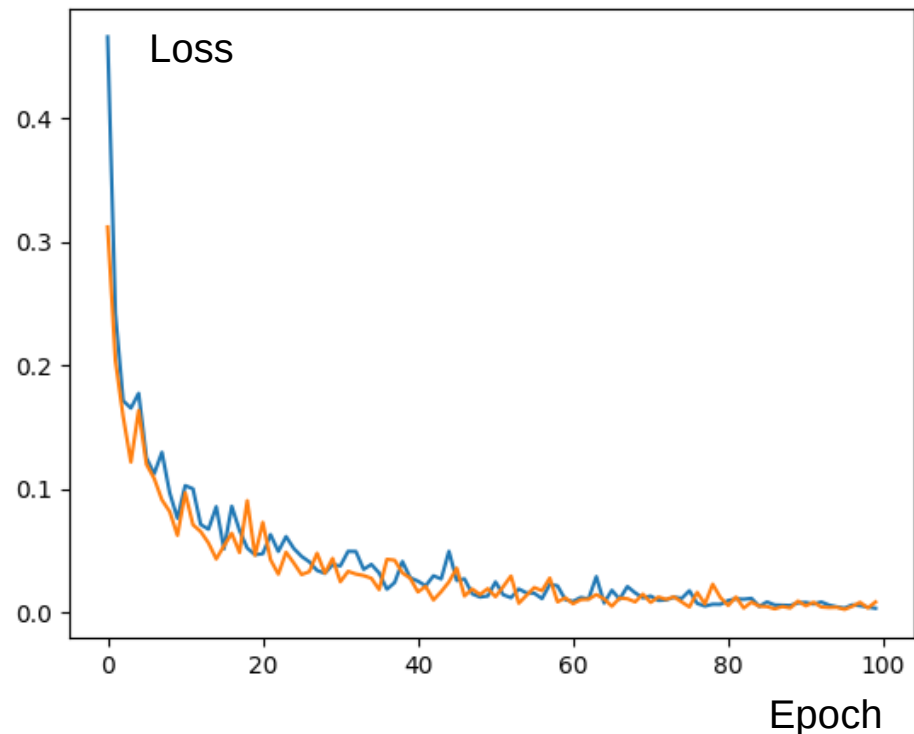
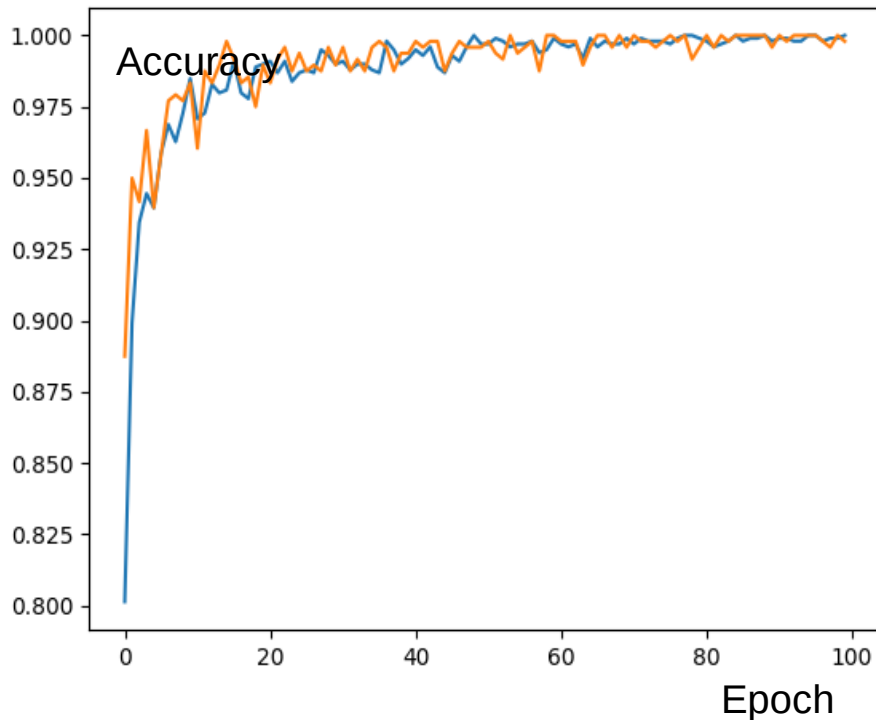
For consistency will re-label old data-set  
Clearly labelling style has changed

