Performance Work Statement (PWS)

National Oceanic and Atmospheric Administration (NOAA)

NOAA Fisheries

Alaska Fisheries Science Center (AFSC)

Marine Mammal Laboratory (MML)

Alaska Ecosystems Program (AEP)

Steller Sea Lion Artificial Intelligence (AI) Image Processing and Counting Effort

1.0 Background

The Alaska Ecosystems Program (AEP) is charged with monitoring the protected Steller sea lion populations in Alaska. AEP conducts annual crewed and uncrewed (UAS) aircraft surveys of known sea lion terrestrial sites during the peak of the breeding season to collect counts. Aerial survey teams collect vertical (or near vertical; i.e., nadir) high-resolution overlapping imagery (to ensure complete coverage) of sea lions hauled out on land and swimming in the water nearshore. The level of overlap is not optimal for creating high accuracy mosaics, and AEP is seeking to work with experts in artificial intelligence (AI) to improve current image registration and detection/classification models used to generate Steller sea lion counts from the high-resolution imagery.

AEP has conducted UAS surveys since 2014, employing various camera makes/models (i.e., images with different resolutions) mounted to a downward facing gimbal to collect vertical images. Typically, there are up to approximately 15 sites per year that are surveyed using the UAS. Flights are conducted manually (i.e., not standardized track lines and therefore varying overlap). These images are higher resolution than the crewed aircraft imagery and typically do not have any metadata associated (e.g., date, time, latitude, longitude, etc.).

Since 2007, the crewed aircraft survey has implemented either a single-camera (only for the first few years) or three-camera array mounted in the belly port of the aircraft (Figure 1). The center camera is vertically oriented while the starboard and port cameras are tilted towards center at an approximate 21° angle (fixed) to provide approximately 10% horizontal overlap between the three-image sets. As the plane flies, there is approximately 60% vertical overlap between image sets. AEP has employed three (3) different camera make/models for these surveys since 2007 (i.e., varying resolution and scale of imagery). There are sites that require a single pass, however, many sites require multiple passes to fully cover a site (with overlap between passes to ensure complete counts). It is also common to have some passes repeated (i.e., high overlap of the same area). AEP attempts to conduct plane passes in a 'racetrack' fashion, however, topography and weather can hinder the ability to do so, resulting in less organized plane passes. Over 100 sites can be surveyed with this method.

