**To Dos**

* Fix up pipeline on Local
* Fix up pipeline on Tacc
* Make diagram of pipeline
* Class diagram of RASR
* Diagram of head
* Scoring metrics (null, all fields, classifier)
* Build new classifier
* Random forest work
* ~~Fix CUDA~~
* ~~Train on colab~~
* ~~Visualize & add in Null data to test data~~
* ~~Check for contrails~~
* ~~Verify pyart resolution with scale~~
* ~~Calculate area of detection - Should be able to get coordinates from detect..~~
* Write up research plan
* ~~Continue to draft research plan~~
* ~~Make class diagram of RASR~~

4/29

* Train LSTM
* Test Classifier model from TACC
* TACC pipeline
  + DT model?
* Make diagrams for presentation
  + Head/full model
  + RASR on TACC

4/24

* Pipeline on TACC
* Train LSTM model

<https://github.com/PacktPublishing/PyTorch-Computer-Vision-Cookbook/blob/master/Chapter10/Chapter10.ipynb>

Key library : <https://github.com/RaivoKoot/Video-Dataset-Loading-Pytorch>

I cannot get this library to import. Why am i so goddamn bad at programming

4/21

* Meeting with Yash
  + RNN
    - LSTM cell is one layer
    - Only a single recurrent layer
    - Have to pass it an end signal somehow
      * Could use a transformer
      * Would output something as a series of unknown length
    - Trying to figure out how to consistently pass
  + CNN
  + For dataloader
    - Test/train - radars - slices & labels
    - Write custom dataloader?

4/20

* LSTM overview <https://cnvrg.io/pytorch-lstm/>
  + <https://blog.floydhub.com/long-short-term-memory-from-zero-to-hero-with-pytorch/>
* CNN, LSTM <https://machinelearningmastery.com/cnn-long-short-term-memory-networks/>
* RNN Image classification:
  + <https://medium.com/@nathaliejeans/how-i-classified-images-with-recurrent-neural-networks-28eb4b57fc79>
* Questions:
  + How do we prepare the data?
  + Do we not need multiple lstm layers?
  + Do data vectors need to be same length? Is that different than number of images?
  + How to test?

4/12

* Talk to yash about productifying lstm head asap
* Presentation include
  + RASR flow down
  + TAcc pipeline
  + Expectation of how many detections to get per day based on radar sweep count and ARES detections
* Currently preparing data for training on other types of data
  + Need to relabel all sets
  + Sort into validate, test, train
  + Throw onto google

4/10

* Finished scoring on Null
  + Looks like detecto model generalizes worse than we thought
  + Can help by filtering out background by biasing the network to look for normal signals as “None”
* Pulled all data from NOAA for raw fields

4/8

* Scoring on Null
* Train on all Fields
* Train on TACC
  + Conda on TACC
* Conda env and training in TACC idev – unsure if working on slurm submissions
  + Also do not know if folder is shared with Kieth..

4/4

* All pyart fields <https://arm-doe.github.io/pyart/notebooks/the_pyart_radar_object_and_indexing.html>
  + Create alternatives for all detections in all fields
  + Manually pull all radar data & visualize
* Random split here <https://pytorch.org/docs/stable/data.html#torch.utils.data.random_split>

4/3

* <https://gradientscience.org/background/> cool experiment
* <https://www.analyticsvidhya.com/blog/2021/06/binary-image-classifier-using-pytorch/> from scratch
* Classifier written but taking up too much space
* Scoring metrics
  + Train on other dimensions
  + Message yash about memory issue
  + Working on TACC

3/30

* SSL
  + <https://towardsdatascience.com/semi-supervised-learning-how-to-assign-labels-with-label-propagation-algorithm-9f1683f4d0eb> – good ref
  + Label prop
  + Good reference <http://www.acad.bg/ebook/ml/MITPress-%20SemiSupervised%20Learning.pdf>
  + And <https://pages.cs.wisc.edu/~jerryzhu/pub/ssl_survey.pdf>
* Null
  + How does \_model build loss
  + Research ex https://arxiv.org/ftp/arxiv/papers/2008/2008.06986.pdf
* Vanishing gradient problem
* RCNN
  + Setup loss function
  + Propagate backwards through autograd
  + Model & loss function for discussion
    - Read up on this
* To do
  + ~~Test null trainer on local~~
  + Scoring metrics for w & w/o null
  + Train on all dimensions of Data (?)
  + Classifier
  + Get system running on TACC crontab, overview of semi-supervised data
    - Train a RF model to pre-label detections

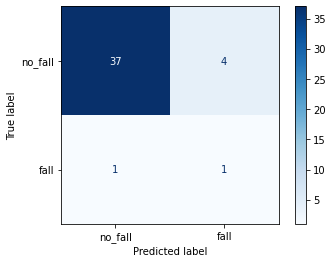
3/29

* Detecto Alternative
  + <https://machinelearningmastery.com/how-to-train-an-object-detection-model-with-keras/>
* Background on RCNN
  + <https://tryolabs.com/blog/2018/01/18/faster-r-cnn-down-the-rabbit-hole-of-modern-object-detection>
* Is zero groundtruth even supported?
  + Fix here <https://github.com/pytorch/vision/pull/1911>

3/26

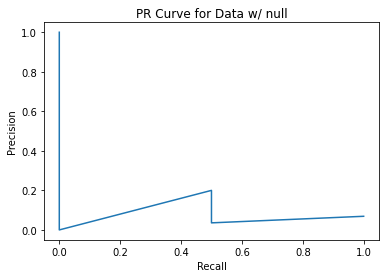
* Stats research
  + <https://stats.stackexchange.com/questions/263539/clustering-on-the-output-of-t-sne>
    - Great article on t sne
  + <https://www.geeksforgeeks.org/difference-between-pca-vs-t-sne/>
* I think detecto is not using the null data
  + Sigh
  + I hate this library so much

3/23

* Steps going forward
  + Train CRNN
  + F1 metrics in testing
  + Baseline demo
  + Semi-supervised model?
  + Make more null data
* CV
  + <https://scikit-learn.org/stable/modules/cross_validation.html>
  + Grid search <https://scikit-learn.org/stable/modules/grid_search.html#grid-search>
* Conv net
  + <https://colah.github.io/posts/2014-07-Conv-Nets-Modular/>
  + <https://arxiv.org/abs/2005.08144>
* Scoring
  + <https://towardsdatascience.com/evaluating-performance-of-an-object-detection-model-137a349c517b>
  + <https://scikit-learn.org/stable/modules/model_evaluation.html> scikit functions
  + Baseline pr
    - <https://stats.stackexchange.com/questions/251175/what-is-baseline-in-precision-recall-curve>
* 
  + For rough model, with labeled data
* Threshold tuning
  + <https://machinelearningmastery.com/threshold-moving-for-imbalanced-classification/#:~:text=Optimal%20Threshold%20for%20Precision%2DRecall%20Curve,-Unlike%20the%20ROC&text=It%20describes%20how%20good%20a,is%20the%20same%20as%20sensitivity>.
  + IMPORTANT
* Semi supervised
  + <http://vigir.missouri.edu/~gdesouza/Research/Conference_CDs/IEEE_ICCV_2009/contents/pdf/iccv2009_065.pdf>
* TODO
  + Generate null data, test dataset
  + Score model for null and without and compare
  + Work on RNN

