Jun training data

* In “flow” folder
  + 947 items
  + Camera A, Camera C, H-Flume: 5/10/2023-5/12/2023
  + Camera A, B, C and H-Flume, 5/24/2023
  + Camera B, 6/4/2023-6/6/2023
* In “no flow” folder
  + 1600 items
  + All cameras, April-June 2023?
* In “snow” folder
  + 1158 items
  + All cameras, April 2023

Training procedure

* 20% validation split:2964 files for training; 741 files for validation
* In evaluatefromfile.py, there is a parameter to batch the data (32), does that mean that the evaluation is being done differently than training?

|  |  |  |
| --- | --- | --- |
| Model | Val loss | Val accuracy |
| Edge128batch24 | 0.011784999631345272 | 0.9959239363670349 |
| Edge128batch128 | 0.017543483525514603 | 0.9959239363670349 |
| Edge128batch512 | 0.7035024762153625 | 0.70923912525177 |
| Edge256batch128 | 0.059860970824956894 | 0.9850543737411499 |

The “edge128, batch24, epoch30” model trained on 12/5 had a reported val\_accuracy of 0.9987 and val\_loss=0.0149. Running it through the “evaluate from file” routine gave loss 0.0003 and accuracy 1.0; need to check the effect of the batch and random seed on the evaluation process.

* Unlabeled directory on OneDrive has pictures from 8/13/2022 to 8/17/2022

From Santi:

Hi, after thinking more about it, I realized that it would be more interesting to analyze the performance of the model with the night view. A common problem we faced when analyzing the photos is that

1) A lot of streamflow events occurred during night.

2) The night view mode would be activated when it as raining and maybe too cloudy.

 For this I organized the photos and gifs of some representative events for both Camera A and Camera B that can be found in this folder:

[AGU Streamflow Events](https://o365coloradoedu.sharepoint.com/:f:/s/CEAE-HUB/EnniYzxfV2ZOmJQAFoHEUZkBwoUeAv0ZedEdeEj2UFAZ4w?email=joseph.kasprzyk%40colorado.edu&e=n170pv&xsdata=MDV8MDJ8fDc3MTllNTI3MmNiYTQ5NjYzZjdhMDhkZDE1YmU1YjA1fDNkZWQ4YjFiMDcwZDQ2Mjk4MmU0YzBiMDE5ZjQ2MDU3fDB8MHw2Mzg2OTA2Mjk0NzE1MTg1ODZ8VW5rbm93bnxWR1ZoYlhOVFpXTjFjbWwwZVZObGNuWnBZMlY4ZXlKV0lqb2lNQzR3TGpBd01EQWlMQ0pRSWpvaVYybHVNeklpTENKQlRpSTZJazkwYUdWeUlpd2lWMVFpT2pFeGZRPT18MXxMMk5vWVhSekx6RTVPamt3TVdRME5EbGtaV0kxWmpRME5qTTVPV0V5WkRJMFltUXpPR0ZqWmpjMVFIUm9jbVZoWkM1Mk1pOXRaWE56WVdkbGN5OHhOek16TkRZMk1ETXlNakk1fGY1MjBmNWIwYzE4NDQyODQzZjdhMDhkZDE1YmU1YjA1fDBjYjIxYzM5Y2NmZjQ1ZmU4OTUwOWQ5MjE1NWE5MWY0&sdata=UWFHU1BKNHA1ZWRMT3N6K0pQUjVnZlNqdkpNRndSanlIdnNRcGJFNTFQVT0%3D&ovuser=3ded8b1b-070d-4629-82e4-c0b019f46057%2Cjoka0958%40colorado.edu)

 Camera A

Night Events

Storm 119. It is a very long event with no vegetation but it is mostly in night view.

Storm 108. It is a very long event with medium vegetation and it is all night view.

Day Events

Storm 97. It is a very short event, with light vegetation and it happens during the day.

Storm 94. It is a very long event with light vegetation and it happens during the day.

Both Day and Night

Storm 109. It is a long event with medium vegetation that starts during the day and extends all night.

Camera B

Nigh Events

Storm 100. Very long, night view.

Storm 118. It is a short event with night view.

Day Events

Storm 109. Not very high flow, day view.

Storm 112. Long event that occurs for a long period during the day and then extends to the night.

Not Sure

Storm 116. The event happens with vey low light, may be night view.

When I mention Storm ### I mean the number of Column C ("Storm Number") of the Excel spreadsheet

* New labeled dataset
  + For training/validation…
    - Storm 100
    - Storm 112
* Unlabeled data
  + Events found: 2,666 predicted flow of 14,596 datapoints
    - 8/13/2022 to 8/17/2022
    - 6/1/2023 to 6/9/2023
    - 6/13/2023, time 1950 only
    - 6/22/2023 to 6/28/2023
    - 7/1/2023, time 1927 only
    - 7/6/2023
    - 7/8/2023 to 7/17/2023
    - 7/19/2023 to 7/31/2023
  + True Positives
    - 8/15/2022 14:40 - true
    - 8/16/2022 10:35 - true
    - 6/1/2023, 12:50 - true
    - 6/3/2023, 21:10 - true
    - 6/5/2023, 14:00 - true
    - 6/6/2023, 15:40 - true
    - 6/22/2023, 15:30 - true
    - 7/6/2023, 16:35 - true
    - 7/8/2023, 15:55 - true
    - 7/20/2023, 5:10 - true
    - 7/21/2023, 20:40 - true
    - 7/24/2023, 14:45 - true
  + False Negatives
    - 6/16/2023, 12:35 - false
    - 6/21/2023, 22:05 - false
    - 6/30/2023, 10:45 - false
    - 7/4/2023, 12:55 – false
    - 7/7/2023, 18:50 - false
  + Events not in Camera B
    - 6/10, 11, 12, 15, 16, 21: events not in Camera B
    - 7/31/2023, 15:20, not in camera B – false positive

I trained a model using only Camera B, "storm 100" and "storm 112". From the images that were in Santi's folder, I manually labeled the pictures from these storms. This was treated as a training/validation set where 80% was used for training and 20% was used for validation. The validation results were really good; ~97% accuracy or something like that

I used this model on an unlabeled dataset from Camera B. 14,596 pictures were selected: a few days from August 2022, and the entire months of June and July, 2023. The model predicted that 2,666 of the images had flow.

Because we don't have labels for every single image, I can't say something definitive about the accuracy. Instead, I used the manually classified storms from Santi in that period to judge the performance of my model. I assumed that if there was one datapoint within a particular day that was modeled as having "flow," that would be defined as the model "capturing an event."

There were 17 manually classified events; 12 were properly captured by the model and 5 were missed by the model. There were also many days where the model said there was flow and it does not appear that there was an event that was manually classified.