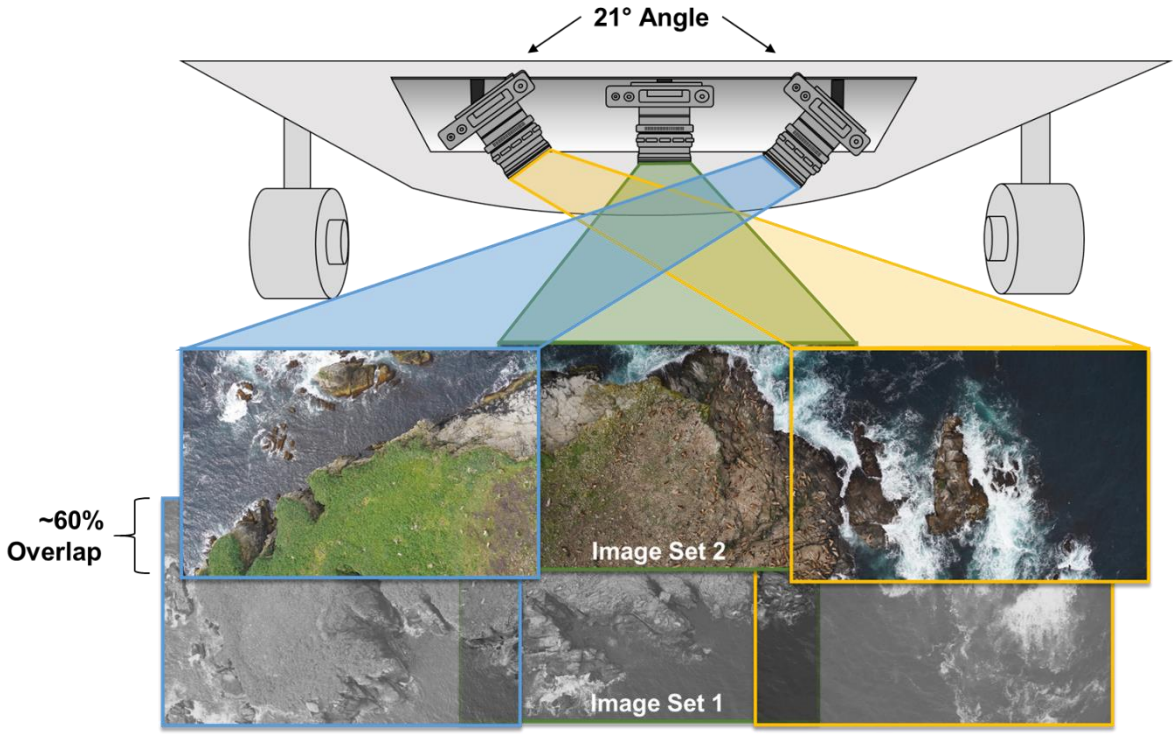

Figure 1—Schematic of the three-camera array mounted in the belly port of the crewed aircraft that collects 3-image sets with ~60% vertical overlap and fixed ~10% horizontal overlap as the plane flies.

Currently, our traditional count workflow involves two (2), full-time biologists to process and count all imagery independently, and when complete, compare and reconcile counts. The traditional method involves first manually mosaicking imagery by puzzling images together and drawing lines to eliminate areas of overlap to prevent double counts of sea lion individuals. The counters then review each image, placing a color-coded dot on each sea lion. Each color represents an age and sex classification: adult male, sub-adult male, adult female, juvenile, pup, dead non-pup, and dead pup. Once counting is completed by the independent counters, counts are summed by site for live non-pups (adult male, sub-adult male, adult female, and juvenile) and pups, and compared between counters. Any counts that are >5% different (or a difference of >20 for non-pups and >10 for pups) are revisited by counters, requiring them to evaluate images in order to correct their count. This traditional count method can take up to four (4) months for the two (2) full-time biologists.

To streamline this process, AEP has been working on developing an artificial intelligence approach for image processing and counting. Specifically, to work with AI experts to develop an image registration method to account for overlapping portions of the images to eliminate double counts, as well as counting and classifying individual sea lions. AEP contracted with Kitware, Inc. from 2019 to 2022, who developed several detection and classification models and two (2) image registration models, along with several pipelines, utilities, scripts, and improvements to Kitware's DIVE graphical user interface (GUI; the VIAME Toolkit and specific "Sea Lion Models"openly available on Github).

Websites for VIAME Toolkit and Sea Lion Models via Github

Kitware's VIAME Toolkit:
https://www.viametoolkit.org/

Kitware's Sea Lion Models via Github:
https://github.com/VIAME/VIAME/wiki/Model-Zoo-and-Add-Ons#sea-lion-models

2.0 Scope/Tasks

AEP assessed the highest accuracy detection and classification models and both image registration models, identifying the most optimal for further development: the FUSION (all class) detection and classification model and the Suppression image registration model. During model assessment, AEP identified areas of improvement to guide future development of existing models, scripts, utilities, pipelines, and the DIVE desktop GUI. The Contractor shall further develop and improve the detection and classification model (FUSION) and image registration model (Suppression); further develop utilities, pipelines, and scripts required to improve and evaluate models' performance; and finally, improve the DIVE (desktop, not the web) interface to support model implementation and improve user experience. Details are specified in Tasks below. Kitware developed several pipelines, utilities, and scripts to perform various tasks and support the FUSION and Suppression models; for example, suppress water detections, ignore regions, register frames, and generation of tile mosaics. Further development of these--or development of new--pipelines, utilities, and scripts may be required to support the Tasks below. The Contractor shall not develop any new detection and classification or image registration models. Additionally, there are existing scripts for model scoring and performance, and receiver operating characteristic (ROC) curve generation utilized for evaluating models to understand accuracy and performance. AEP and the Contractor shall work together to discuss and finalize objectives, project plan, and prioritization of and within each task.