PR AUC

0.07616995073891626



0.07616995073891626

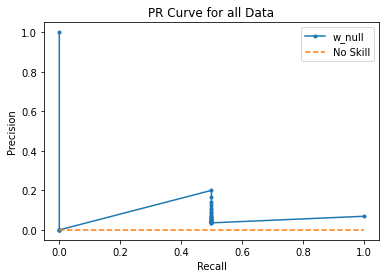
[0.44987318 0.4719926 0.48891723 0.5011641 0.5092642 0.5104745

0.51067805 0.51631474 0.5259497 0.52769285 0.5288597 0.54846716

0.5588337 0.56398594 0.5729823 0.5821309 0.58358824 0.62806416

0.653476 0.6827111 0.6880867 0.6995536 0.7062828 0.74027836

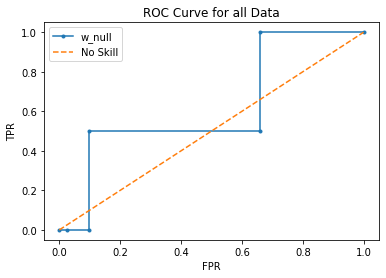
0.7528562 0.75670373 0.7590501 0.77295357 0.86239 ]



0.6219512195121951

[1.86239 0.86239 0.75670373 0.7528562 0.4719926 0.44987318

0.09963012]



3/13

* Meeting w/ Yash
  + What is the next step with the CRNN
* Find F1 metrics with test dataset
* Self supervision
  + Great ref <https://amitness.com/2020/03/illustrated-pirl/>
  + <https://paperswithcode.com/paper/self-labelling-via-simultaneous-clustering-1>
  + <https://paperswithcode.com/paper/unsupervised-representation-learning-by-1>
  + <https://paperswithcode.com/paper/barlow-twins-self-supervised-learning-via>
  + Self classifier here <https://analyticsindiamag.com/a-beginners-guide-to-self-supervised-classification/>
  + BYOL
  + Big agnostic self learners <https://arxiv.org/pdf/2006.10029.pdf>
* Semi supervision
  + <https://amitness.com/2020/03/fixmatch-semi-supervised/>

3/12

* <https://blog.roboflow.com/missing-and-null-image-annotations/>
* Null images added

3/10

* Docker ready to go, firing off tomorrow
  + **NEED TO ADD IN FIXED DETECT**
  + Cleaning up everything was a mess, expecting some readability/usability issues..
  + <https://docs.docker.com/get-started/04_sharing_app/>
  + <https://docs.docker.com/language/python/build-images/>
* Fixed and unit tested refactor
* Create null data and cross validation over the weekend
  + Read papers on self supervised schema
  + Start working on the lstm?
  + Ask clarice at work if she has any advice lol

**3/4**

* Roll off of 3/2
* Dockerizing
  + <https://docs.docker.com/develop/develop-images/dockerfile_best-practices/>
  + <https://docker-curriculum.com/>
  + Scraping on docker <https://github.com/Createdd/Writing/blob/master/2021/articles/dockerizeWebscraping.md#why-dockerizing-your-scraper>
  + Got docker running after an afternoon of work
* To Dos (After docker is up)
  + Create null data
  + Fix and unit test refactor
  + Cross validation??

**3/3**

* Still no luck with docker
  + <https://medium.com/@chadlagore/conda-environments-with-docker-82cdc9d25754>

**3/2**

* No meeting with Keith
* Colab
  + Adding libraries
  + V1 CNN working
* Dataset
  + Need to add null data
  + Clean up training repo
    - <https://detecto.readthedocs.io/en/latest/usage/quickstart.html>
* Dockerizing
  + Sheesh.
  + Debugging
* Script maintenance
  + Run master and get .json outputs

