Scottish Government
Riaghaltas na h-Alba

A New Application for Automated Video Identification of Marine Species (AVIMS)

Scottish Government Marine Directorate

Scottish Marine and Freshwater Science Vol 14 No 7

M Mackiewicz, G French and M Fisher

Annex 1 – User Guide

A New Application for Automated Video Identification of Marine Species (AVIMS)

Annex 1 – User Guide

Scottish Marine and Freshwater Science Vol 14 No 7

M. Mackiewicz, G. French and M. Fisher

Published by the Marine Directorate of the Scottish Government

ISBN:
ISSN: 2043-7722
DOI: 10.7489/12473-1

The Marine Directorate of the Scottish Government is responsible for the integrated management of Scotland's seas. Scottish Marine and Freshwater Science is a series of reports that publishes the results of scientific research and monitoring carried out by the Marine Directorate. It also publishes the results of marine and freshwater scientific work that has been carried out for the Marine Directorate under external commission. These reports are not subject to formal external peer-review.

This document is an Annex to the Main Report which presents the results of scientific work commissioned by the Marine Directorate and carried out at the School of Computing Sciences, University of East Anglia.

The full report should be quoted as:
Mackiewicz, M., French, G. and Fisher, M. (2023). A new application for automated video identification of marine species (AVIMS). Scottish Marine and Freshwater Science Vol 14 No 7, 18pp. DOI: 10.7489/12473-1.

© Crown copyright 2023

You may re-use this information (excluding logos and images) free of charge in any format or medium, under the terms of the Open Government Licence. To view this licence, visit:
http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/
or email: psi@nationalarchives.gsi.gov.uk.

Where we have identified any third party copyright information you will need to obtain permission from the copyright holders concerned.

A New Application For Automated Video Identification Of Marine Species (AVIMS)

Annex 1 – User Guide

M. Mackiewicz, G. French and M. Fisher

School of Computing Sciences, University of East Anglia,
Norwich Research Park, Norwich. NR4 7TJ

Abstract

In the course of its environmental monitoring activities, the Scottish Government's Marine Directorate collects a large amount of underwater video to, for example, obtain information on the numbers of fish in rivers or on species living on the seabed. Manual analysis of this footage is laborious and costly, but Machine Learning algorithms can now be used to automate such image analysis. The Marine Directorate commissioned the University of East Anglia to develop a web-based application to allow staff to create, train and execute machine learning-based (semi-)automated analysis of video footage without a need to interact with the underlying computer code. The application was tested using three diverse sets of video footage and found to be usable by staff without computer science or coding experience. The tool was able to detect and count sea pens in footage from towed underwater vehicles, salmon smolts at sea in underwater footage from towed fishing gear and adult salmon and sea trout in footage from underwater or overhead cameras at fixed locations on rivers. Improving the accuracy of the models at detecting and counting organisms of interest will require the use of larger annotated datasets in further training of the algorithms, but the current application provides a basis for further developing these.

Contents

Abstract....1
Contents....2
Abbreviations....3
Introduction....4
Training Resources....4
Workflow....5
References....11
Contact....11

Abbreviations

AI	Artificial Intelligence
AVIMS	Automated Video Identification of Marine Species
CNN	Convolutional Neural Network
GPU	Graphics Processing Unit
GUI	Graphical User Interface
mAP	Mean Average Precision
OS	Operating System
UEA	University of East Anglia

Introduction

This user guide relates to a new application for Automated Video Identification of Marine Species (AVIMS), developed to meet the requirements of:

> The Scottish Ministers' invitation to tender for a new application for Automated Video Identification of Marine Species (AVIMS) case/216380.

The software system provides a web-based tool set to enable Artificial Intelligence (AI) models derived from video footage captured on-site to be trained, evaluated and deployed. The system was originally conceived to target the following survey domains:

- Sea Pens
- Rocky Reef
- River - infrared
- River - daylight
- Vaki light boxes
- Fan mussels and horse mussels
- Smolt Trawl
- AVIMS Training Sessions

New surveys can be added as needed.

The tool is located on a server at the University of East Anglia (UEA) and will be maintained by Dr M. Mackiewicz, UEA School of Computing Sciences until 31 December 2021. Access is restricted to authorised users To discuss possible access please contact craig.robinson@gov.scot.

In view of the range of other training resources available (see next section) this user guide aims to provide an overview of the workflow.