To summarize, the Contractor is required to perform the following objectives:

1. Further develop FUSION (all class) detection and classification model
2. Further develop Suppression image registration model
3. Supporting scripts, pipelines, utilities, or similar to support models 1-2, above.
4. Tile mosaic capabilities (for three-image sets) that can be used in the DIVE (desktop) GUI with objectives 1-3, listed above.
5. Single frame (for single and three-image sets) capabilities that can be used in the DIVE (desktop) GUI with objectives 1-3, listed above.
6. DIVE desktop GUI capabilities to view, edit, process, and evaluate (i.e., attribute fields) detections and annotations generated from models and products from objectives 1-3, listed above, for images configurations from objectives 4 and 5, listed above.
7. Virtual training covering how to use products, how to train models, how to use DIVE with these products and for editing and evaluation of model outputs.

Details pertaining to Task needs are described in paragraph form and bulleted lists, with specific Task Deliverables listed as letter bullets listed at the end of each Task. Timing and metrics of success can be found in the Schedule of Deliverables and Performance Requirements Summary sections below. Key tasks include the following:

2.1 Task 1: Project Communication

A. Quarterly Virtual Meetings: The Contractor shall coordinate with AEP Technical Point of Contact (TPOC; will be assigned upon award), to schedule virtual quarterly meetings with AEP and the Contractor (i.e., technical project manager; at minimum, other teammates shall join when applicable) to discuss efforts to date (with demos/visuals, if applicable), goals for the following quarter, project priorities, timelines, data needs from AEP (e.g., training data, flight logs, survey data), and any other pertinent topics. This task includes a "kick-off" meeting at the start of the base and option year (if applicable) when priorities and timelines will be discussed for the year. At the end of the base and option (if applicable) year, there will be a "wrap up" meeting that counts towards the quarterly meeting requirement.

B. Monthly Reports: The Contractor shall provide brief (approximately 1-2 pages) monthly reports from the technical project manager on progress, issues encountered, recommendations, efforts accomplished the month prior, and any plans for the following month. The technical project manager shall address any follow-up questions from AEP regarding the monthly summary reports. Monthly summary reports shall be submitted with the monthly invoices.

C. Iterative Products: The Contractor shall share iterative versions of model, pipelines, utility, scripts, and/or other products (e.g., DIVE GUI updates) from Tasks 2-4 to AEP as available to allow AEP to conduct assessments with sample data. This will assist AEP in providing meaningful feedback on products to support and inform further development throughout the effort, and allow AEP to stay informed on progress. The Contractor shall only share iterative products that AEP will be able to implement on a test set of data, accompanied by instructions on how to implement and use the iterative products.

2.2 Task 2: Tile Mosaic Capabilities

Kitware developed a script to create "tile mosaics" from the image sets captured with the three-camera array (crewed aircraft survey; Figure 2A). AEP assessments discovered tile mosaics lacked functionality in several ways:

- Scripts frequently failed, and regularly created inaccurate tile mosaics (i.e., distorted mosaics), especially for images with homogenous substrate (e.g., water and vegetation; Figure 2B).
- The Suppression image registration model failed to create suppression zones on an average of 50% (range: 0-95.2%) of tile mosaics.
- Of sea lions that the detection and classification model missed (i.e., not annotated), 93% were due Suppression model errors. This is a higher percentage than instances with single-frame imagery, which resulted in 23% of missed animals.
- Tile mosaics were not compatible with the desktop DIVE desktop GUI (only the web GUI).