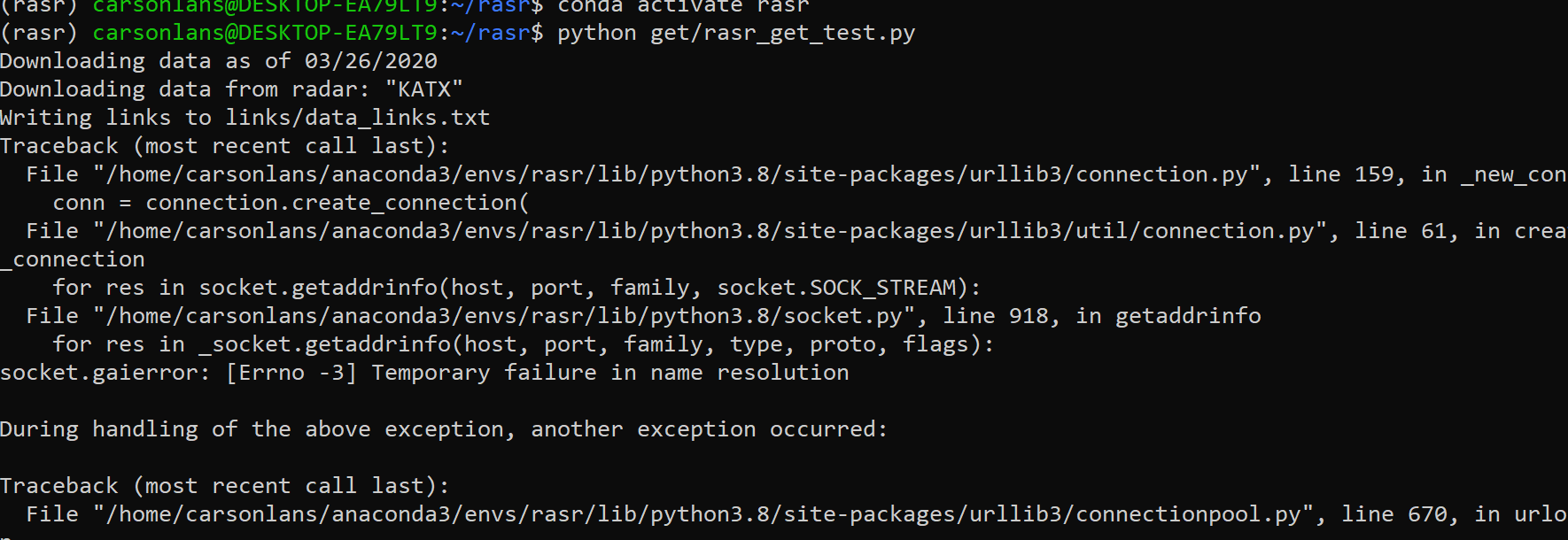
**3/1**

* Dockerizing
  + w/ conda <https://pythonspeed.com/articles/activate-conda-dockerfile/>
* Training on Colab
  + <https://towardsdatascience.com/importing-data-to-google-colab-the-clean-way-5ceef9e9e3c8>
  + Basic models <https://medium.com/paper-club/how-to-set-up-google-colab-colaboratory-for-building-pyro-models-8e51129e772a>
* TO DO - by EOW
  + Attempt dockerization of master (refactor?)
  + Unit test refactor
    - Check out output files for location, bounding data
  + Training on Google colab
* Backburner
  + Null data
  + Verify py art scale
  + Cross validation - kfold split

**2/23**

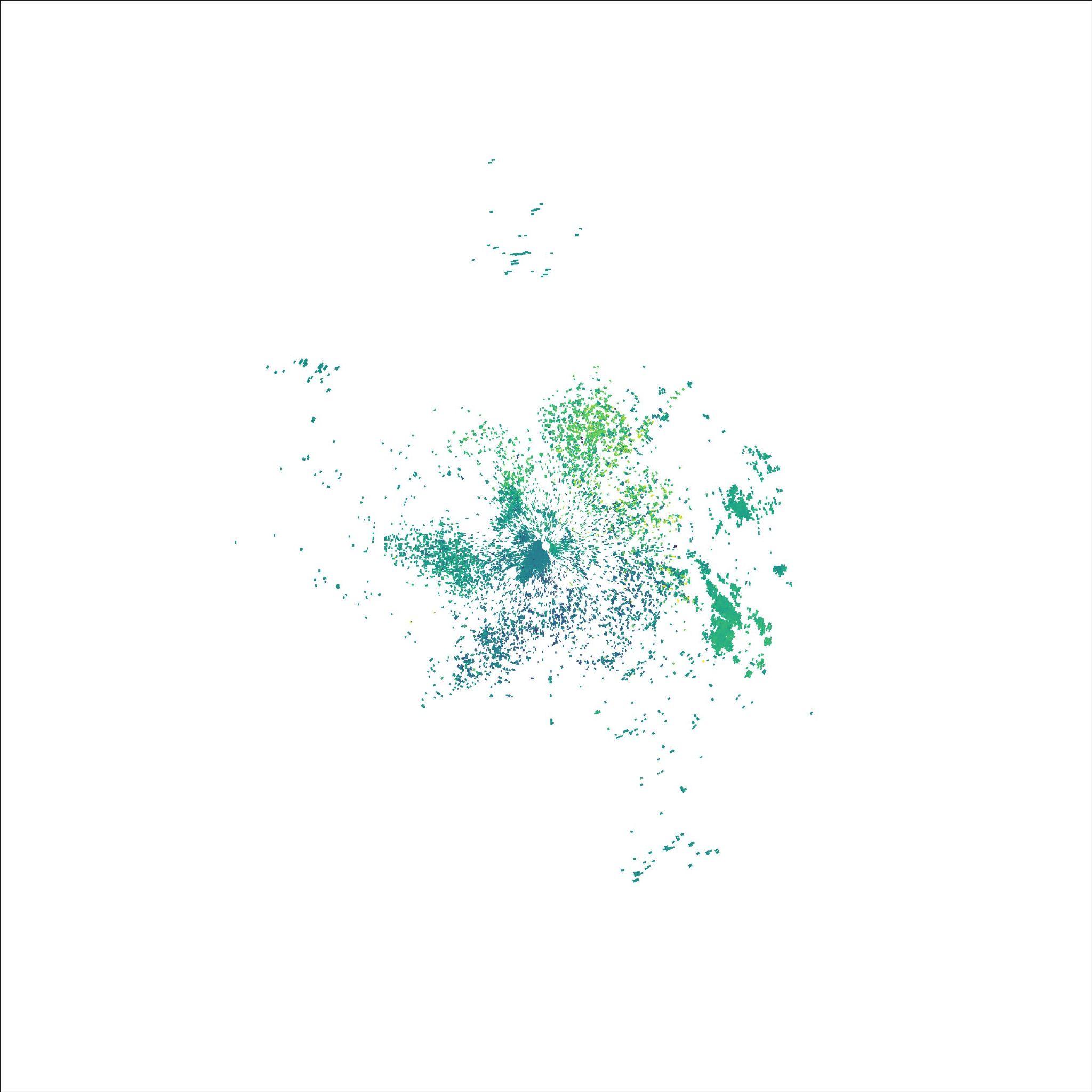
* Meeting with Keith
  + Fix connection bug
* Fix Detect bug
* Need to verify resolution
* Attempt at dockerizing when all done

