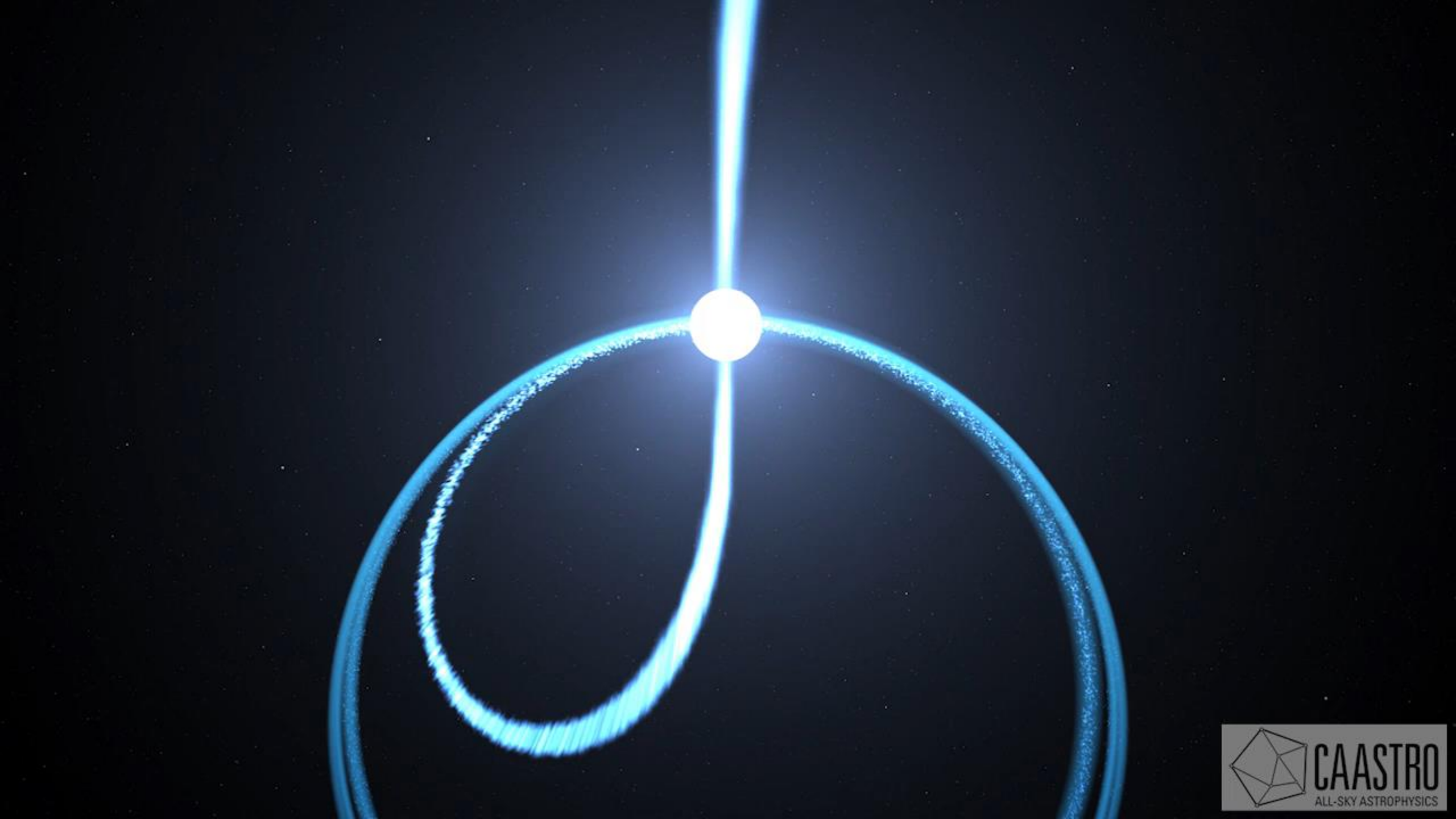
PSRFINET: Radio Frequency Interference Detection in Pulsar Data with Deep Neural Networks

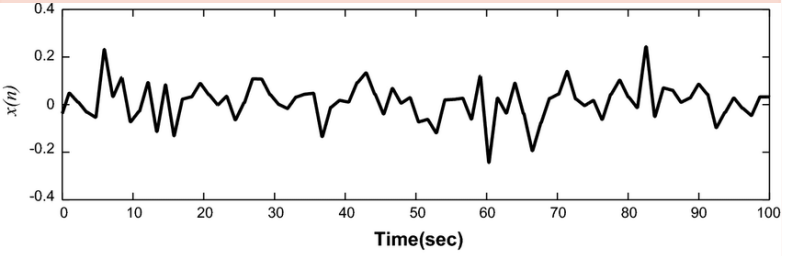
Afiq Abdul Hamid	**Institute of Radio Astronomy and Space Research (IRASR)**
Willem van Straten	IRASR
Anthony Griffin	High Performance Computing Research Laboratory (HPCRL)