PreSelection only has ~5%  
False Pit ID

# Top Vs Bottom Surface

2D image  
recognition, in single  
channel

- Simple problem, Used as benchmark previously when signal low. T vs B , ultimately ~ intensity gradient
- Demonstratable improvement in Stability and Learning



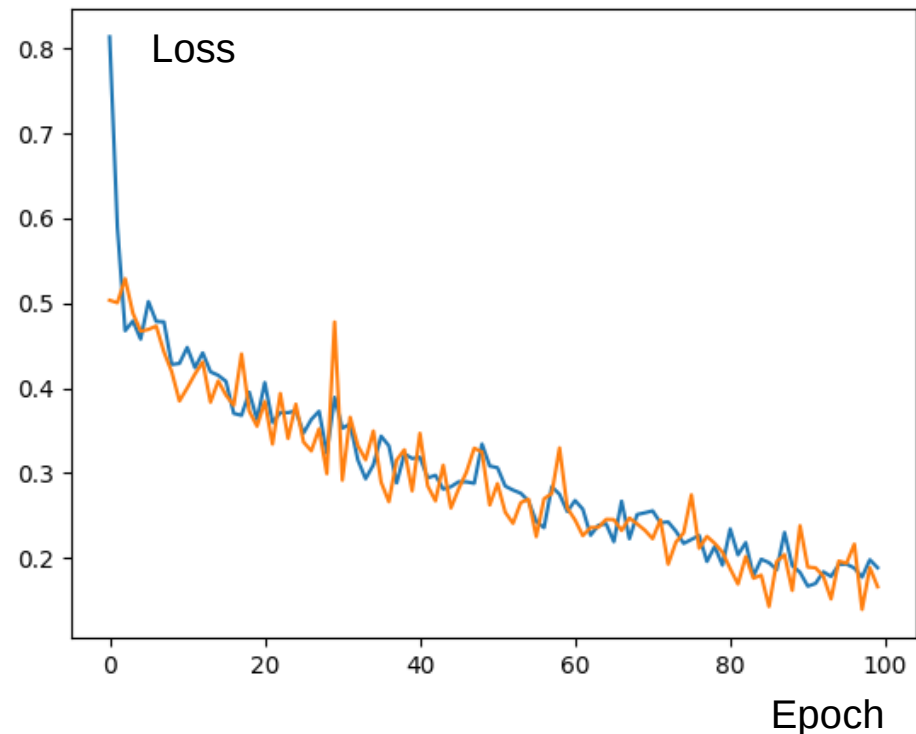
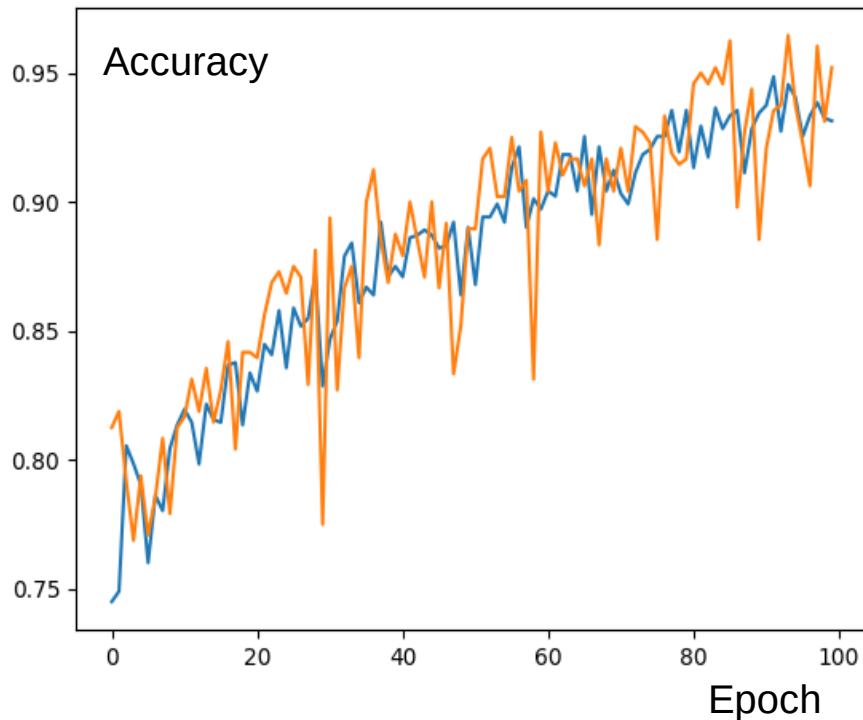
- No overtraining divergence!
- NB# Not generalisable result since top + bottom alone doent represent full problem domain