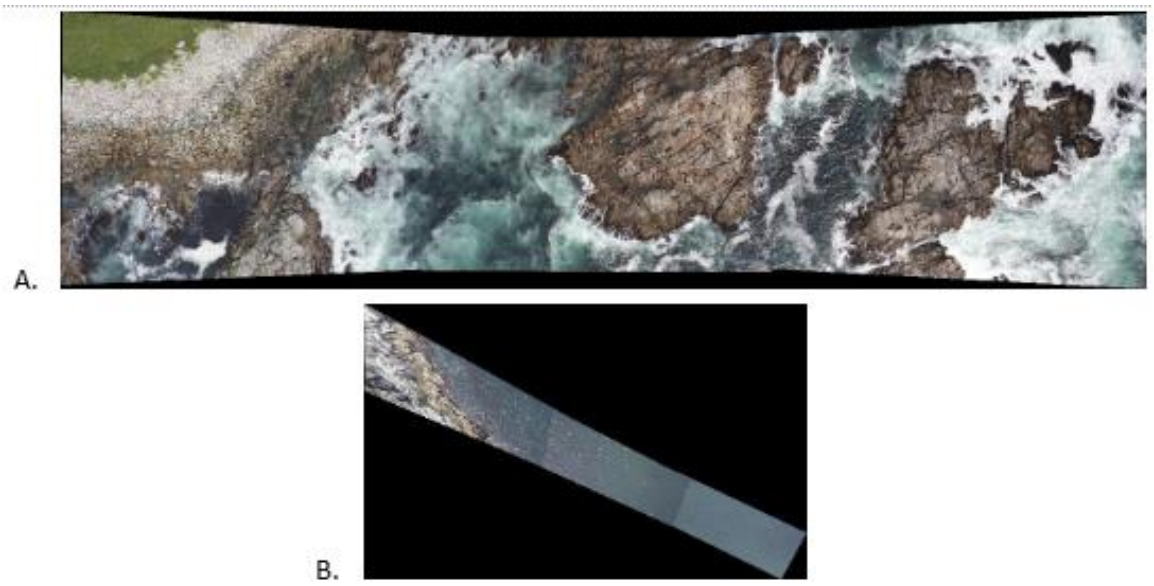

Figure 2—Examples of tile mosaics of the image sets captured by the three-camera array that are acceptable (A) and poor (B) quality.

A. Improved Tile Mosaic Capabilities: The Contractor shall improve the ability to create accurate tile mosaics from three-image sets consistently.
B. DIVE Desktop GUI: The Contractor shall develop DIVE desktop GUI to be compatible with tile mosaics (e.g., ability to view, edit, and evaluate model detections with tile mosaics in DIVE) matching existing capabilities for single frames.

2.3 Task 3: Suppression Image Registration Model

The Suppression image registration model performed well in some instances with single-frame imagery and for sites that only required a single plane pass. In instances with single-frame imagery, where there were multiple, overlapping plane passes (or manual UAS flights), the model did not perform at identifying sea lions that have already been "seen" by the models in previous passes. Within a single pass, approximately 8.6% of sea lion annotations were duplicates (i.e., same sea lion annotated on multiple images). As stated above, of sea lions that the detection and classification model missed (i.e., not annotated), 93% were due Suppression model errors. With tile mosaic imagery, these percentages were even higher at 25% duplicates and 93% missed. An average of 50% (range: 0-95.2%) of the tile mosaics failed to have suppression zones written at all.

To assist in the development of this feature, AEP will provide flight logs (.csv files) for imagery collected by the crewed aircraft with location, roll, pitch, yaw, heading, altitude, etc., information that can be linked to image frame numbers. This information is largely unavailable for UAS imagery. If relevant to this task, information on the camera makes/models can be found in image metadata (i.e., image resolution and scale information).

A. Improved Suppression Image Registration Model: The Contractor shall further develop the Suppression image registration model to improve performance for single-frame imagery. The Contractor shall develop the model (or additional script) to include registration for multiple, overlapping plane passes (e.g., loop closure) and manual UAS flights for single-frame (single and three-image sets) and tile mosaic imagery options. Ensure the Suppression model is functional with single-frame and tile mosaic imagery inputs in the DIVE desktop GUI, allowing users to run

models, and view, edit, and evaluate (e.g., attributes) annotations. In order to improve and support this model, further development of, or development of new, scripts and/or pipelines may be required.

2.4 Task 4: FUSION (All Class) Detection and Classification Model

The FUSION (all class) detection and classification model detects individuals and classifies them by age and sex classes, including dead animals (adult male, sub-adult male, adult female, juvenile, pup, dead non-pup, and dead pup). In assessing the FUSION (all class) detection and classification model, AEP identified several areas where the model requires improvement: increase accuracy of detecting individuals in large, overlapping aggregations; detecting animals in the water (hereafter referred to as "water detections"); and reducing false positive detections in water (e.g., waves, seaweed, light reflections, etc.). To ameliorate the latter, the model should be developed to separate land detections from water detections.