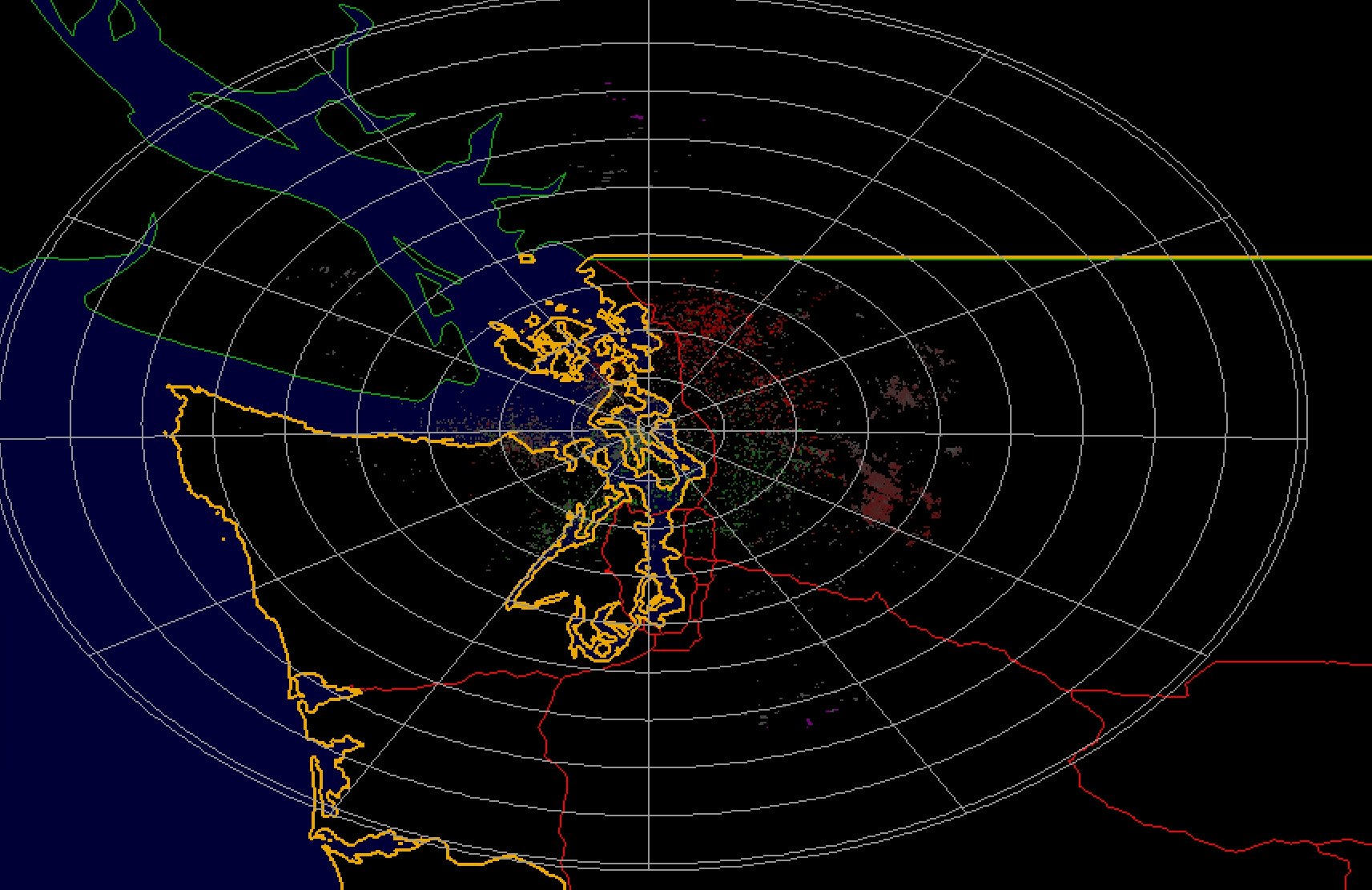
**2/22**

* NEXRAD Data not pulling on my local
  + Working for keith
  + I am getting this issue 
* Spinning up on Google colab
* RASR detect on refactor *also* seems to be buggy..
* TO DOs
  + Dockerize
  + Document (diagram)
  + Fix rasr-detect output
  + Run on TACC (?)
  + Get Null data
  + Run CV analytics

