Danelle Cline Sorted By Project

500055 - Video Annotation DM

901009 - CANON

- Meeting July 9 (Week 28)
- Discussion w/Kevin on ODSS script migration. (Week 44)

901502 - Passive Acoustic Sensing

- Teleconference with AWS about open data sharing program (Week 5)
- Shared notebook examples and helped setup notebook in GCP for Cal Poly student Daniel DeLeon to train models. (Week 5)
- Cleaned up wandb model reports organizing by project number instead of name (Week 5)
- Applied and was approved by AWS open data program to freely share acoustic data - all 95 TB! (Week 6)
- Assisting student Daniel DeLeon with embedded TensorRT code project. Having success using Google Colab+GCP for this effort. (Week 6)
- Was invited and joined the OrcaSlack channel (Week 6)
- Testing Keras inference speed improvements (Week 6)
- Instrumented topic model with wandb logging, ran two sweeps for K-means over different PCEN parameters, first broad then fine. Next steps are to run full model sweeps with best PCEN parameters. (Week 9)
- Working on streamable PCEN implementation notebook (Week 10)
- Retraining blue whale call classifiers with improved PCEN parameters (Week 10)
- Reviewing summer intern applications (Week 10)
- Restarted rsync and Shoutcast audio stream after reboot following IS patch failures. Added reboot check to crontab. (Week 13)
- Found and fixed an issue in decimation for edge case with multiple small discontiguous files (Week 17)
- Reprocessing 2020 data with UTC aligned day in preparation for AWS Data Sharing and pre/post COVID soundscape analysis. (Week 17)
- AWS data sharing teleconference to discuss rollout of 2020 data with tutorials (Week 19)
- Computing 2kHz decimated time series with UTC daily files (Week 19)
- Working on HTCondor pool CentOS8 upgrade (Week 19)
- CentOS8 migration of HTCondor pool master machine condormastervm to htcondormaster (Week 20)
- Teleconference with AWS team about open data sharing program (Week 20)
- Weekly team meeting and monitoring of decimation jobs (Week 20)
- Centos8 upgrade (Week 21)
- Finished decimating 5-year time-series and now investigating some 10-minute gaps in decimation likely due to time drift. (Week 21)
- Code optimization of blue whale code (Week 23)
- Found issue with missing data in decimation files. Resubmitting backlog jobs to recompute those. (Week 23)
- McGill and Zhang are working on further optimizations to the decimation filters. This should be useful for various workflows that need different levels of decimated data. (Week 23)
- Fixed an issue with wav cache read collision in decimation job (Week 24)
- Created postgres development database for MARS wav file metadata using prisma/graphql. Really liking the schema evolution and tools in the software stack. (Week 24)
- Added coherence image enhancement to Blue whale classifier and regenerating data for training. (Week 24)
- Reviving topic model code in preparation for student project efforts. (Week 24)
- Student discussions (Week 24)