CAASTRO
ALL-SKY ASTROPHYSICS

Pulsar Data

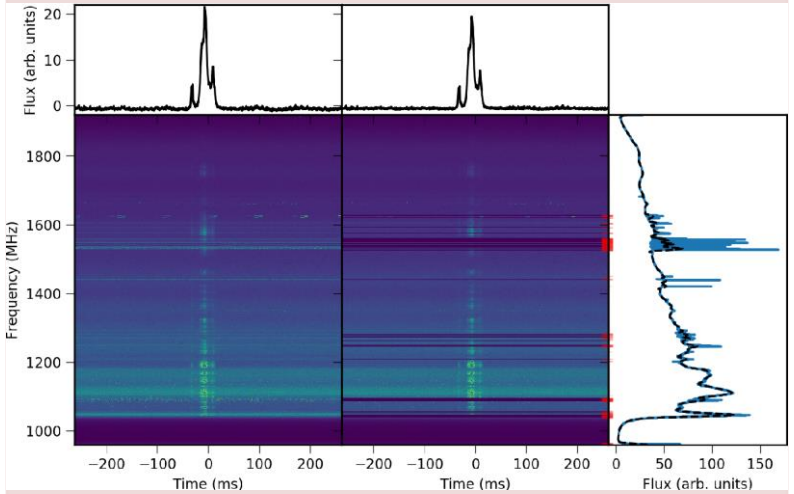
Baseband data

- Raw voltages sampled at $f_{Nyquist}$
- High storage space requirements



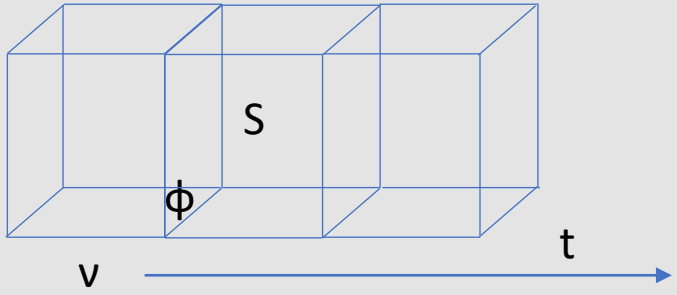
Filterbank data

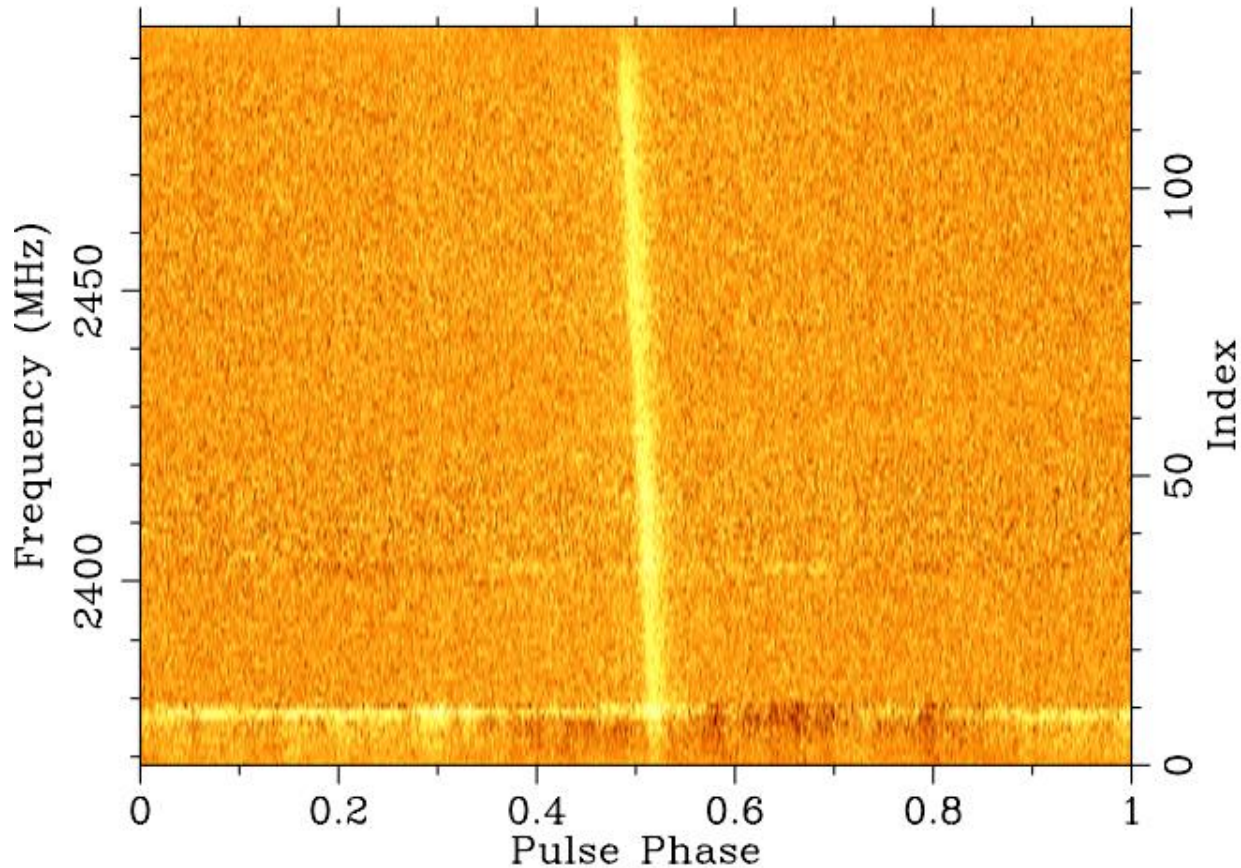
- Baseband channelised with PFB



Folded data

- 3D Data cube: ν x ϕ x S.
- Appended in time.



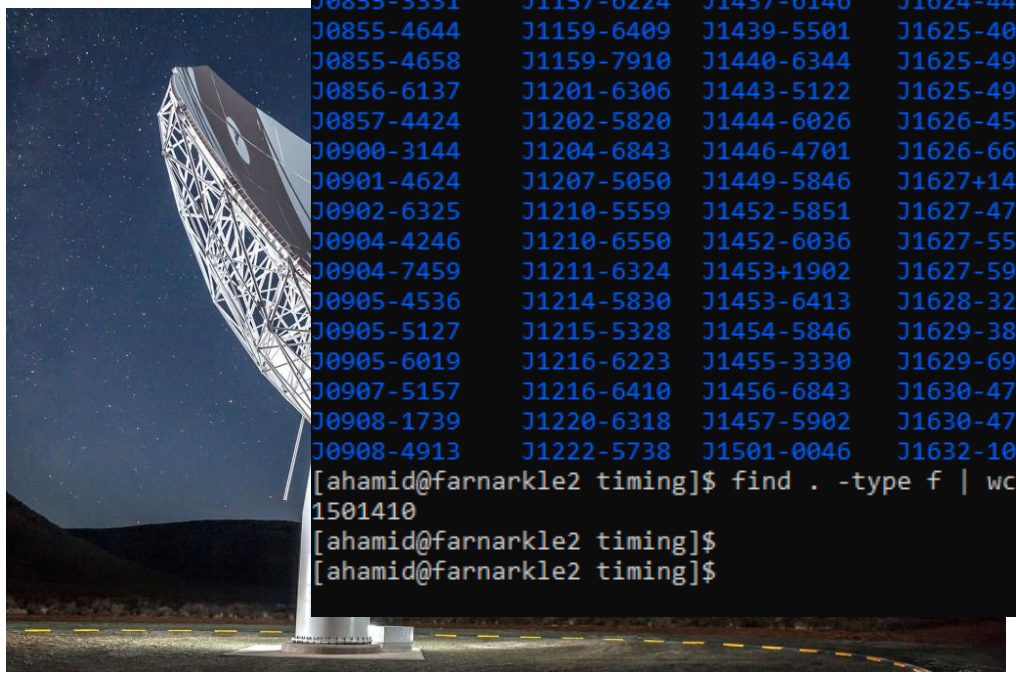

Frequency (MHz)
2400
2450
0
0.2
0.4
0.6
0.8
1
Pulse Phase
0
50
100
Index

Why P[illegible] A[illegible]

