Phase 2 response:

Thank you for the feedback on our proposal for 2019 and on our progress in 2018 and preceding years. We appreciated the opportunity to discuss this feedback in person, to confirm and help us craft an appropriate way forward. We are working with project 901900 and 5000055 to coordinate our response and our plans, as we always were.

Based upon this input, we propose to:

1) Convene an institute-wide working group, with active participation, leadership, and, where needed, authority from management team members, with an objective to create a comprehensive, agreed-to vision and plan for institutional imagery data at MBARI.

   The scope will include (Fig. 1):
   a. Articulation of objectives and requirements for video acquisition, management, curation, analysis, and life cycle. Science, Engineering, Operations, Communications.
   b. Acquisition plan (cameras, video formats, data volumes)
   c. Management (storage mechanisms, costs, formats, life cycle).
   d. Analysis (institutional human annotation, detail of annotation, opportunities for machine-aided annotation and analysis). The plan will include institutional requirements for annotation, including outline vs detail, and associated costs and ML opportunities.
   e. Automation (opportunities and explorations for machine-learning (ML) based workflows to aid human analysis).
   f. Institutional video policy.

   This plan will cover short-term activities (how we move forward now), as well as decadal-scale projections of future activities, options, limitations, opportunities, and requirements related to how video data will be collected, analyzed and retained moving forward at MBARI (see high-level view, Fig. 2). It will include the plan for introduction of next-generation video capability (4K), integration of machine learning tools into our workflows, and expected outcomes and expense and labor costs in all areas.

   The VideoTAG team provides focus and expertise to the creation of this plan; the Management Team provides leadership, authority, and ultimate responsibility for creation, ratification, adoption and execution of the plan.

   The timeline for this activity:

   Initiate and ramp as soon as possible, with a goal of articulation and presentation in time for 2019 DOEST and 2020 proposal activities.

2) Focus remaining 2018 and 2019 machine-learning related (ML) activities on assessing technology readiness to directly impact video lab workload and outputs. Specifically, we propose to allocate resources to build upon our findings* to date (D. Cline, 200 hours) to:
   a. Gather requirements for ML workflow that off-loads significant effort from video lab staff applied to outline-level annotation of video.

b. Evaluate application of unsupervised learning techniques to develop ML system to train outline-level (course-level) annotation of video from ROV descent, ascent, and non-transect benthic activities.
c. Evaluate reinforcement learning applied to ML system to continuously improve performance (accuracy, precision) while system is used to perform annotations.
d. Collect example video clips generated from ML systems to form core of MBARI VideoNet Monterey Bay underwater video guide.
e. Collect models used from ML systems to form MBARI model zoo.

This activity is not new for the VideoTAG team; it is simply an articulation of a strategy we have been pursuing over the last year. We recently acquired the NOAA funded SRI International (Sarnoff Lab) Flask (Automated "Fish" (Fauna) Classification with User-Guided Annotation Tool), an output from the NOAA AIASI (Automated Image Analysis Strategic Initiative) that MBARI (Edgington) has been actively participating in for the last 5 years. We are poised to evaluate this tool as a case-study to see if Deep Learning unsupervised methods, with human in the loop, and reinforcement learning can be applied to our video lab requirements. It is our expert opinion that outline-level descent, ascent, and non-transect benthic are the best candidates for evaluations given the available technology – detailed benthic transect (with accurate counts) are more challenging for ML, and midwater video analysis the most challenging, though of course we continue to look for breakthroughs and collaborations to address those data as well. Collected video clips classified by ML evaluations naturally form the basis for MBARI VideoNet curated collection of sample video clips, to drive ML efforts internally and externally, as well as education and outreach. Evaluated models naturally form the basis for MBARI model zoo, for internal and external use.

All of this work, of course, falls under the shadow of the comprehensive video plan of (1) above and the requirements codified in that plan.