Adam Optimiser .0001 lr

# Eg, Signal VS top + botom

- Much Smoother Training
- Statistical fluctuations / noise an scitll be seen in the smaller validation set (orange)

2D image  
recognition, in single  
channel



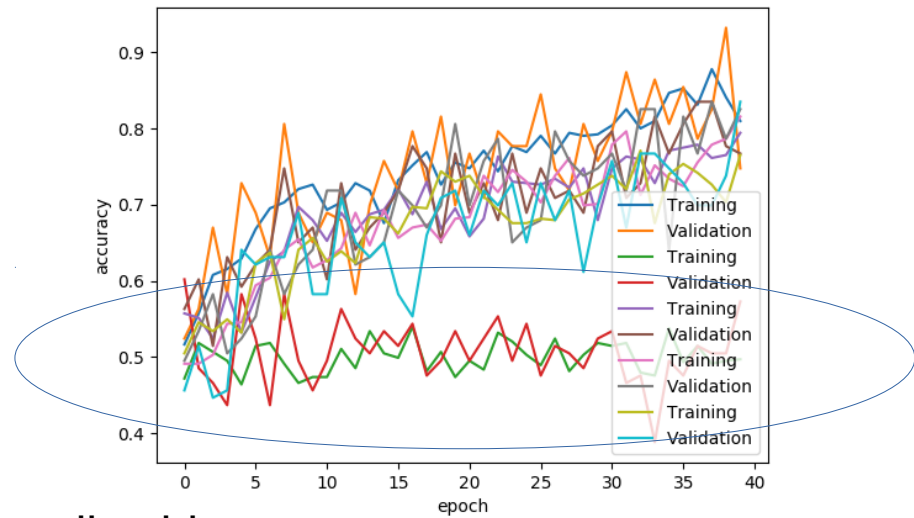
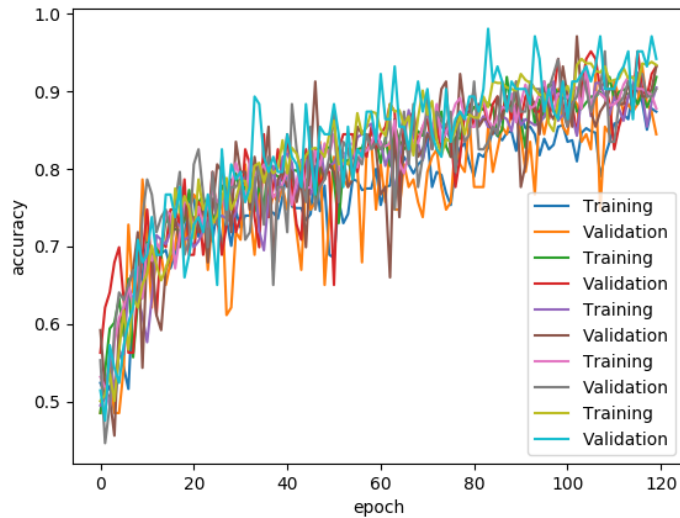
- Accuracy ~ 90-95% # Not same as sig eff,
- No overtraining divergence
- NB# Not generalisable result  
background doent represent full problem domain