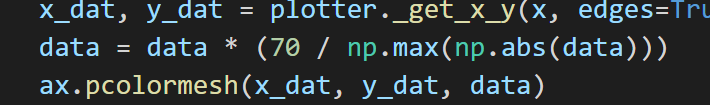
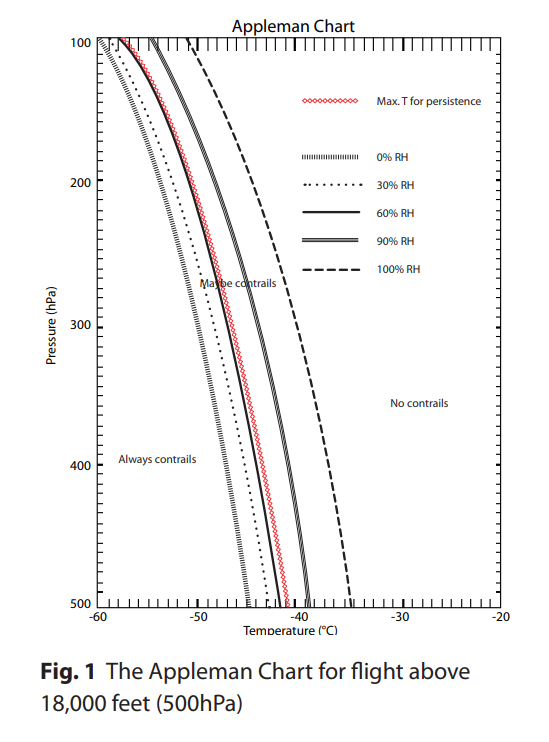
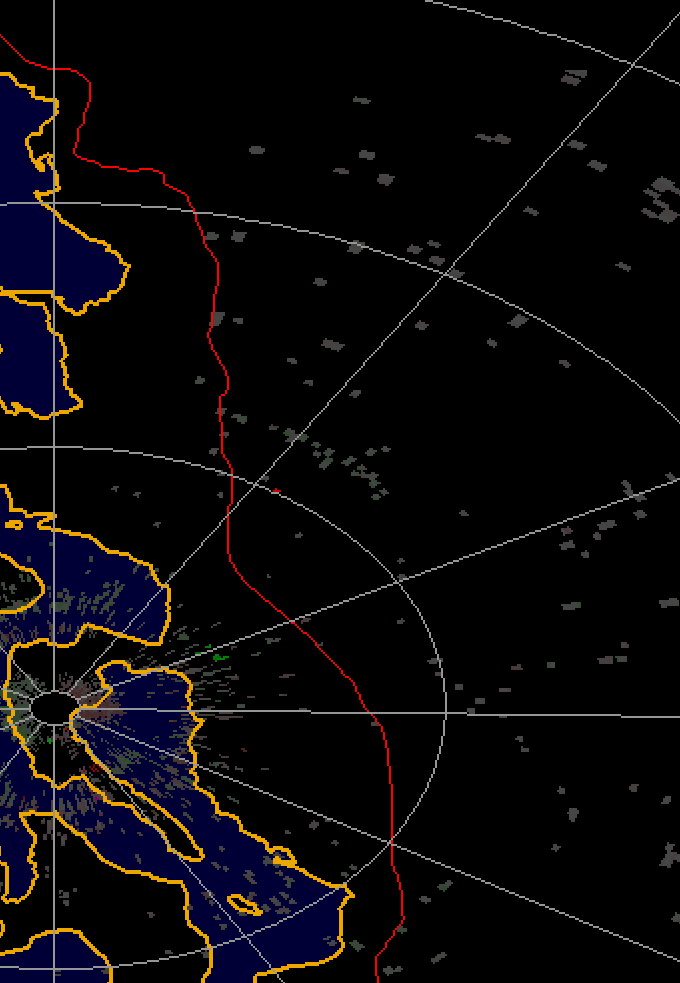