*Findings - this is what we have learned about ML from our years of study and application:

a. Methods that report the highest accuracy are based on deep neural networks.
b. Transfer learning is the key method used to train in the absence of large samples.
c. Metrics are key to understanding and evaluating models, e.g. Accuracy, F1 score, mean Average Precision.
d. Model development is highly iterative and requires good data science skill.
e. Gathering training data is a barrier to adoption.
f. Benthic video is easier to automate than mid water video. This could be because the species are generally larger, opaque and sessile. That is not to say that automating mid water is untenable – it's just a harder computer vision problem.
g. Distribution of data matters. Choosing training data from the same distribution will produce better results, and there may be cases where oversampling of rare species produces better results.
h. There is a performance trade-off in accuracy versus computational time in different deep networks for object detection. More accurate models take longer to run the prediction step (inference).
i. Automatic classification can assist in the annotation of benthic megafauna for density, biomass evenness, and diversity estimation.

3) A project update is in our plan for 2018, though we are concerned that may not be an effective vehicle to demonstrate our progress and results. Aspects of such an update were presented at the ML workshop prior to Phase 1 proposals. We will continue our endeavors to communicate ML capabilities, limitations, and best practices to interested MBARI users (e.g. continue working w/ Smith postdoc Jennifer M. Durden to further ITS classification progress; w/ the Video, Midwater, Benthic, Bioinspiration labs to identify use cases, create training data sets, share progress, and establish efficient workflows). Additional suggestions or forums are welcome.

4) Our CalPoly activities and Paull's wind farm project 901023 are not directly related – Cal Poly is focused on dataset and algorithm development and ends April 2019 (Cal Department of Energy grant); Paull's project is focused on data collection and analysis in 2019-2020 (a multi-agency effort). While those collected videos may be useful for training classification models to meet the long-term Cal Poly goal of developing ML algorithms to process data from future wind farm site surveys, we do not expect the ML tools will be available for many more years.

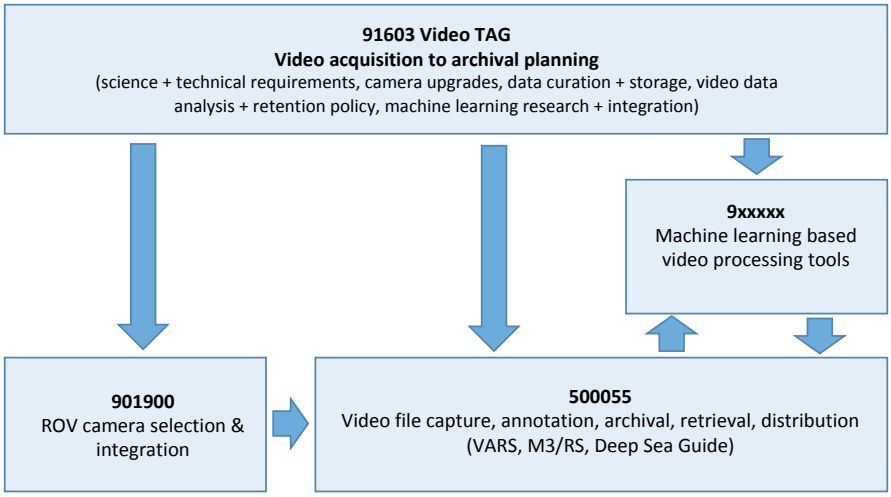

Figure 1. A new strategic vision for video data at MBARI: the VideoTAG team will lead the effort, working with key projects, stakeholders, and experts throughout the institute, to create a comprehensive plan for how MBARI will capture, manage, and use video data over the next decade. -- This model is a draft --