Adam Optimiser .0001 lr

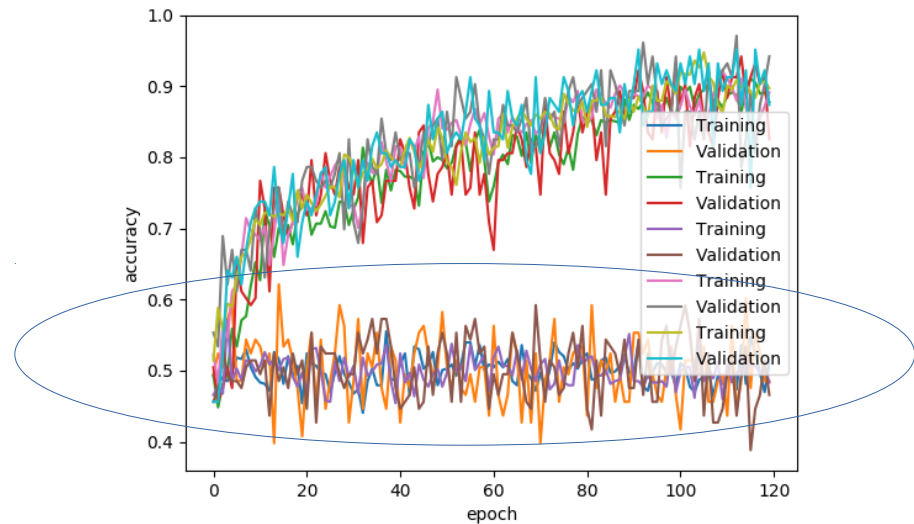
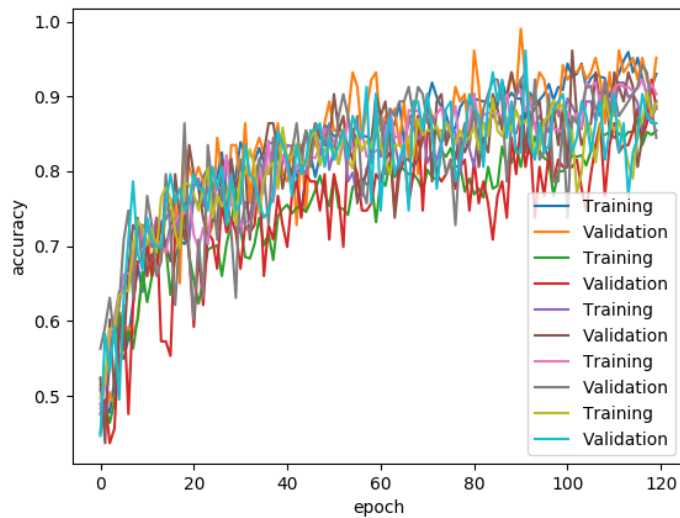
# Issues

- Previous 'Failure mode'  
No / minimal signal,  
High uncertainty over signal samples  
∴ Bias towards background classification
- 'Binary Accuracy' poor metric for moedal  
want strong signal acceptance preference  
Long term → custom loss function
- Short term, Balance sig vs bkg  
Remove background, or Augment signal
- //////////////////////////////////
- Generally observing good performance  
However sometimes network 'Fails' to  
learn, doesnt learn any consistent pattern  
/ rules in training or validation
- Double checked Training,  
Testing, Validation
- Not overtraining,  
consistent convergence  
without major test train  
divergence
- Hypothesis, could be 'bad'  
data examples, that dont  
provide consistent rules to  
learn from / non-  
representative