The model will also be improved by including scale information using the camera make and model sensor information, especially for UAS imagery. Five (5) different camera makes/models have been utilized to-date. Camera make/model upgrades are expected to occur in the future and require a way to account for changes in image resolution (i.e., scale). Finally, AEP will provide new and upgraded (previously used annotations now with higher resolution imagery) training data to retrain the model. According to our most recent assessment (confusion matrix) with 2024 imagery, we found that this model was 86% accurate in classifying animals. The model missed 13.39% pups and 3.63% non-pups (missed 6.19% total). We are seeking improved accuracy in classification and detection of missed animals, especially pups.

A. Improved FUSION (All Class) Detection and Classification Model: For the FUSION (all class) detection and classification model, the Contractor shall evaluate the model to assess further improvements based on their specialized expertise and from the information outlined above. Improve the detection of individuals in large and overlapping aggregations. Utilize false positive training data to increase accuracy (i.e., hard negative mining). Create the ability to separate sea lions detected on land versus in the water. Create capabilities to account for imagery scale from camera make and model sensor and resolution information (provided in image metadata) for camera makes and models used in the past, present, and future. Integrate the option to account for high and/or low performance GPU processing (i.e., to reduce "chipping"). The FUSION model is currently composed of three (3) different models: The Contractor shall trial the addition of new, additional models to include in the FUSION model and assess whether it will improve performance. Communicate with AEP Technical Point of Contact on other potential areas of development to improve model performance. Re-train the model with upgraded (to higher resolution imagery) and new training data (provided by AEP). Ensure the FUSION model is functional with single-frame and tile mosaics with the DIVE desktop GUI, allowing for users to run the model and view, edit, and evaluate (e.g., attributes) annotations. In order to improve and support this model, further development of, or development of new, scripts and/or pipelines may be required.
B. Model Performance Evaluation Capabilities: Provide scripts to evaluate model performance (e.g., scoring, accuracy, and ROC curve) run with AEP validation data, as needed. These scripts shall work effectively with single-frame and tile mosaic imagery inputs and when the Suppression image registration model is used.

2.5 Task 5: Virtual Training

A virtual training led by the Contractor is required to learn about final product capabilities, how to implement and utilize products, how to train detection and classification models, about improvements to the DIVE desktop GUI, and any other pertinent information related to AEP's mission. The training will be scheduled just before the end of the contract (first year or option year, if applicable). The training length should be at least four (4) hours long and no more than eight (8) hours to address and adequately cover all pertinent topics. The training will not need to include basic information about VIAME or DIVE and AEP can provide a small data package (e.g., sample of imagery, etc.) for the Contractor and participants to use as an example data set.

A. Virtual Training: The Contractor shall coordinate with AEP to conduct an open and virtual training of contract products (e.g., deliverables) and relevant processes (e.g., training the detection and classification model) required to utilize products. The Contractor shall coordinate with AEP on planning the training at least two (2) weeks in advance and shall lead this training within the last four (4) weeks of the end of the contract's period of performance.

3.0 Schedule of Deliverables

The following is a schedule for all Task Deliverables, including administrative Deliverables, required during the period of performance of this contract:

Task	Description	Quantity	Due Date	Deliver To
Task 1A	Quarterly virtual meetings	Base Yr: 4 Option Yr: 4 (if applicable)	Coordinate meeting date ≥2 weeks prior	AEP TPOC and Participants
Task 1B	Monthly reports	Base Yr: 12 Option Yr: 12 (if applicable)	Within first 10 days of the month	AEP TPOC & COR
Task 1C	Iterative products (from Tasks 2-4)	Base Yr: Varies Option Yr: Varies (if applicable)	As available; at least once per year	AEP TPOC via GitHub
Task 2A	Improved Tile Mosaic Capabilities	Base Yr: 1 Option Yr: 1 (if applicable)	Four (4) weeks prior to end of contract	AEP TPOC via GitHub
Task 2B	DIVE desktop GUI (with tile mosaic capabilities)	Base Yr: 1 Option Yr: 1 (if applicable)	Four (4) weeks prior to end of contract	AEP TPOC via GitHub
Task 3A	Improved Suppression Image Registration Model	Base Yr: 1 Option Yr: 1 (if applicable)	Four (4) weeks prior to end of contract	AEP TPOC via GitHub