- Humpback whale song topic model sweeps for Dirichlet hyperparameters: gamma, alpha/beta. Code for that is here: https://bitbucket.org/mbari/soundscape-humpback-topicmodel. Expect many more model sweeps in the coming weeks to fine-tune the final parameters. Looking at coherence in addition to perplexity as a measurement of quality. (Week 26)
- Summer intern discussions on coherence, PCEN preprocessing etc. (Week 26)
- Staged 5-year decimated data in AWS to share with collaborator Will Oestreich on Blue whale A/D call analysis. It was surprisingly fast to upload the entire 2kHz dataset. (Week 26)
- Experimenting with prisma/graphql database with sound metadata. (Week 26)
- Conference with Carrie Wall, an oceanographer who leads the fisheries acoustic archives at the NOAA National Centers for Environmental Information (NCEI). Discussed our blue whale research and methods for automated analysis. Data is in hand and experiments are planned to test with our signal processing and machine learning methods. (Week 28)
- Testing single day case study John identified as a high priority for decimation testing. Day has time drifts and multiple irregular drop-outs. Getting close to a robust solution to handle the messy data. (Week 28)
- Started on testing Yanwu's optimized filter parameters in our decimation notebook. These should result in a speed-up in our current rate limiting step of decimating the high-frequency data. (Week 28)
- Topic modeling sweeps for summer intern and ongoing discussions 3x per week about progress on coherence results. Many thanks to Ben Raanan and Carlos Rueda for helping when needed with this effort. (Week 28)
- Project proposal input (Week 28)
- Tested optimized decimation filter with shorter filters and disseminated to team for inspection. New filters are ~20% faster in the most rate-limiting step for the soundscape analysis. We think the faster decimation is working but needs a time adjustment for the filter lengths which P. McGill will investigate. (Week 31)
- Meetings with summer intern, Madison. She's doing a great job exploring topic coherence methods. Added similarity test for her to find two runs with similar topics based on ARI for her coherence analysis. (Week 31)
- Doing some clean-up of blue whale batch inference code. (Week 31)
- Fixing an issue with dockerized deployment of blue whale model in preparation for deployment on 5yr time-series with VGGskeleton model. Processed a few days for collaborator, Will Oestreich to review; he reports 100% accuracy on test files. Model was around 92% on the validation set. (Week 32)
- Summer intern discussions and drafting up collaboration agreement. (Week 32)
- Ongoing decimation filtering discussions with the team. (Week 32)
- AWS data sharing program correspondence. Many thanks to Kevin Gomes and the IS department for their help as we forge into the great Cloud. (Week 35)
- Added coherence and VGGish embeddings plots of humpback topic model output. The dispersion of the VGGish embedding looks like a better discriminator than coherence for model optimization. (Week 35)
- Blue whale A call predictions across the entire 5-year time-series produced >200K calls. Next week blue processing jobs for whale D call predictions will be put in the queue. (Week 35)
- Troubleshooting issues with streaming audio failure. (Week 36)
- Trying to keep up with the AWS Data Sharing chatter (Week 36)
- Blue whale D call predictions ready to process. Just need to free up GPU resources to start. (Week 36)
- Discussion with Daniel DeLeon about inference tests results on Aria 10 FPGA and ideas for his masters studies. (Week 36)
- Restarted shoutcast stream which stopped after running fairly robustly for some time. There were no errors in the log file so it remains a mystery. (Week 37)
- Fixed email notification on recording machine which was failing due to secondary authentication checks. (Week 37)
- AWS upload throttle test (Week 39)
- Discussion with Cal Poly about potential student projects (Week 39)
- Team meeting (Week 39)
- Testing ported tensorflow to pytorch blue whale classification code (Week 41)
- Assisting with versioning and sharing click data with Cal Poly AI class (Week 41)

- Discussion and planning with John Ryan and Kevin Gomes on how to best capture metadata and share with the AWS Open Data Sharing program (Week 41)
- Working on cleaning up notebooks to share with POGO workshop participants & AWS Open Data Sharing program (Week 41)
- Working on scripts to describe and share our entire 5-year archive in the Amazon Open Data Sharing program. Cloud push and testing planned this week for 2020 data. (Week 44)
- Found a normalization issue in the blue whale call prediction code which reduced the effective prediction accuracy. Reprocessing the blue A and D calls with this patch. Accuracy is now 93 and 95% for each model respectively. Models take approximately 30 minutes to train. (Week 44)
- Finished metadata and upload scripts for AWS Open Data Sharing program initial 2.5 month data set. Upload of raw data underway. So far everything has worked well. Prototyped web design for communicating with ITD external web changes to serve data tutorials. John and I are working on prototyping notebooks in Colab and EC2 Jupyter stacks to demonstrate how to use the data. (Week 46)
- Humpbacks have been busy singing in the evenings. Check-out the live stream if you need some relief from the chaos. (Week 46)
- Call planned next week with folks at the Earth Species Project to discuss our plans for data sharing. (Week 46)
- Team meeting (Week 51)

901603 - Video TAG: Vis Ocean Chng

- Project abstract review (Week 23)
- Abstract reviews (Week 24)

901809 - Deep Sea Bioinspiration

- Cleaned up wandb model reports organizing by project number instead of name (Week 5)
- Requested 2020 task list from K.Katija for allocated 240 hours (Week 6)
- Generated improved video of bathos detection for external website (Week 13)
- Need to get details from the project on model training needs (Week 17)
- Need to inquire again about model training needs. All went pretty quiet with this project over the Covid period. (Week 23)
- No requests yet on allocated time (Week 32)

901915 - Improving MBARI Data Impa

- Meeting to discuss TimescaleDB and STOQS (Week 6)
- Meeting (Week 14)

901915 - Improving MBARI Data Impact

902002 - Fathomnet Detect and Classify

- Tested a few of the APIs. Had some difficulty fetching images. Tar of the archive would probably be easier. (Week 36)
- Started compiling notes about improvements and nice-to-have features (Week 37)
- Interview scheduled 2pm 9/24 to speak to web designers (Week 39)
- Tested code to remove black borders and deinterlace (Week 41)
- Noticing some duplicate annotations (Week 41)
- Need a way to filter out equipment and transect versus close-ups (Week 41)