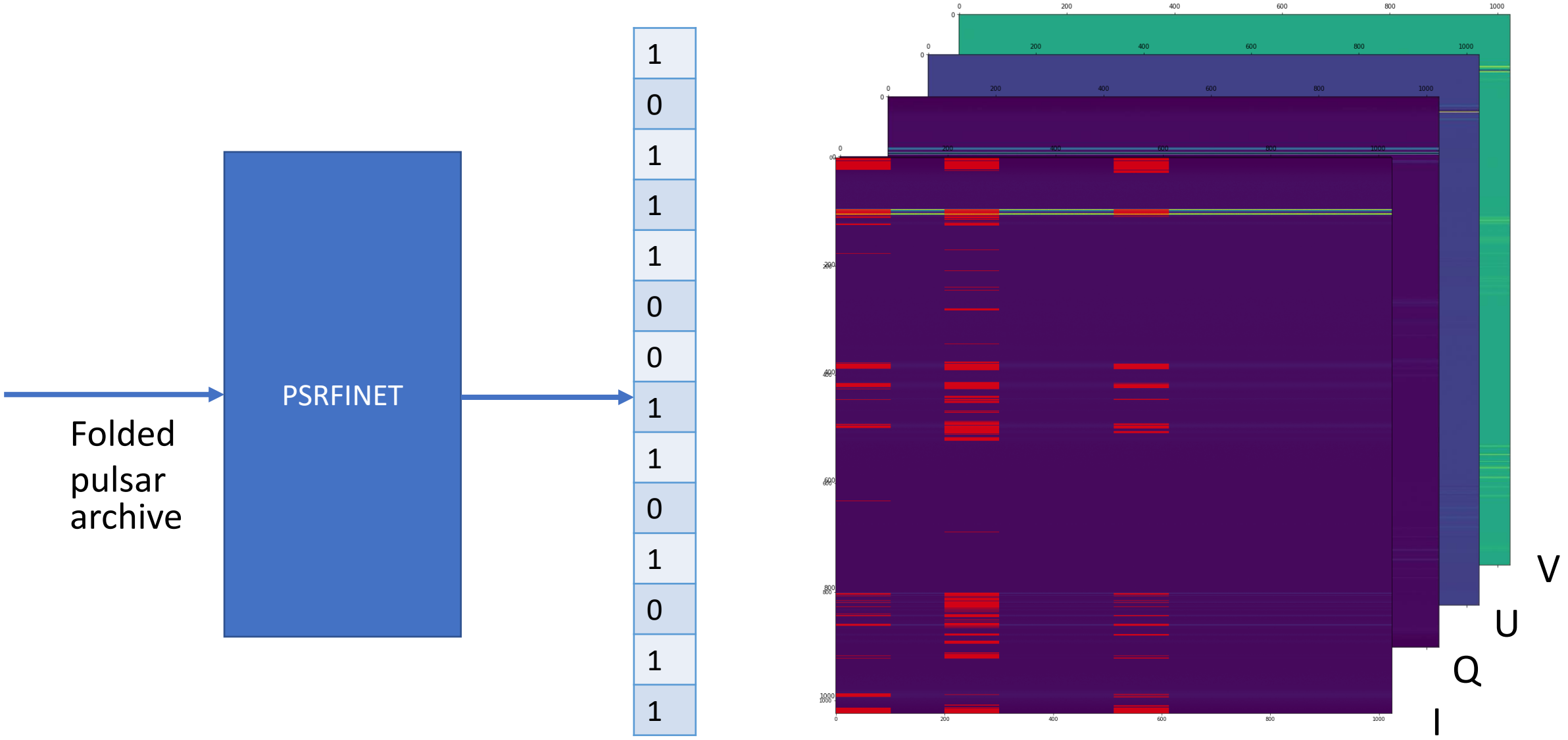
- **Big Data**
- GPU Acc[illegible]
- 3rd party

ahamid@farnarkle2:/fred/oz005/timing

```
J0823+0159    J1141-3107  J1424-5556  J1617-4608  J1725-3853  J1817-3618  J1847-0605  J1914+0219  J2136-1606
J0824+0028    J1141-3322  J1424-5822  J1617-5055  J1726-3635  J1817-3837  J1848+0351  J1914+0631  J2139+2242
J0828-3417    J1141-6545  J1424-6438  J1618-3921  J1726-4006  J1818-0151  J1848+0604  J1914+0659  J2140-2310A
J0831-4406    J1142+0119  J1425-5723  J1618-4624  J1727-2739  J1818-1422  J1848+0647  J1914+1122  J2144-3933
J0834-4159    J1142-6230  J1425-5759  J1618-4723  J1727-2946  J1818-1607  J1848+0826  J1915+0227  J2144-5237
J0835-3707    J1143-5158  J1425-6210  J1620-5414  J1728-0007  J1819+1305  J1848+1516  J1915+0738  J2145-0750
J0835-4510    J1143-5536  J1427-4158  J1621-5039  J1730-2304  J1819-0925  J1848-0023  J1915+0752  J2150-0326
J0836-4233    J1144-6217  J1428-5530  J1621-5243  J1730-3350  J1819-1008  J1848-0123  J1915+0838  J2151+2315
J0837+0610    J1146-6030  J1429-5935  J1622-0315  J1731-1847  J1819-1114  J1848-0601  J1915+1009  J2154-2812
J0837-2454    J1146-6610  J1430-6623  J1622-3751  J1731-3123  J1819-1458  J1848-1150  J1915+1410  J2155-3118
J0837-4135    J1147-6608  J1431-4715  J1622-4332  J1731-4744  J1819-1510  J1848-1243  J1915+1606  J2205+1444
J0838-2621    J1148-5725  J1431-5740  J1622-4347  J1732-3131  J1819-1717  J1849+0409  J1915+1647  J2214-3835
J0840-5332    J1151-6108  J1432-5032  J1622-4802  J1732-4156  J1820-0427  J1849+2423  J1916+0844  J2215+1538
J0842-4851    J1152-6012  J1433-6038  J1622-4950  J1732-5049  J1820-0509  J1849-0317  J1916+0951  J2222-0137
J0843-5022    J1154-6250  J1434-6006  J1622-6617  J1733-3716  J1820-1346  J1849-0614  J1916+1030  J2229+2643
J0846-3533    J1156-5707  J1435-5954  J1623-0841  J1733-4005  J1820-1529  J1849-0636  J1916+1225  J2234+0611
J0847-4316    J1156-5909  J1435-6100  J1623-0908  J1734-0212  J1820-1818  J1850+0026  J1916+1312  J2234+0944
J0849-6322    J1157-5112  J1437-5959  J1623-4949  J1734-2415  J1821+0155  J1850+0124  J1916-2939  J2234+2114
J0855-3331    J1157-6224  J1437-6146  J1624-4411  J1734-3058  J1821+1715  J1850+0242  J1917+0834  J2236-5527
J0855-4644    J1159-6409  J1439-5501  J1625-4048  J1735-0724  J1821-0256  J1850+0423  J1917+1353  J2241-5236
J0855-4658    J1159-7910  J1440-6344  J1625-4904  J1735-3258  J1821-0331  J1850+1335  J1917+1737  J2243+1518
J0856-6137    J1201-6306  J1443-5122  J1625-4913  J1737-0811  J1821-1432  J1850-0006  J1918+1444  J2248-0101
J0857-4424    J1202-5820  J1444-6026  J1626-4537  J1737-3102  J1822+0705  J1850-0026  J1918+1541  J2253+1516
J0900-3144    J1204-6843  J1446-4701  J1626-6621  J1737-3555  J1822+1120  J1851+0233  J1918-0642  J2307+2225
J0901-4624    J1207-5050  J1449-5846  J1627+1419  J1738+0333  J1822-1400  J1851+0418  J1918-1052  J2310-0555
J0902-6325    J1210-5559  J1452-5851  J1627-4706  J1738-2647  J1822-4209  J1851+1259  J1919+0021  J2317+1439
J0904-4246    J1210-6550  J1452-6036  J1627-5547  J1738-2736  J1823+0550  J1851-0029  J1919+0134  J2317+2149
J0904-7459    J1211-6324  J1453+1902  J1627-5936  J1738-2955  J1823-0154  J1851-0053  J1919+1314  J2322+2057
J0905-4536    J1214-5830  J1453-6413  J1628-3205  J1738-3211  J1823-3021A J1851-0114  J1919+1645  J2322-2650
J0905-5127    J1215-5328  J1454-5846  J1629-3825  J1739+0612  J1823-3021F J1852+0008  J1919+1745  J2324-6054
J0905-6019    J1216-6223  J1455-3330  J1629-6902  J1739-1313  J1823-3021G J1852+0013  J1920-0950  J2330-2005
J0907-5157    J1216-6410  J1456-6843  J1630-4719  J1739-2903  J1823-3106  J1852-0118  J1921+0137  J2346-0609
J0908-1739    J1220-6318  J1457-5902  J1630-4733  J1739-3023  J1824-0127  J1852-0127  J1921+0812
J0908-4913    J1222-5738  J1501-0046  J1632-1013  J1739-3131  J1824-0132  J1852-0635  J1921+1419
[ahamid@farnarkle2 timing]$ find . -type f | wc -l
1501410
[ahamid@farnarkle2 timing]$
[ahamid@farnarkle2 timing]$
```