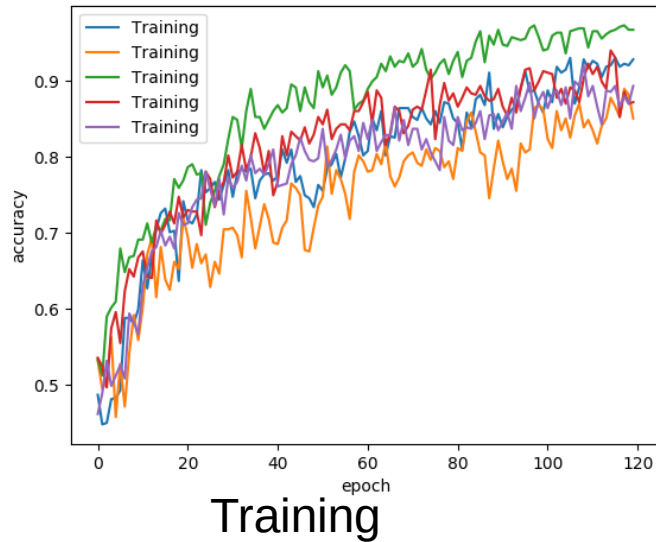
# 5 fold cross-validation



Divergence in learning behavior. Unpredictable



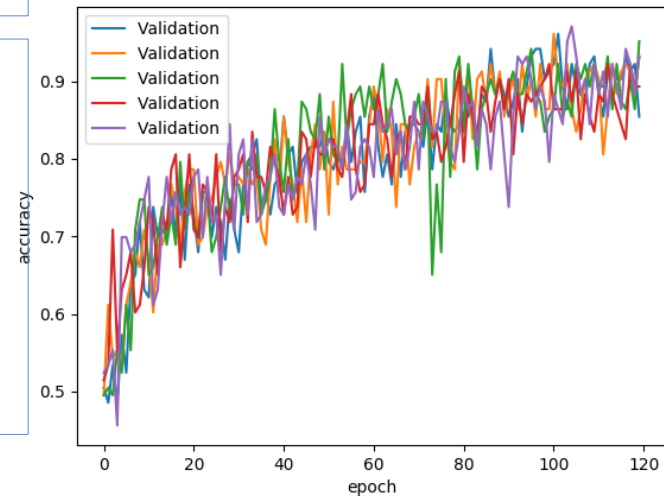
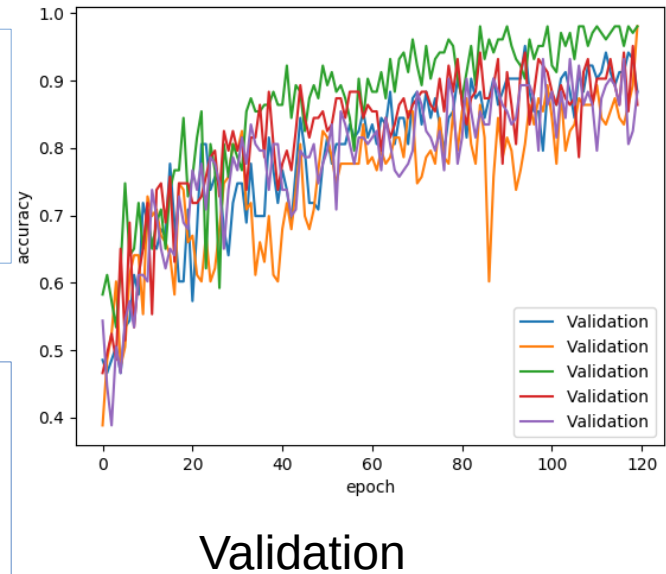
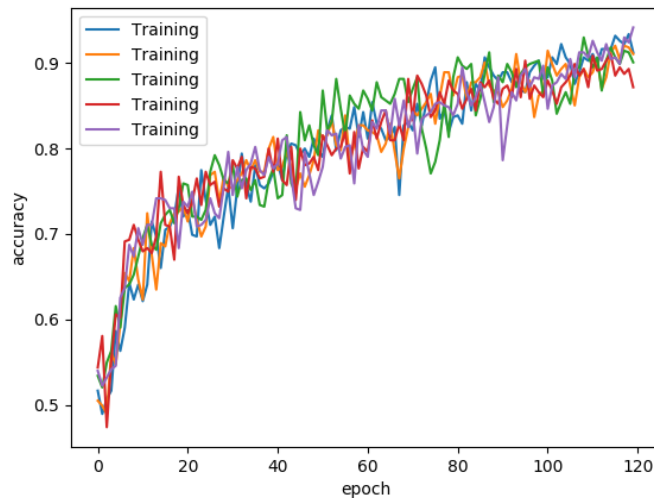
# Training, Validation Consistency