902005 - VARS Annotation Assistance

- Lead project “kick-off”. Great progress already on both Sagemaker and ML aware video player development!! (Week 5)
- Feb 12 meeting planned to brainstorm with video lab about workflow. (Week 5)
- Working on bug in tensorflow detection model inference served in MLFlow (Week 5)
- Cleaned up wandb model reports organizing by project number instead of name (Week 5)
- Team is working towards benchmarking cost/performance of cloud compute and preparations for testing with new ML-aware video player (Week 6)
- Meeting scheduled for February 12@2pm with video lab to brainstorm about the future use of machine learning for enhancing and augmenting video and image analysis at MBARI. Contact Lonny Lundsten if interested in joining. (Week 6)
- Fixed bug in Tensorflow batch inference (Week 6)
- Successful deployment of MLFlow in dockerized container in AWS in Cloud9 (Week 6)
- Testing jupyter notebook with example benthic video processing pipelines using TensorFlow + MLFlow. A similar notebook will be developed for midwater. Ostensibly, the notebooks we are developing can also be adapted for Station M and drone still image processing. SageMaker has some great examples for those of you interested in this space. Please come talk to me or Duane Edgington for more details. (Week 9)
- Spun up MLFlow in AWS for development testing (Week 9)
- Added nose unit tests to tfrecord utility (Week 9)
- Adding semantic versioning and bitbucket pipelines for automated releases (Week 9)
- Working on system requirements (Week 9)
- Sent out ideas to the team on user interface design and discussions with Carlos about design ideas (Week 9)
- Working on Sagemaker notebook to demo fast video processing workflow (Week 10)
- Added unit test to tfrecord (Week 10)
- Reviewing summer intern applications (Week 10)
- About half way through sprint to load localized data into VARS database VAA-32. (Week 13)
- Example notebooks for benthic video detection, classification and tracking coming along. (Week 13)
- Working on system requirements. VAA-23. (Week 13)
- Finishing up ML tracker code and continuing work on VARS data loaders. Hoping to finish this up by next week (Week 14)
- Working on high-level system requirements doc (Week 14)
- Working on the beginnings of AWS web portal. Climbing learning curve of Vue, GraphQL and cloud deployment. (Week 17)
- Dockerizing tracking code. (Week 17)
- Finished data transforms for 2 of 4 datasets to handoff to FathomNet team https://bitbucket.org/mbari/vaa-localize-clean/; working on using the dockerized tools developed for assisting annotation to clean the remaining datasets so they can easily be loaded into VARS and also be used for tracking experiments. (Week 19)
- Team meeting and continued prototyping on Vue/Graphql app. (Week 19)
- Working with IS and KGomes on AWS web login. Joe Gomez made big strides in that last Friday spinning up a webapp in AWS redirected to our usual login. Thanks, Joe! (Week 19)
- Working on python notebooks for data localized data clean-up using the tools designed for assisting annotation. Two data sets ready for ingest so far. This set may serve well for summer intern work (Week 20)
- Continued evaluation of amplify/vue/graphql web app (Week 20)
- Student project discussion (Week 20)
- Team meeting scrubbed this week (Week 20)
- Meeting with UCSC Creative coding lab; this group is interested in helping with human-machine learning visualization aspects of the project. (Week 21)
- Meeting with summer intern, Peyton Lee, Univ Washington student, to discuss intern project. (Week 21)
- Finished week-long GraphQL/Prisma online training - all 110+ videos and exercises. Whew. (Week 23)