Our Approach: PSRFINET



Constraints

N subintegration (0.7 training, 0.15 validation, 0.15 test)

L band (v = 1284 MHz, vBW = 856 MHz)

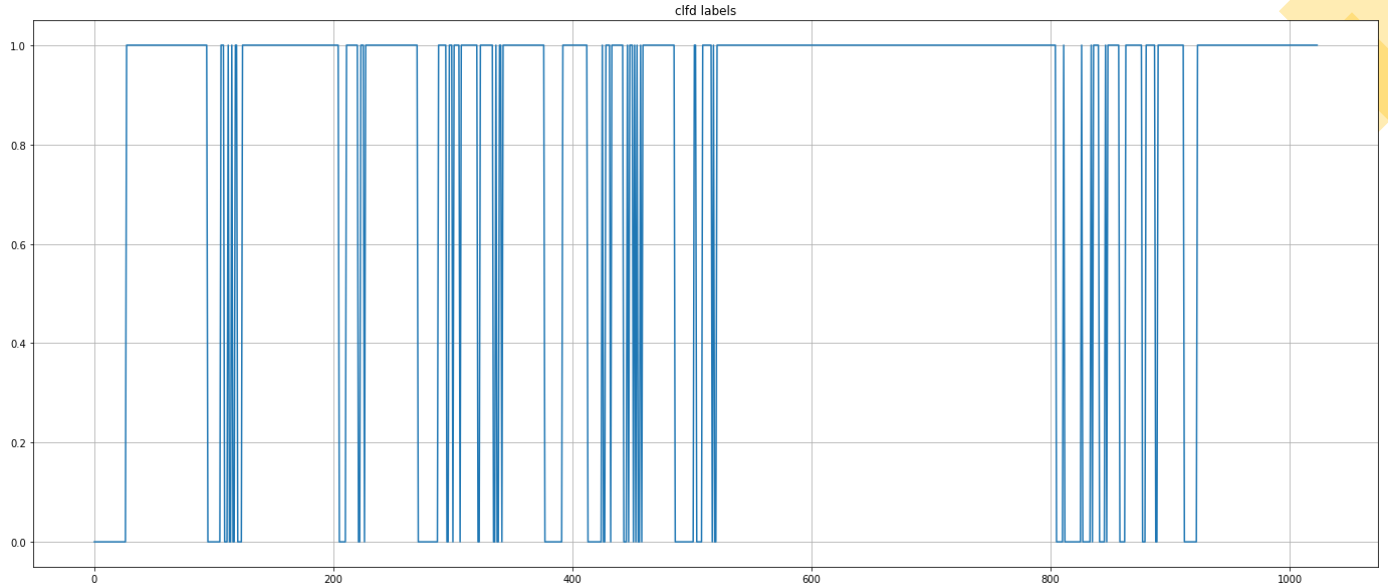
8 Second Subintegrations

1024 channels

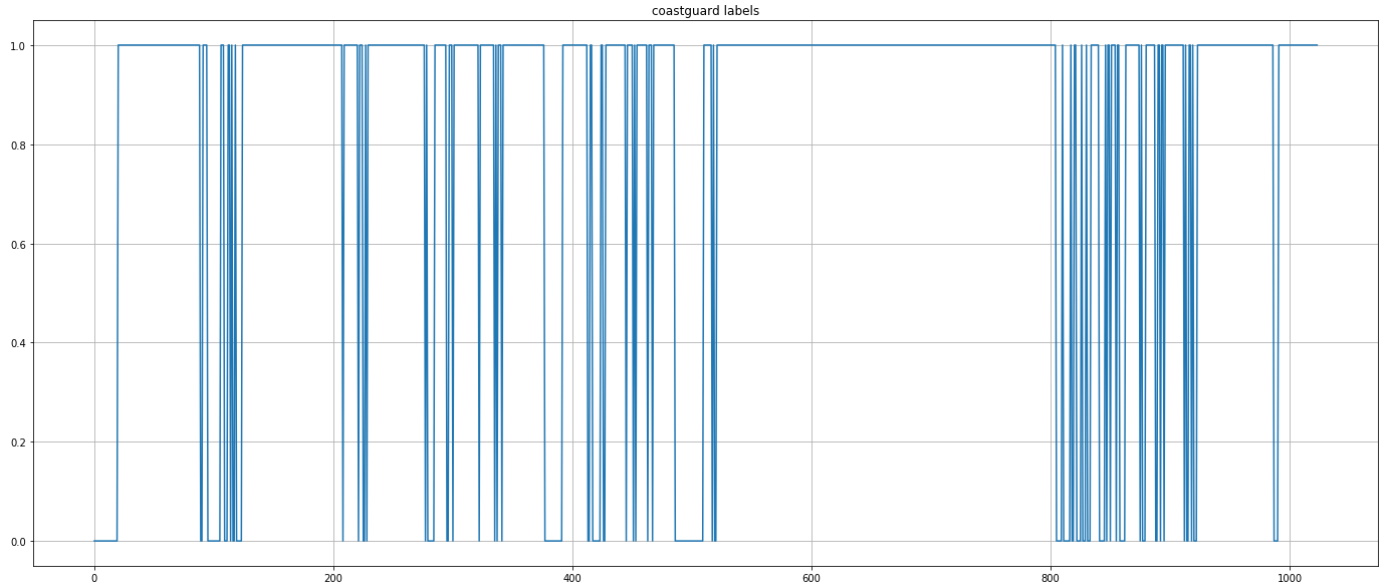
SNR Limits

- 14 (Upper)
- 4 (Lower)

- Ground Truth
  - Clfd (Morello et al, 2018)
    - [Q1 − qR, Q3 + qR]
    - R = Q3 - Q1
    - (std, ptp, lfamp, q=1.75)



  - Coastguard (Lazarus et al , 2016)
    - (surgical, cthresh = 5, sthresh = 5, cut edge = 0.1)
    - (bandwagon, cthresh = 0.8, stresh = 0.8)



  - Mixed (50/ 50)

Residual Neural Network Keras Resnet 50

Model	Size (MB)	Top-1 Accuracy	Top-5 Accuracy	Parameters	Depth	Time (ms) per inference step (CPU)	Time (ms) per inference step (GPU)
Xception	88	0.790	0.945	22,910,480	126	109.42	8.06
VGG16	528	0.713	0.901	138,357,544	23	69.50	4.16
VGG19	549	0.713	0.900	143,667,240	26	84.75	4.38
ResNet50	98	0.749	0.921	25,636,712	-	58.20	4.55
ResNet101	171	0.764	0.928	44,707,176	-	89.59	5.19
ResNet152	232	0.766	0.931	60,419,944	-	127.43	6.54
ResNet50V2	98	0.760	0.930	25,613,800	-	45.63	4.42

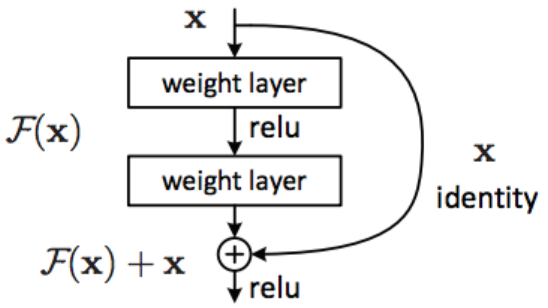


Figure 2. Residual learning: a building block.