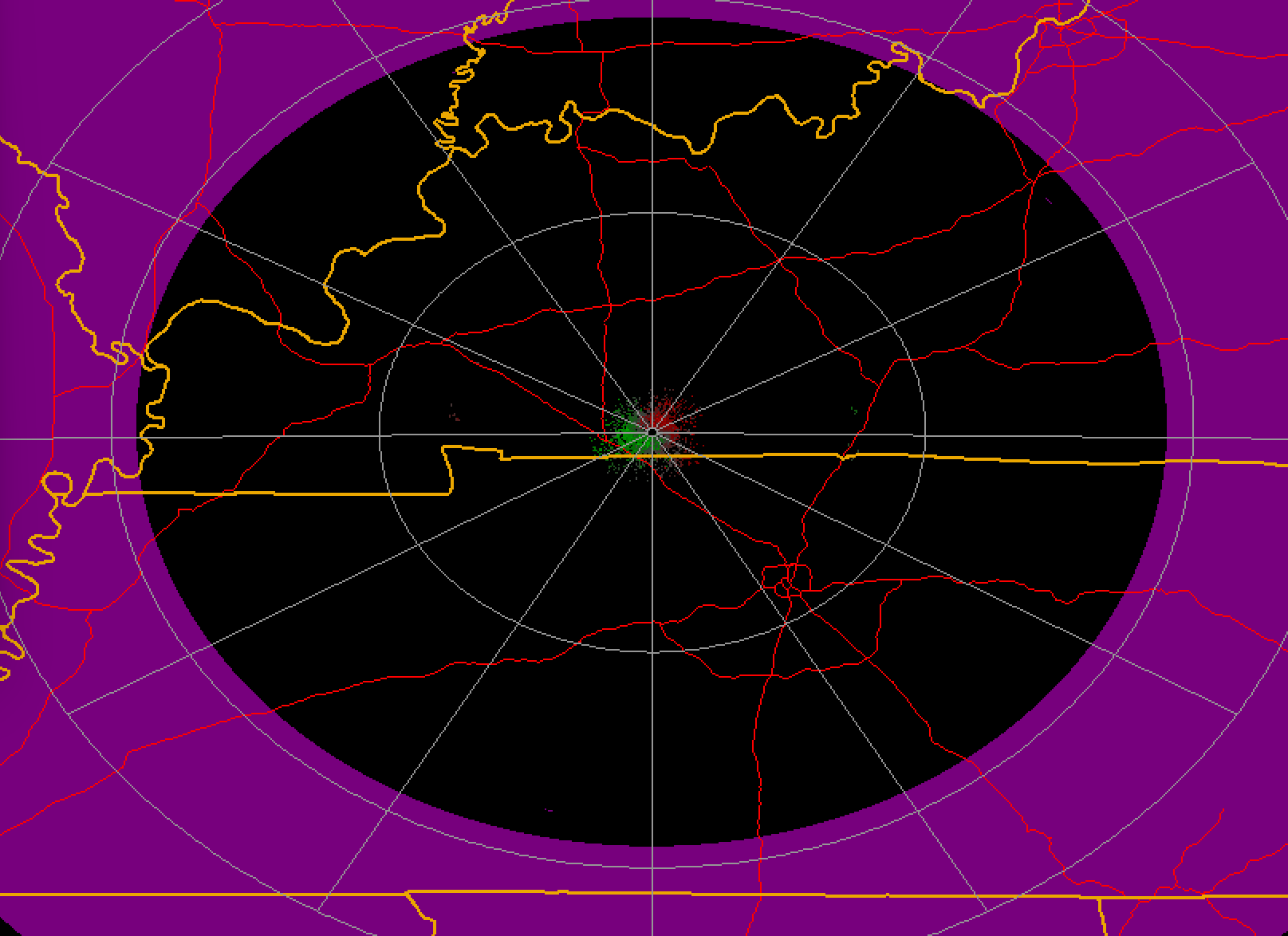
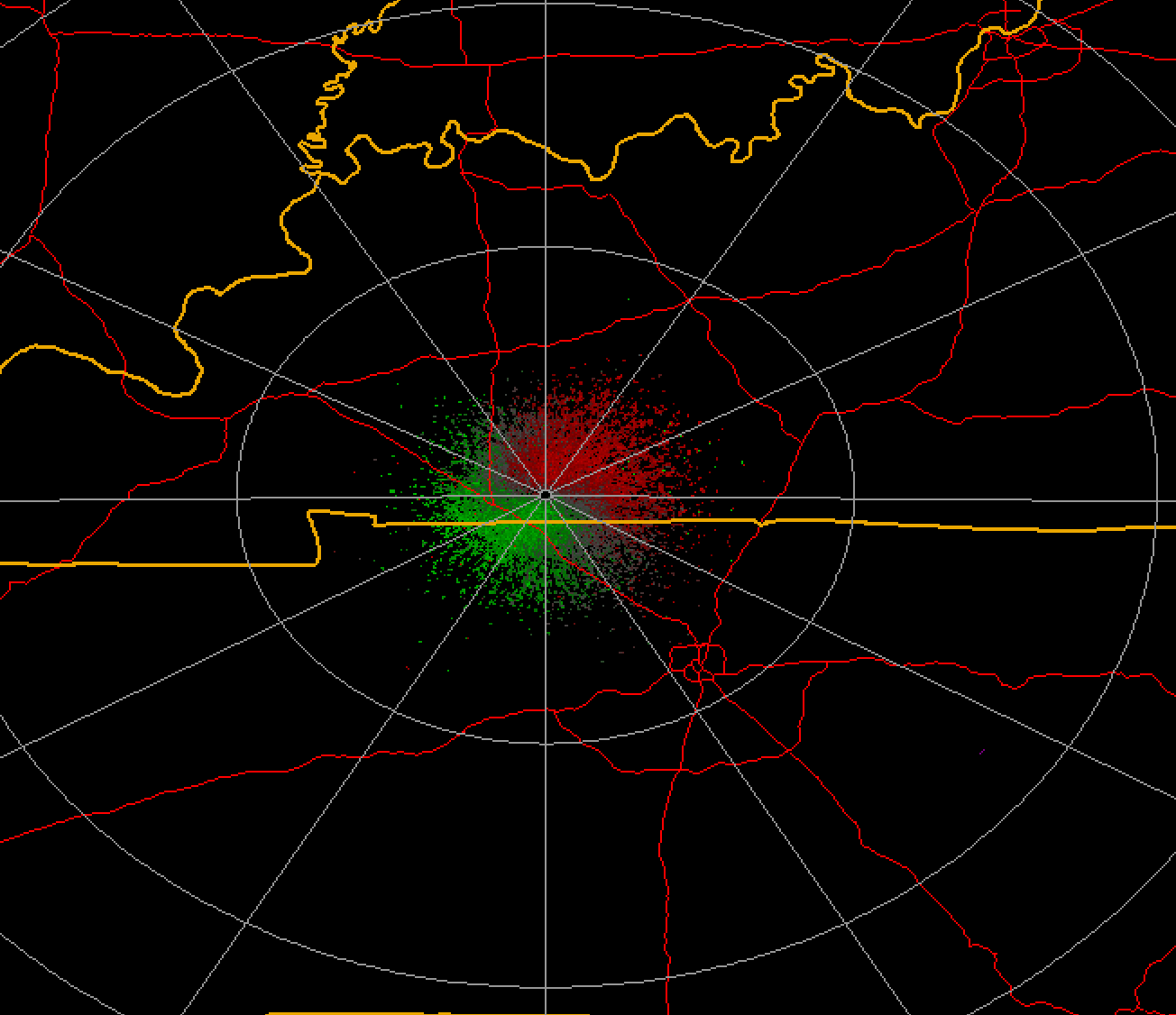
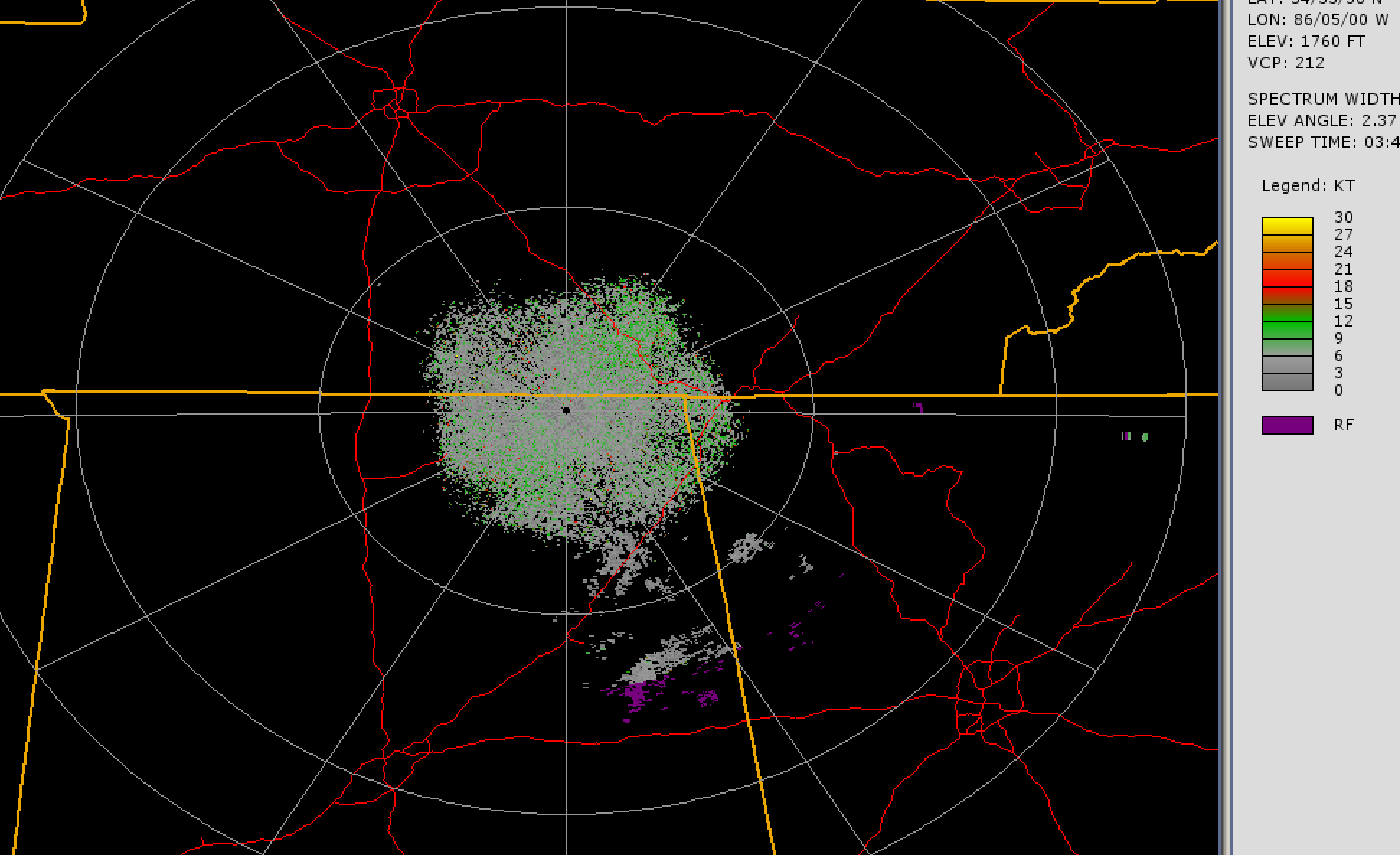
**2/16**