- Discussion with UCSC Creative coding lab about possible collaboration on user interface and machine learning efforts (Week 24)
- Student discussions (Week 24)
- Summer intern discussions on track metrics using benthic/midwater test track data. Python metric tracking package is working on our test data and should address key ML performance questions. Work is progressing quite well with Peyton! Our focus will shift to creating a 2-hour test sequence shortly. (Week 26)
- Setup CVAT and MLFlow development servers (Week 26)
- Working on schema design and experimenting with Cortex machine learning API as a possible tool to simplify devops and compute scaling needs. (Week 26)
- Discussion with UCSC Creative Coding lab scheduled for Tuesday June 30th. (Week 26)
- Code housekeeping to get all of the dockerized components in the workflow working with semantic versioning and automated docker builds and pushes to docker hub. (Week 28)
- Project proposal input (Week 28)
- Conference with USCS creative coding lab about AII/ML human-performance literature review and suggestions for improved VARS workflows using AI/ML (Week 28)
- Created AWS s3 bucket for training data collections versioned with DVC and working on code for reproducible preprocessing. (Week 28)
- Fixed an issue with metrics failing to log correctly following object detection training. (Week 31)
- Meetings with summer intern, Peyton who is working on evaluating performance metrics for our multiple object video tracker for the VAA project. We are using Amazon Web Services AWS and CVAT for this effort. (Week 31)
- Summer intern meetings and working on fixing an issue with object detection model deployment failing in AWS but working okay on prem. (Week 32)
- After evaluating ML stacks project plans are to process collection of benthic dives for proof-of-concept. Training libraries will include Cal Poly and FathomNet labeled images for top species that we have enough localizations for. (Week 35)
- Instrumenting TensorFlow 2 Retinanet module for Metaflow and midwater multiple object track tests. (Week 35)
- Experimentation with Cortex Machine Learning Stack (Week 35)
- Setup detection test for NetFlex Metaflow with TensorFlow2 Retinanet. Metaflow works on laptop and local GPU. Next steps are to test in the Cloud. (Week 36)
- Creating data registries for all available data for upcoming process tests. This versions the data for reproducibility and simplifies cloud testing. (Week 36)
    - https://bitbucket.org/mbari/registry-benthic2017 (Week 36)
    - https://bitbucket.org/mbari/registry-fathomnet (Week 36)
    - https://bitbucket.org/mbari/registry-i2map (Week 36)
    - CalPoly is next. (Week 36)
- Instrumented wandb logging in new object detection code in preparation for training in SageMaker. This will be a more efficient model (smaller/faster). (Week 37)
- Added data registry for FathomNet and Cal Poly data collections (Week 37)
- Created merged repo for taxa common to MARS BOEM sites with data filtering, merging, and correct VOC formatted outputs (Week 37)
- Sleuthed out why some data was missing in our Cal Poly data set. (Week 37)
- Metaflow/Cortex evaluation - looking like Cortex is a better fit (supports auto scaling, batch modes, easier to debug). (Week 37)
- Beginnings of OpenVINO docker image for model optimization (Week 37)
- CUDA deinterlacing docker image (Week 39)
- SageMaker training almost ready for sweeps (Week 39)
- Integrating C++ inference engine into tracking code. (Week 39)
- Instrumentation of efficientdet for spot training almost complete. Testing in SageMaker Studio. Many issues in the entire stack from the Docker/CUDA libraries, SageMaker and Wandb logging to resolve. (Week 44)
- Drafting documentation on JSON output for the team to work through VARS ingest issues. (Week 46)

- Finishing up local mode SageMaker tests which has sped-up the debugging process considerably. (Week 46)
- Waiting for approval to use a bigger instance for model training from AWS. (Week 46)
- Working on the integrated video inference and tracking code. So far so good. (Week 46)
- Team meeting to discuss 2020 milestones and 2021 plans. No other meetings planned for 2020 and the team expects to begin VARS integration and testing in early 2021. (Week 51)
- Finishing up database design for GraphQL API. This will serve as a data store for the machine learning visual event tracks and other ML related data needed to manage the ML lifecycle. Onboarding planned with Duane this week. Expect to deploy sometime next week. Wish me luck with the podman blues. (Week 51)
- Training a larger D4 EfficientDet model to use to populate a test database. (Week 51)
- Successfully tested several FathomNet and Sagemaker API calls in a web app taking shape that I'm calling DeepSeaAI. This will be both a dashboard and tool for submitting and monitoring the ML pipelines. (Week 51)

Out of Office

- Vacation day Monday 3/2.
- May 27-29 for much needed break. With the exception of a day, Danelle has been working during the entire quarantine period.
- Vacation day
- Off Fridays for summer duration
- Vacation July 15-24th
- Out of office Aug 24-28.

Overhead

- Performance review (Week 5)
- ACloudGuru training on CloudWatch logging with group members (Week 6)
- Attended ER meeting and tuned into presentations on bluejeans (Week 9)
- Danelle was asked to participate in an Expert Committee on Oceanography for the Canada Foundation for Innovation's 2020 Innovation Fund competition. This meeting, originally scheduled for Toronto in June, has been moved to a virtual meeting in light of COVID-19. Three grants with various ocean science themes with data management, compute, machine learning and outreach components will be reviewed. (Week 13)
- Out of Office (Week 13)
- Vacation day Monday 3/2. (Week 13)
- Engineering meeting (Week 14)
- Correspondence with Cal Poly about NSF Convergence Accelerators - AI-Driven Innovation opportunity. (Week 17)
- Discussions with Cal Poly collaborators about AI Data Sharing NSF Convergence Accelerator proposal. (Week 19)
- AWS Online US/Canada Summit tutorials and keynotes (Week 20)
- Vue/Prisma bootcamp. About 80% complete with this online training class. (Week 21)
- Updated miniomlflow stack. Pretty useful for those interested in experimenting with buckets and using MLFlow for managing machine learning models. (Week 23)
- Volunteer for SACNAS virtual conference most of last week. Targeted students interested in ML/AI based on poster presentations for summer internships. (Week 44)