Not a Test / Train Divergence, Cluster similarly with larger test spread

Unlikely a bad portion of data, as would show up as bad k-fold

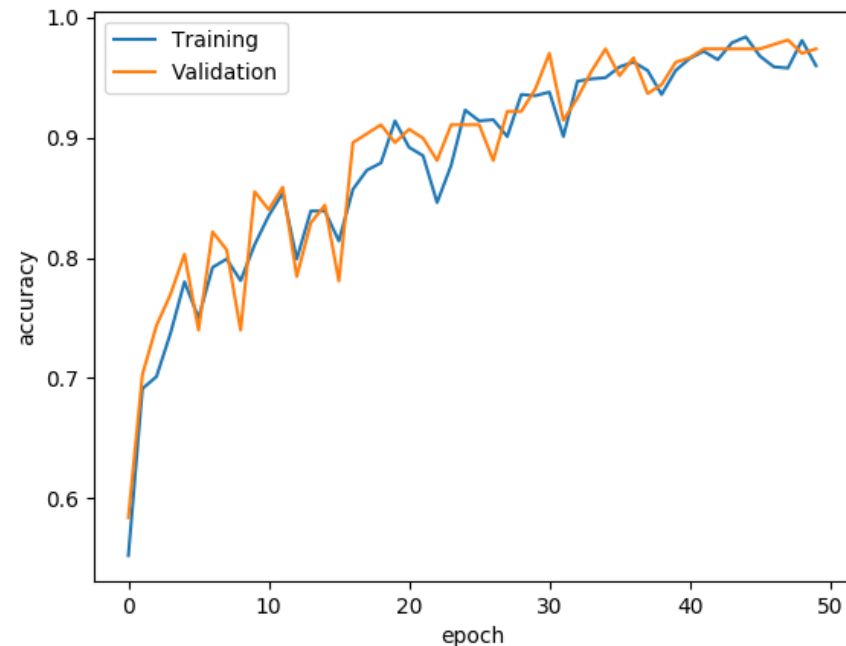
Failure to learn appears to just be random  
Not problem as ultimately will only keep models that work



All models, inc good ones do better at background rejection  
Than signal acceptance – We would really like Signal Acceptance

# Conclusion

- 'Consistent' accuracy ~ 90% (neglecting Outlying trainings)  
Inc completely unseen data
- Consistently better at rejecting / classifying background than accepting signal
  - # Seen in multiple runs
  - # Suggests 'Signal as Anomaly' hypothesis - could be useful
  - V. Good False-positive rate boosts 'nominal' score, But balances poorer True-positive, False-negative
- Would rather MAX T-P



Still ..... ~ 90 % Should be able to run in inference on unlabeled regions of foil,  
Still ... have more data from one scanning of one foil (calib)  
Should be able to find New signal examples in this data,



# Questions / next

- Multi-Class labelling maybe Bkg is too easy to spot / biased
- Test 3 / 4 way labelling

- What happens if include all channels as examples of the relevant class?  
How important is angle

- Hyper-params, parameter scans
- Different optimisers
- Depth /2d,3dK,3dC

- Go back to patches / large area labelling
- Can we spot 'sig' anomalies