* <https://www.youtube.com/watch?v=chQNuV9B-Rw> - for colab upload
* For null annotations in VOC
  + <https://blog.roboflow.com/missing-and-null-image-annotations/>
  + Need to show image generation working
* Size bounding box in eval.py
  + Looking like cartesian coordinates already found?
  + Spit this out in alert
* Privateer Tag-up notes
  + Questions
    - Documentation standards to make baton pass/integration seamless
  + [LINK](https://docs.google.com/document/d/1mn9ddPQ_cVPI1V22ghhG-YH_75Y0HHxhFibogFshCtM/edit) to doc & action items
* Pulled the number from my directories from RASR's last run, a single site's data (there are 159) in a 24-hr period (this is the interval I've been running RASR) is about 0.5 GB. For anchoring us moving forward.
* There's certainly a trade-off of the ingest interval & compute resources for running the algorithm, I will try to dig into that to save Mike some time in the future.
* Image res comparison:
  + KATX at 1.32 deg sweep





Same data from NOAA Toolkit

* Notably, NOAA toolkit has a few more sweeps/slightly different labels than given by py art
* PyArt resolution documentation:
  + <http://arm-doe.github.io/pyart/API/generated/pyart.graph.RadarDisplay.set_limits.html#pyart.graph.RadarDisplay.set_limits>
  + Dat\_to\_img sets limits at 250 km square – radar res goes up about 50km more for longest cases
  + <https://arm-doe.github.io/pyart-docs-travis/_modules/pyart/graph/radardisplay.html> Functions here
  + Images have location data in RASR that is currently unused when converting raws to images
  + 
  + Script scales everything to max value of 70 (for velocity scale)
    - Why (?)
  + Pyart resolution seems to be the same, aspect ratio unmodified in script
  + <http://arm-doe.github.io/pyart/notebooks/the_pyart_radar_object_and_indexing.html>
    - Helpful
* Contrail checking
  + Literature review
    - <https://www.connectedpapers.com/main/b0781b3746a6fc429fe1d62bdd5e13f6fde112e7/Operational-detection-of-contrails-from-NOAA%20AVHRR%20data/graph>
    - 
    - 110 Hpa at 50,000 ft
    - Based on Air traffic patterns, it is difficult to consistently discern contrails from the NEXRAD data alone
    - 
    - From Kentucky site, does not look like Radar consistently goes to 18,000 feet
    - 
    - 
    - ^here range ring is 90 km, at 2.4 elevation is roughly 20,000 ft elevation
    - 
    - Here in huntsville, up to about 20,000 feet is detected, but no clear pattern of contrails can be seen
      * Where the display is colored pink (coded as "RF" on the color legend on the left side), the radar detected an echo but was unable to determine the wind velocity, due to inherent limitations in the Doppler radar technology. RF stands for "Range Folding". <https://www.wunderground.com/prepare/understanding-radar>
    - It is concluded that simple relations based on radar reflectivity will likely yield ambiguous estimates of ice mass content because of characteristic vertical variations in cloud content that evolve with cirrus growth. The weak signals (<~-40 dBZ) produced by the micron-sized particles present in cirrus cloud-top enerating layers, and in persisting contrails, make their detection by most radars problematic.
  + Seems to be do-able with GOES/Lidar fusion
* TACC automation
  + Conda environment not working on bash
  + General architecture:
    - Schedule\_rasr.sh is called by crontab, calls rasr\_TACC
    - rasr\_TACC contains the slurm job, sends to single compute node on frontera
    - run/rasr\_activator is a bash script containing the full rasr suite of commands called in conda environment
    - Diagram [here](https://app.diagrams.net/#G1sA-p1rzuXuD2B7vhzLwxS1otOJkrnPGX)
    - RASR\_get doesn’t seem to be working
      * <https://stackoverflow.com/questions/14132789/relative-imports-for-the-billionth-time> great note on relative imports
* Training locally/colab
* Modifying output info/alert
* Containerizing
  + <https://pythonspeed.com/articles/activate-conda-dockerfile/>

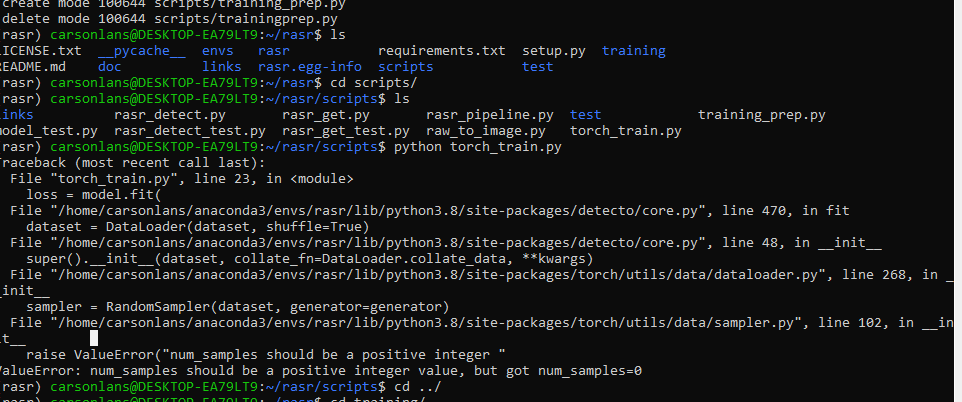
**2/15**

* <http://arm-doe.github.io/pyart/> pyart resolution
* Trying to throttle down the batch size

**2/11**

* Research plan
  + <https://journals.ametsoc.org/configurable/content/journals$002fapme$002f44$002f5$002fjam2235.1.xml?t:ac=journals%24002fapme%24002f44%24002f5%24002fjam2235.1.xml>
  + <https://journals.ametsoc.org/view/journals/mwre/130/4/1520-0493_2002_130_0852_hearfd_2.0.co_2.xml>
* CRNN
  + <https://www.analyticsvidhya.com/blog/2020/11/a-short-intuitive-explanation-of-convolutional-recurrent-neural-networks/> explainer