	VIDEO TAG 901603		CORE ANNOTATION 500055	MACHINE LEARNING for IMAGERY potentially a new, separate project in the future	ROV Camera Upgrades 901900
Core functions	**Lead strategic vision/plan for a sustainable institutional video program at MBARI:** **Video acquisition** **Video curation** **Video life cycle/data policy** **Video analysis** **Image Machine learning**	>>>	**Video recording (M3RS)** **Video curation (management + storage) (M3)** **VARS software** **Annotation policy, methods, implementation** **Ancillary functions (DSG, Embargos, QC/QA)**	**Machine learning tool development for internal imagery analysisys work/ problems**	**ROV camera selection** **ROV integration > acquisition**
Core Team	Duane Edgington Danelle Cline Nancy Jacobsen Stout Lonny Lundsten Todd Ruston Mark Chaffey Management Team (Doug Au, Jim Barry, others as needed)		Nancy Jacobsen Stout Brian Schlining (VARS SE) Lonny Lundsten Todd Ruston (Storage)	Danelle Cline Duane Edgington Brian Schlining Nancy Jacobsen Stout Lonny Lundsten Research representation	Mark Chaffey (design) Lonny Lundsten (M3RS integration) Dave French (ROV integration) Steve Haddock (Sci. impact)
Stakeholders (additional or future team membrs, labs, projects)	All of MBARI		Kyra Schlining Susan von Thun Linda Kuhnz Megan Basset	Ben Yair Raanan John Ryan Steve Rock Lab Bioinspiration Lab Acoustics Lab	Bryan Schaeffer (Ricketts) DJ Osborne (Ventana) Dale Graves (MiniROV) Todd Walsh (maint, calib)
Progress to date	Adisory on video and machine learning projects Investigation of current machine learning technologies Collaborations with external peers on projects specific to deep-sea domain		30-year's of video assets, annotations, and publications VARS and DSG M3 and M3RS (modern file-based system	Deep-learning methods successfully applied to still images Significant experience developing effective training datasets (stills and video) Shifted focus to machine learning technologies for video	COTS camera testing (75%) ROV integration readiness Doc Ricketts (%) Ventana (%) MiniROV (%)
Future high-level goals	Develop and maintain overall strategic vision		VARS software modifications/maintenace Core annotation/backlog (humans + AI)	Integrated machine learing into everyday annotation workflow MBARI Model Zoo (published models and procedures) MBARI VideoNet (currated, hierarchical)	Specify camera upgrades that improve resolution for science Ensure camera technology does not become obsolete
Challenges and identified solutions	Lack of institute strategy and priority for video and machine learning projects Increasing data streams and resource time to manage them		Diversify the pool of software engineers for developing tools to aid processing and analysis	Secure funds for: Engaging active external collaboration Cloud-based, serverless ML evaluations/solutions internal computational capabilities	Incompatabilities among platforms Wide array of COTS cameras to evaluate, many low resolution

Figure 2. The scope of issues and work to be addressed in developing a new strategic vision and comprehensive plan for how video data management. -- This scope of work is a draft --

Phase 1 feedback from the MT:

The Video TAG team is expected to provide leadership in several key areas related to video at MBARI. One area is video data curation, analysis, and management. Another is integrating appropriate machine learning methods into MBARI's video data workflow.

There appears to be great progress made in using machine learning as a tool for analyzing video. This would be a good time to reassess what MBARI's specific needs are and what our options are with respect to machine learning methods.

The team should re-scope their activities with these priorities in mind.

The team should re-engage on the science side with respect to these areas of emphasis:

1) Consider the problem of video data storage. Convene a group to define requirements that includes annotators, science stakeholders, and other experts. The group should define how are we going to manage video for now and into the future. Define what problems are being addressed. A much more comprehensive requirements document is necessary. A project update might be an effective way to assemble this group. Work with the 901900 and 500055 project teams on video storage aspects for 4k and the expected growth in video data storage demands as part of this requirements definition. (E.g. What amount of video should be stored, saved, and collected?)

2) Machine learning activities should be focused on what internal problems need to be solved and how machine learning tools can be integrated into our video processing workflow. While the concept of a Model Zoo appears to be a great idea, it is unclear how this will fit into our workflow and what MBARI-specific problems this might address. Convene a group to determine what specific videorelated problems might be addressed with machine learning methods. This group should also determine the institutional requirements for annotation (e.g., coarse vs. basic video annotation). There needs to be a much more refined requirements document on how machine learning tools will be used.

For Phase 2 please address the following:
• Re-scope activities. Please be sure to connect with the management team to discuss this further if any clarification is needed.
• As presently scoped, it is unclear how the engineering hours are spread across activities. Prioritize and de-scope Danelle's hours. She is available for 200 hours toward this effort in 2019 due to oversubscription.
• Hold a project update on the machine learning efforts. Demonstrate our current capabilities. Show how these capabilities can be used to address MBARI-specific problems.
• Note that there could there be a collaborative opportunity between efforts on the Cal Poly grant and Charlie Paull's wind farm project.