Hyperparameters

- Dataset: 4200
- N epochs: 200
  - (early stopping monitor: val loss)
- Batch Size: 3
- Steps per epoch: 180
- Learning Rate: 0.055
- Class Weights: [0.65 4.25]
  - (Initial Assumption)
- Optimizer: SGD
- Activation: Leaky RELU

Processing and Augmentation

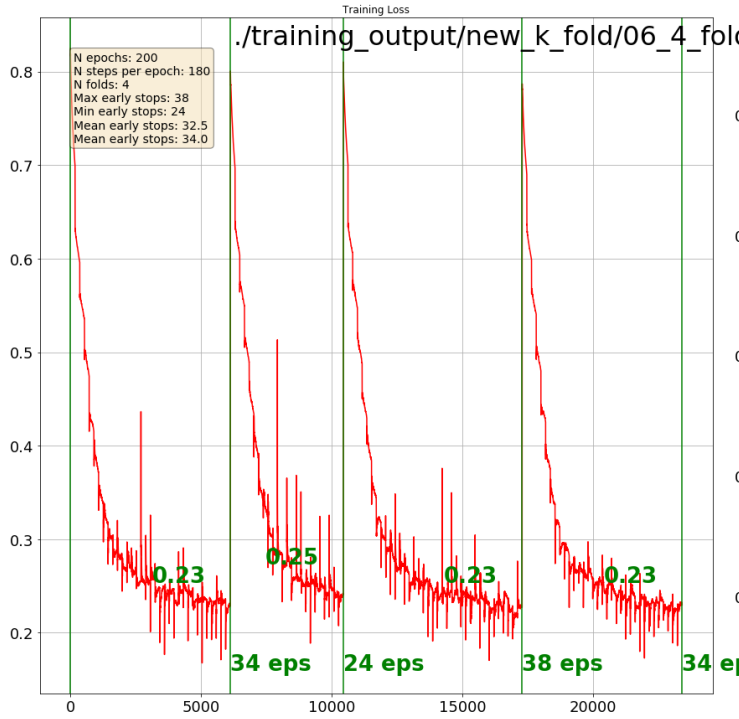
Processing:

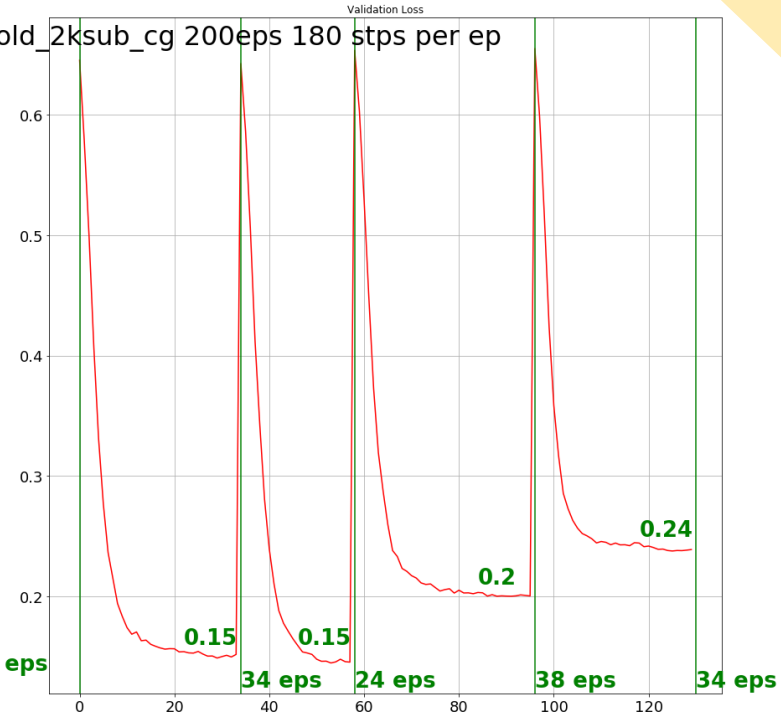
- Transpose and clip band edges (N_{ch} = 984, 872 < ν < 1695 MHz)
- Scaling: 0 < float(data) < 1

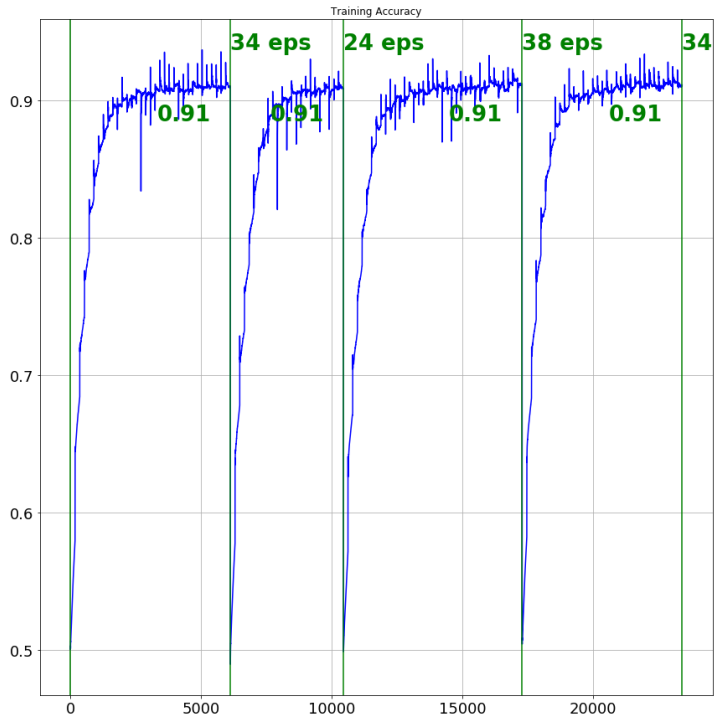
Augmentation:

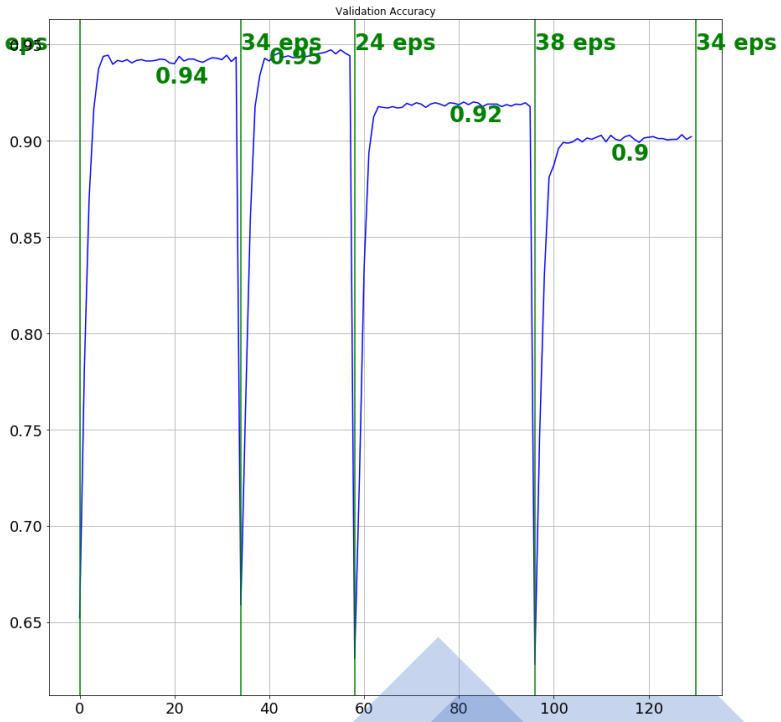
- Adjusting Contrast
- Adding white noise to all pixels:
- Check scaling (rescale)
- Horizontal flip
- Phase translation

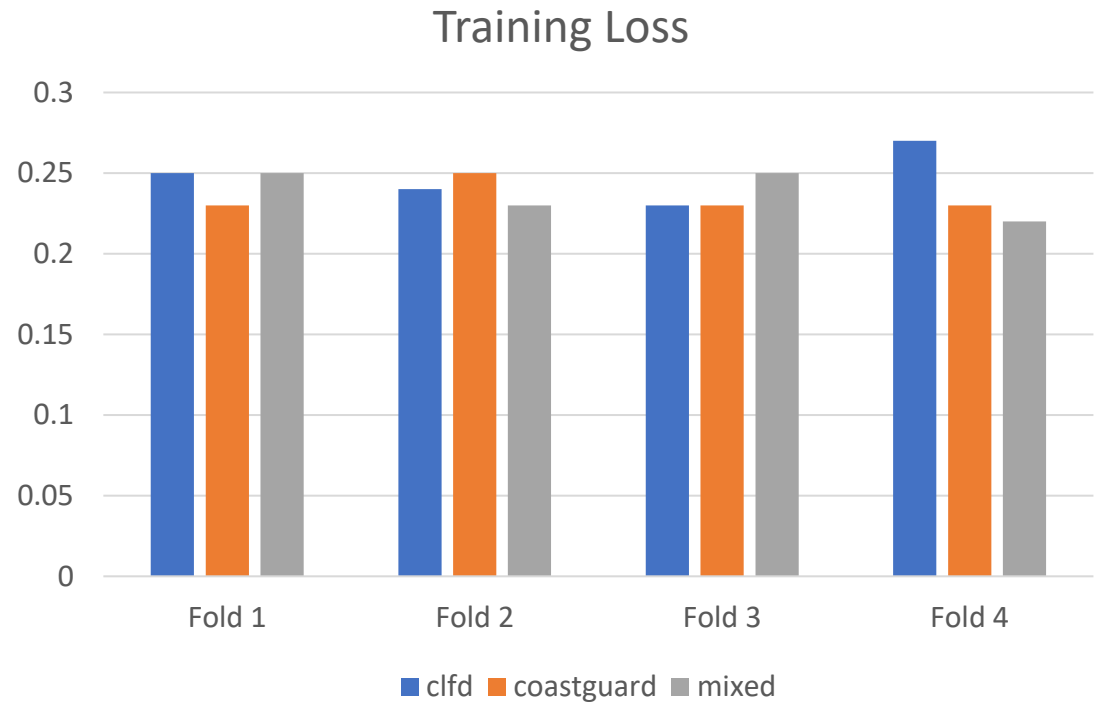
./training_output/new_k_fold/06_4_fold_2ksub_cg 200eps 180 stps per ep