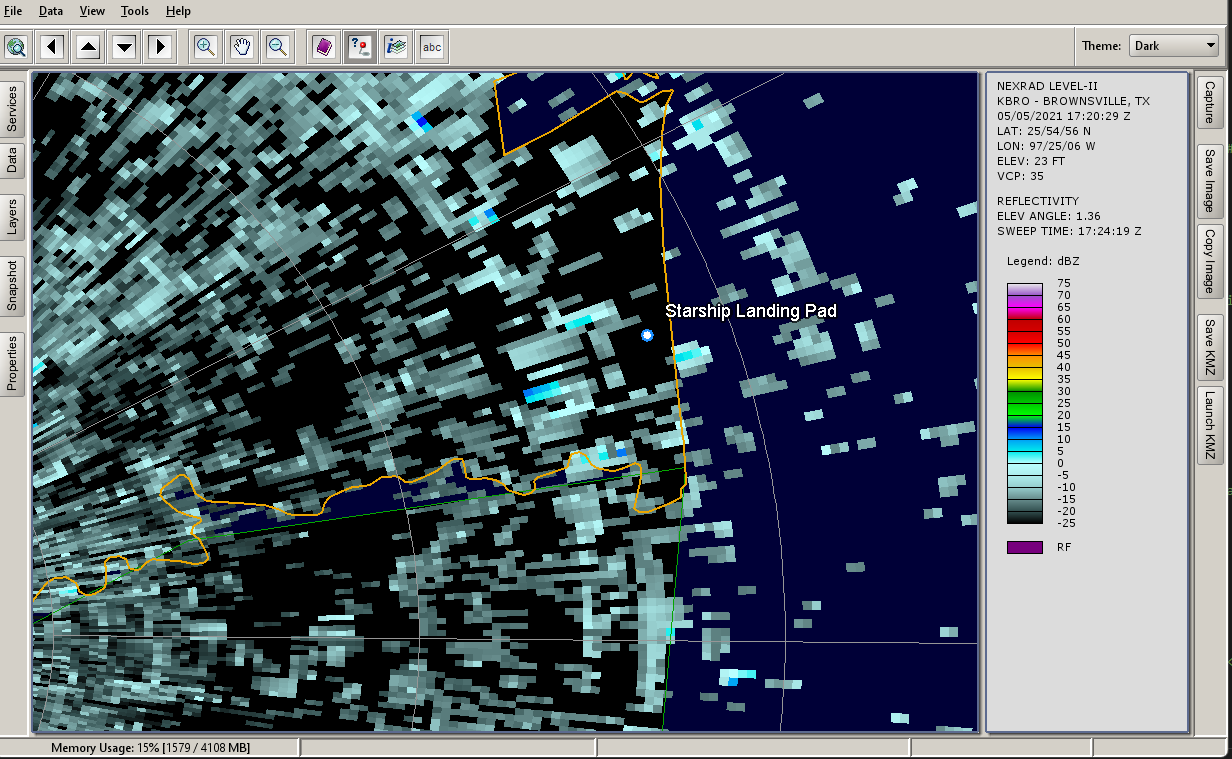
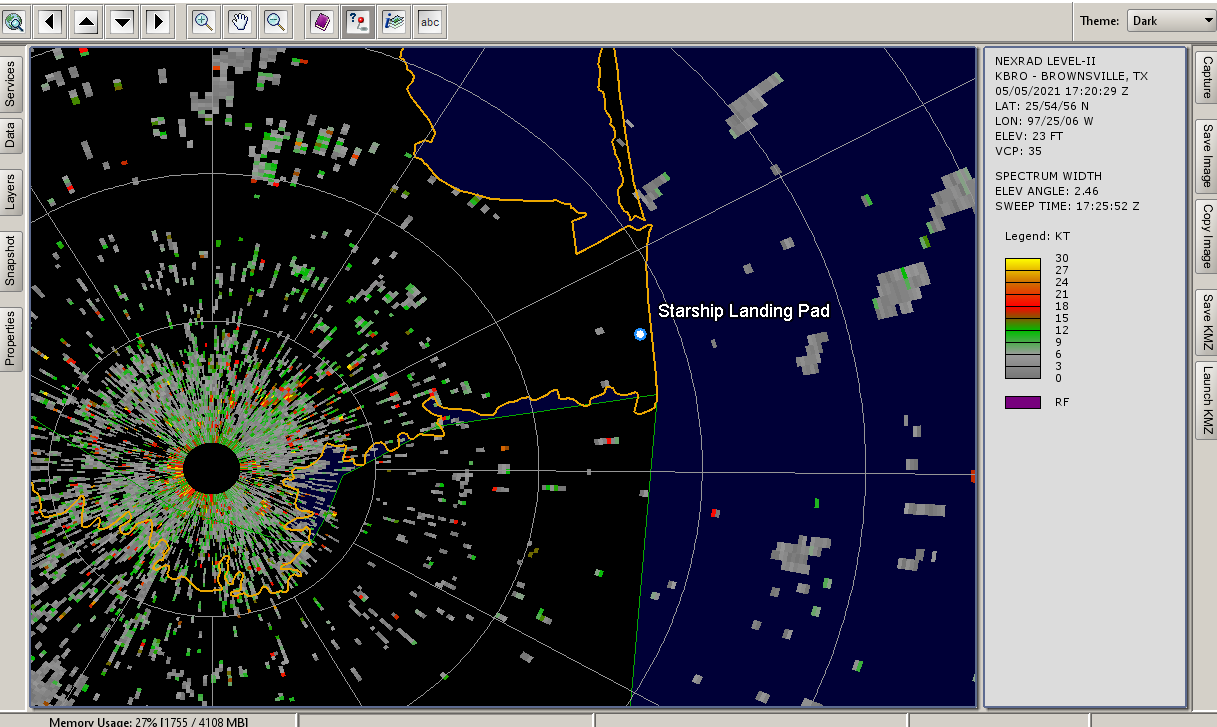
**02/10**

* Training running into new error after getting CUDA to work (?)
  + 

**02/09**

* Met with Keith, discussed classification techniques
* Precision recall might be helpful here, ROC curves useful for when we are training on normal conditional probabilities
  + <https://stats.stackexchange.com/questions/7207/roc-vs-precision-and-recall-curves>
  + This [PR] is because it directly answers the question, "What is the probability that this is a real hit given my classifier says it is?"
  + <https://www.biostat.wisc.edu/~page/rocpr.pdf>
    - Space is highly skewed
  + <https://machinelearningmastery.com/roc-curves-and-precision-recall-curves-for-classification-in-python/>
* Test data needs negatives, we need to be able to generate false positives to understand precision
  + Right now, can only understand sensitivity

**01/26**

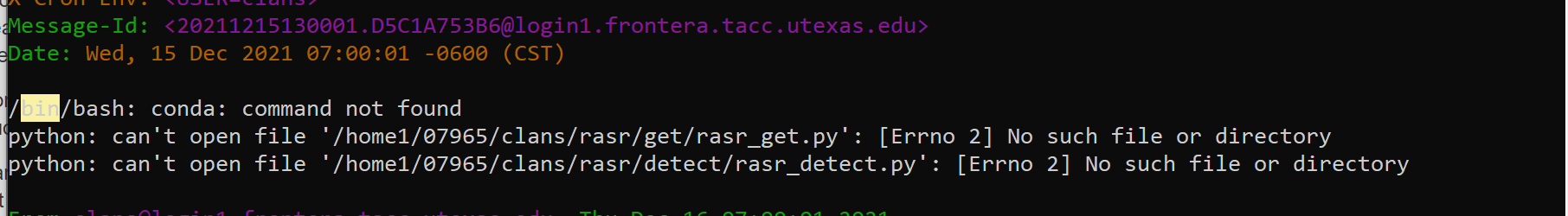
* <https://www.weather.gov/jan/dualpolupgrade-products>
  + Enumerate all parameters
  + How can any of these be useful in the detection algorithm?
  + Play around and see what you can find?
  + Look for contrails, known orbital debris
* Can I see starship?
* 
* Also something weird I noticed with the data is that sometimes NOAA puts the extreme low and high elevations at weird time increments in with bins that aren't the same. so.. I'll have to watch out for that sigh
* [2:41](https://ut-astria.slack.com/archives/C030M274NUQ/p1643229664784989)
* and here is spectrum width about 110 seconds into starship flight, about ~1km below where it was (edited)
* [2:41](https://ut-astria.slack.com/archives/C030M274NUQ/p1643229669562419)
* image.png
* [2:42](https://ut-astria.slack.com/archives/C030M274NUQ/p1643229758296499)
* Reflectivity a little less sparse, but still NADA above starbase.

**01/25**

Meeting with privateer, make visualizations and bounding boxes with confidence and scale

Figure out how to integrate/build to integrate with their data pipeline

**01/20**



Need to figure out path for files in TACC and how to store

Yash cleaned up rasr\_get\_test, not downloading any data on vscode (NVM it is)

Getting geojson not found error when running rasr\_detect\_test

Model trained with error, not .pth saved. (ugh)