Task	Description	Quantity	Due Date	Deliver To
Task 4A	Improved FUSION (All Class) Detection and Classification Model	Base Yr: 1 Option Yr: 1 (if applicable)	Four (4) weeks prior to end of contract	AEP TPOC via GitHub
Task 4B	Model Performance Evaluation Capabilities	Base Yr: 1 Option Yr: 1 (if applicable)	Four (4) weeks prior to end of contract	AEP TPOC via GitHub
Task 5A	Virtual Training	One (1) training at end of contract (base, or option yr, if applicable)	Within last four (4) weeks of end of contract	AEP TPOC and Participants

4.0 Performance Requirements Summary

This contract includes a Performance Requirements Summary (PRS) and Quality Assurance Surveillance Plan (QASP). The PRS plays an integral role in the administration of the contract. In addition to any applicable inspection clauses or other related terms and conditions contained in the contract, the PRS shall serve as a primary tool for inspection and acceptance of services as facilitated by the Contracting Officer's Representative (COR). Evaluation of the Contractor's overall performance shall be in accordance with the performance standards set forth in the PRS, and will be conducted by the COR. The PRS constitutes a material aspect of the contract and will not be changed or otherwise modified without prior written approval of the Contracting Officer.

Task/ Desired Outcome	Performance Standard (PS)	Acceptable Quality Level	Monitoring Method	Incentive / Disincentive
Task 1A	Contractor provides monthly reports with invoices.	PS met >90% of the time; on-time >80% of the time.	COR will solicit feedback from AEP TPOC; use of email to determine delivery; review of reports.	Material-poor performance evaluations; corrective action plan
Task 1B	Contractor provides meeting coordination and attends.	PS met at least 100% of the time; on time >80% of the time.	COR will solicit feedback from AEP TPOC; Use of email to determine scheduling; review of meeting notes.	Material-poor performance evaluations; corrective action plan
Task 1C	Contractor provides iterative pipeline/script/ DIVE products for AEP to review.	PS met at least once per year for each Deliverable in Tasks 2-4	COR will solicit feedback from AEP TPOC; use email to determine request/receipt date.	Material-poor performance evaluations; corrective action plan

Task/ Desired Outcome	Performance Standard (PS)	Acceptable Quality Level	Monitoring Method	Incentive / Disincentive
Task 2 A & B	Contractor creates functional tile mosaic capabilities and usable in DIVE desktop.	PS met when tile mosaics are drawn accurately ≥95% of the time; functional in DIVE at least ≥95% of the time; or Contractor showed good faith effort to meet metrics above	COR will solicit feedback from AEP TPOC.	Material-poor performance evaluations; rejection of invoice until rectified; negative performance may not be considered for future work
Task 3A	Contractor delivers an improved Suppression image registration model.	PS met when the model works with ≥90% accuracy for surveys with multiple passes or Contractor showed good faith effort	COR will solicit feedback from AEP TPOC	Material-poor performance evaluations; rejection of invoice until rectified; negative performance may not be considered for future work
Task 4A	Contractor delivers an improved FUSION (all class) detection and classification model.	PS met when the detection accuracy is ≥95% and false negatives are ≤5%; or Contractor showed good faith effort	COR will solicit feedback from AEP TPOC	Material-poor performance evaluations;rejection of invoice until rectified; negative performance may not be considered for future work
Task 4B	Contractor provides scripts to evaluate model performance	PS met when products can assess model performance for different image inputs and with FUSION and Suppression model	COR will solicit feedback from AEP TPOC	Material-poor performance evaluations; rejection of invoice until rectified; negative performance may not be considered for future work

Task/ Desired Outcome	Performance Standard (PS)	Acceptable Quality Level	Monitoring Method	Incentive / Disincentive
Task 5A	Contractor develops and leads virtual training.	PS met when Contractor leads virtual, open training on Tasks 2-5 and detailed supporting tasks (e.g., training FUSION model).	COR will solicit feedback from AEP TPOC; review of training notes and video.	Material-poor performance evaluations; rejection of invoice until rectified; negative performance may not be considered for future work

Place of Performance

Work will be conducted at other than the government facility (i.e. off-site).

Travel

Travel is not expected or authorized for this task order.