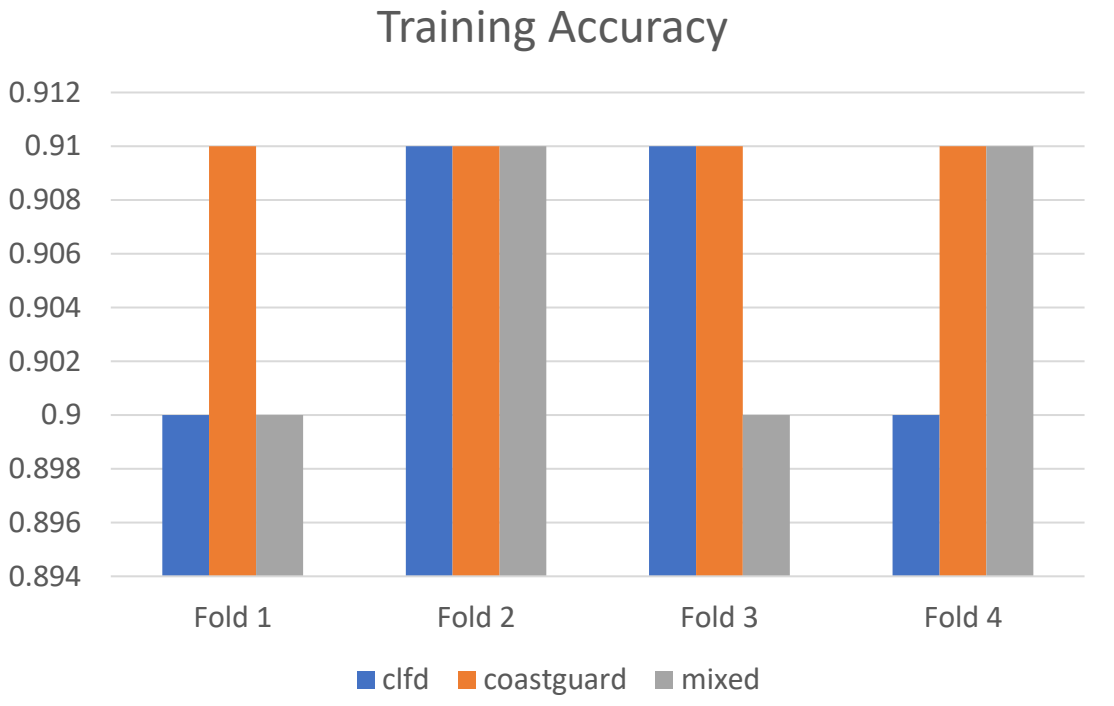


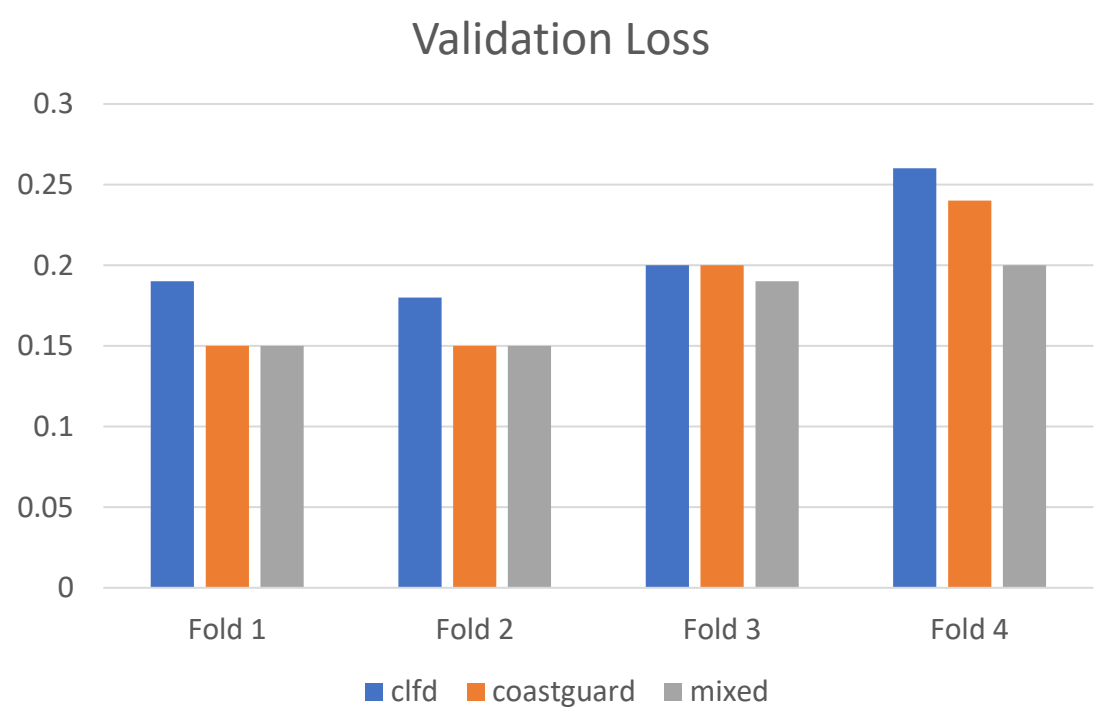


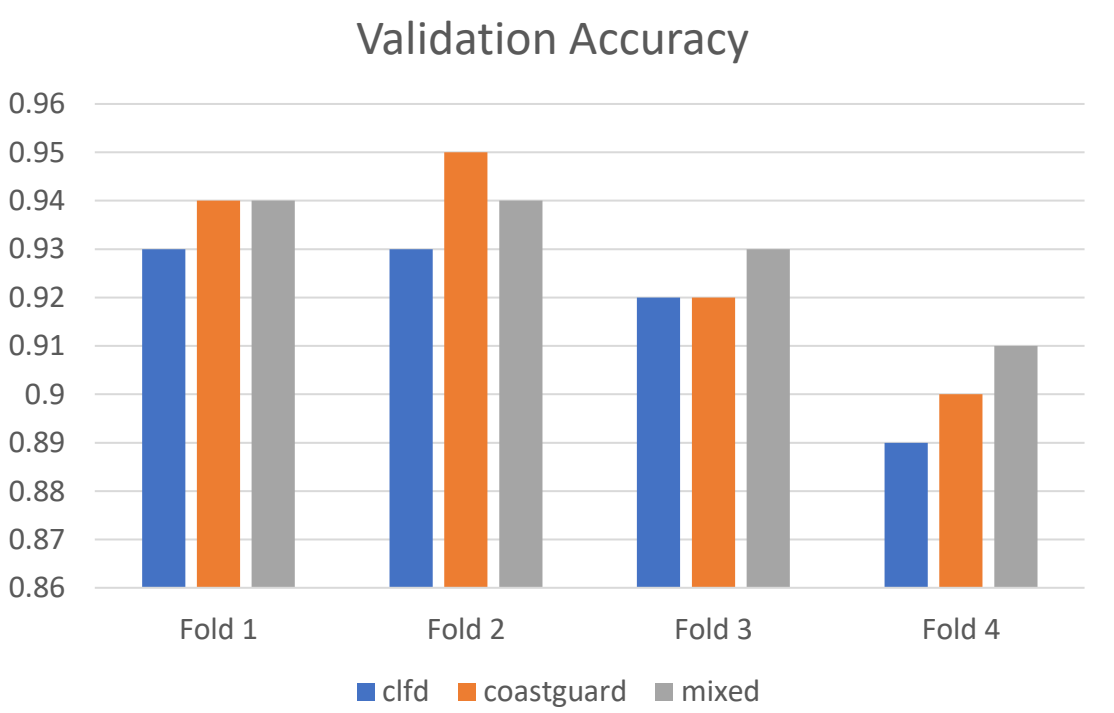





Training Loss
0.3
0.25
0.2
0.15
0.1
0.05
0
Fold 1
Fold 2
Fold 3
Fold 4
clfd
coastguard
mixed

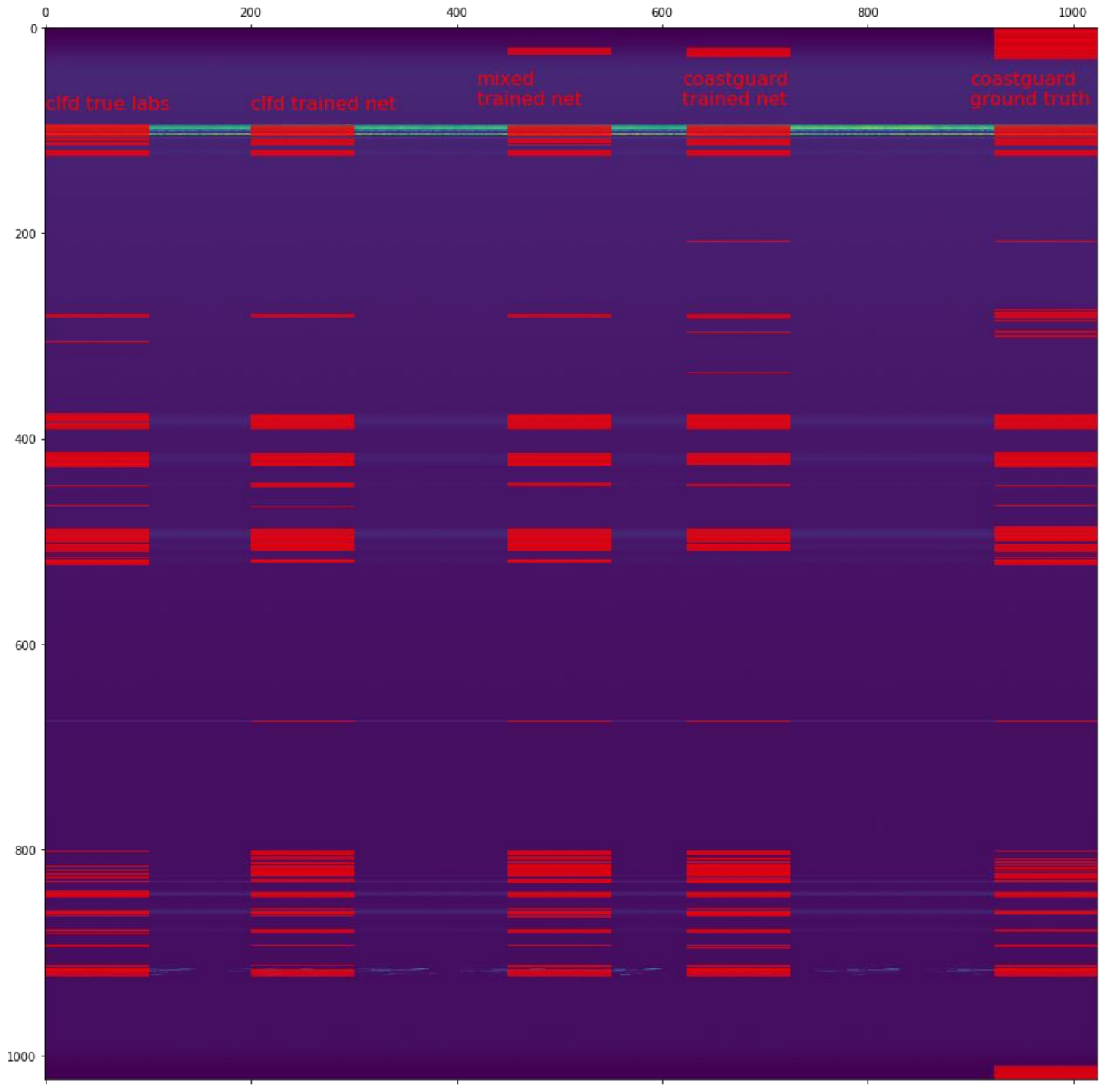

Training Accuracy
0.912
0.91
0.908
0.906
0.904
0.902
0.9
0.898
0.896
0.894
Fold 1
Fold 2
Fold 3
Fold 4
clfd
coastguard
mixed


Validation Loss
0.3
0.25
0.2
0.15
0.1
0.05
0
Fold 1
Fold 2
Fold 3
Fold 4
clfd
coastguard
mixed


Validation Accuracy
0.96
0.95
0.94
0.93
0.92
0.91
0.9
0.89
0.88
0.87
0.86
Fold 1
Fold 2
Fold 3
Fold 4
clfd
coastguard
mixed

Test Results

Clfd	Coastguard	Mixed
AUC ROC = 0.92 PR =0.66 Max F1 =0.63	AUC ROC =0.94 PR =0.73 Max F1 =0.69	CLFD Test Data AUC ROC = 0.92 PR = 0.66 Max F1 = 0.62
		Coastguard Test Data AUC ROC = 0.94 PR = 0.74 Max F1 = 0.69
		50/50 Test Data AUC ROC = 0.94 PR = 0.7 Max F1 = 0.65


clfd true labs
cifd trained net
mixed
trained net
coastguard
trained net
coastguard
ground truth
0
200
400
600
800
1000

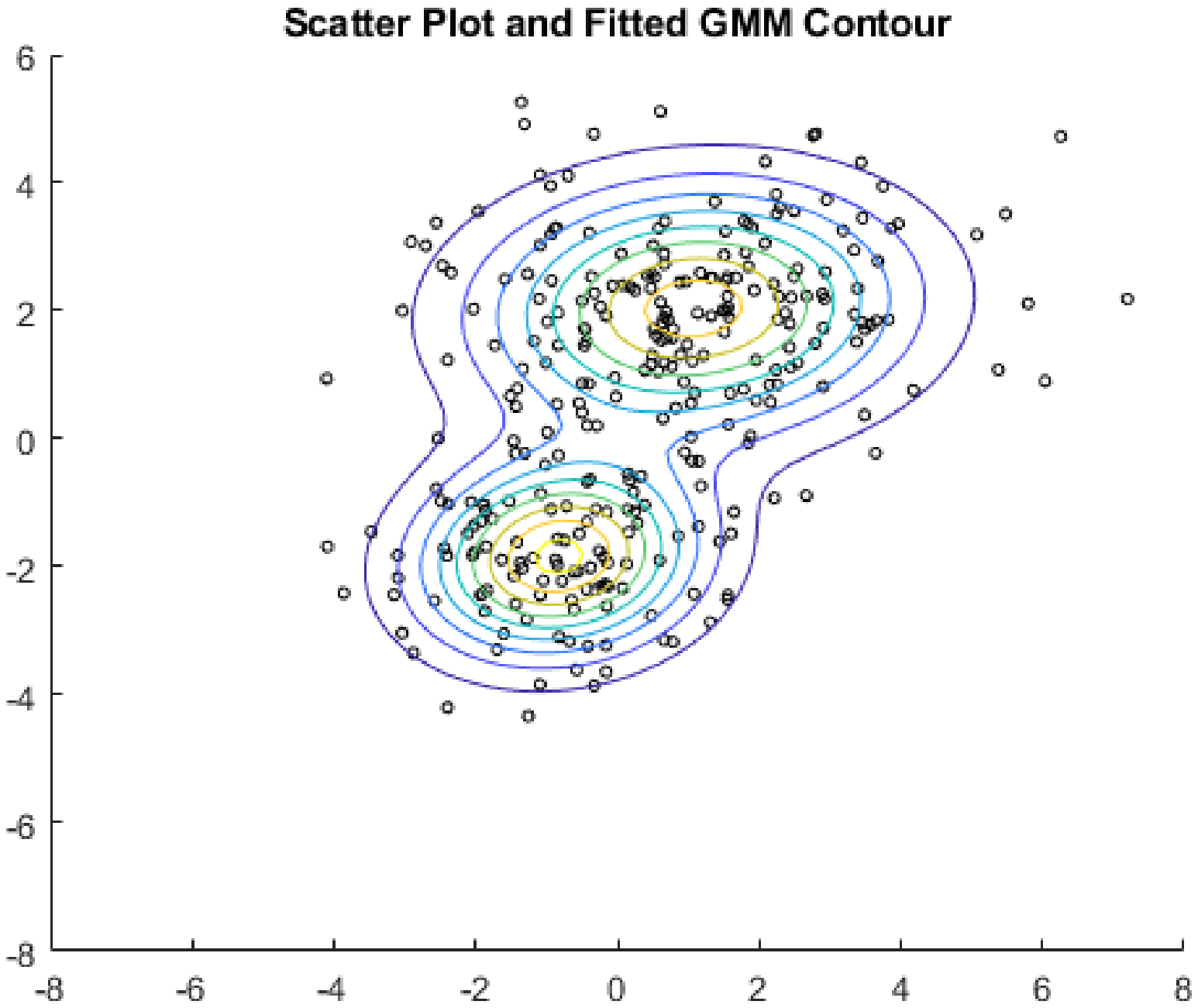
RFI flagging metric

- Anomaly/ Outlier detection from Gaussian Mixture Modelling (GMM).
- Distance = probability density aka Likelihood (L)

1. $PC_j = \{\vec{v}_j \cdot [\vec{P} - a\vec{S} + b\vec{1}]\}$
2. Ngauss of the GMM is evaluated with Akaike Information Criterion (AIC)

$$AIC = 2k - 2\log(L)$$

3. Likelihood is computed from the GMM


Scatter Plot and Fitted GMM Contour
6
4
2
0
-2
-4
-6
-8
-8
-6
-4
-2
0
2
4
6
8

Future Work

- Compare RFI metric results
  - Rank flagging based on likelihood.
- Deeper Networks
- More Data
- More Augmentation
- Fine tuning labels

Thank you