Training Resources

In addition to this guide the following training resources are available:

- AVIMS Tutorial PowerPoint slides.
- An additional 'survey' named 'AVIMS training sessions' contains a selection of video material and instructions to guide new users through

the steps involved in producing a Convolutional Neural Network (CNN) model.
- The system includes comprehensive Help screens which steer the user through the workflow.

Workflow

CNN models are trained by assembling a training set of exemplar objects/creatures of interest, identified by manually annotating video frames using the *Django labeler* tool. This is a time-consuming and tedious step, and before this stage it is worth considering the aim for the survey (survey type).

There are several types of survey that can be addressed by AVIMS; considered here in order of difficulty:

a) Single site, fixed camera: Here, the task involves identifying (counting) objects/creatures of interest captured in video from a single stationary camera. Objects/creatures of interest move through the scene and we wish to train a CNN model to detect them. Although this represents the simplest scenario there is a problem, we should be aware of. Despite the stationary background, the lighting/shadows that affect the scene may change due to time of day, season, prevailing weather, water level etc. Therefore, we need to train on a representative set of video frames.
b) Single camera, multiple sites: In this case the background changes (e.g. the camera may be towed). Again, we need to train on a set of frames representative of the operational conditions we will encounter.
c) Multiple cameras, single site: In this case we envisage several cameras monitoring a single site. We assume all cameras are the same type (ideally the same model) but are mounted to acquire different views. Video from infrared and daylight cameras should not be mixed as the cameras have very different operating characteristics.
d) Multiple cameras, multiple sites: This is the most challenging scenario which will require a very large number of annotated frames and examples.

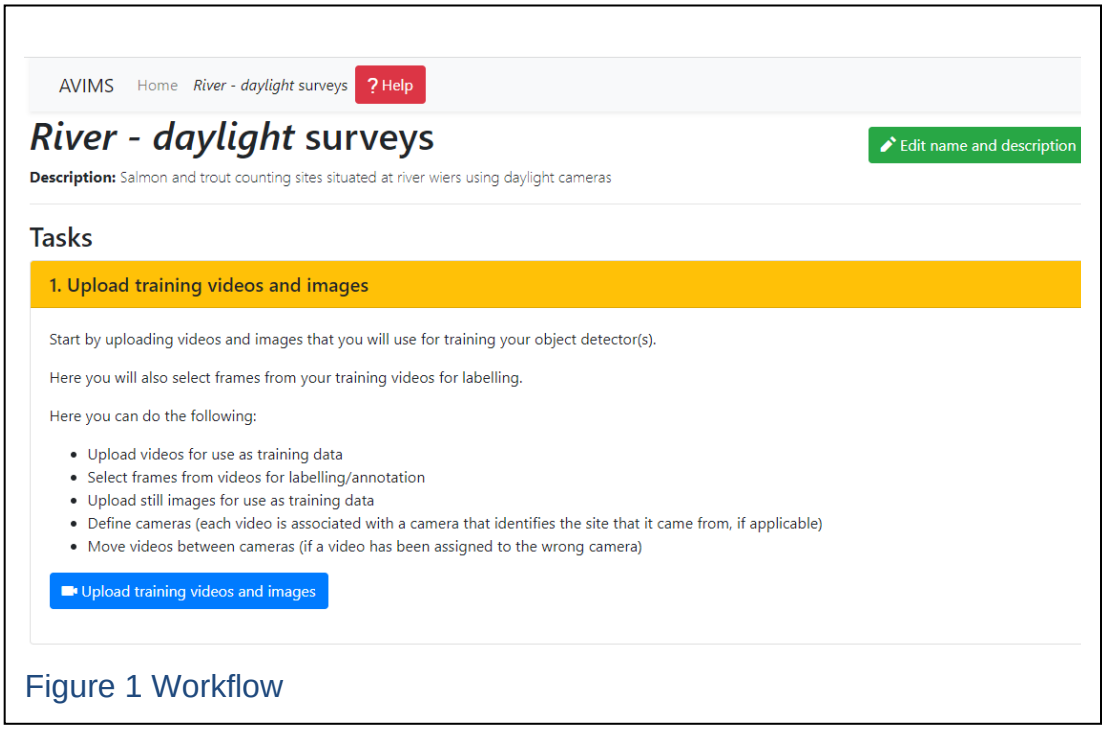

Figure 1 Workflow

To train a CNN model video footage must be uploaded and a labelling schema defined, or an existing schema used/extended.

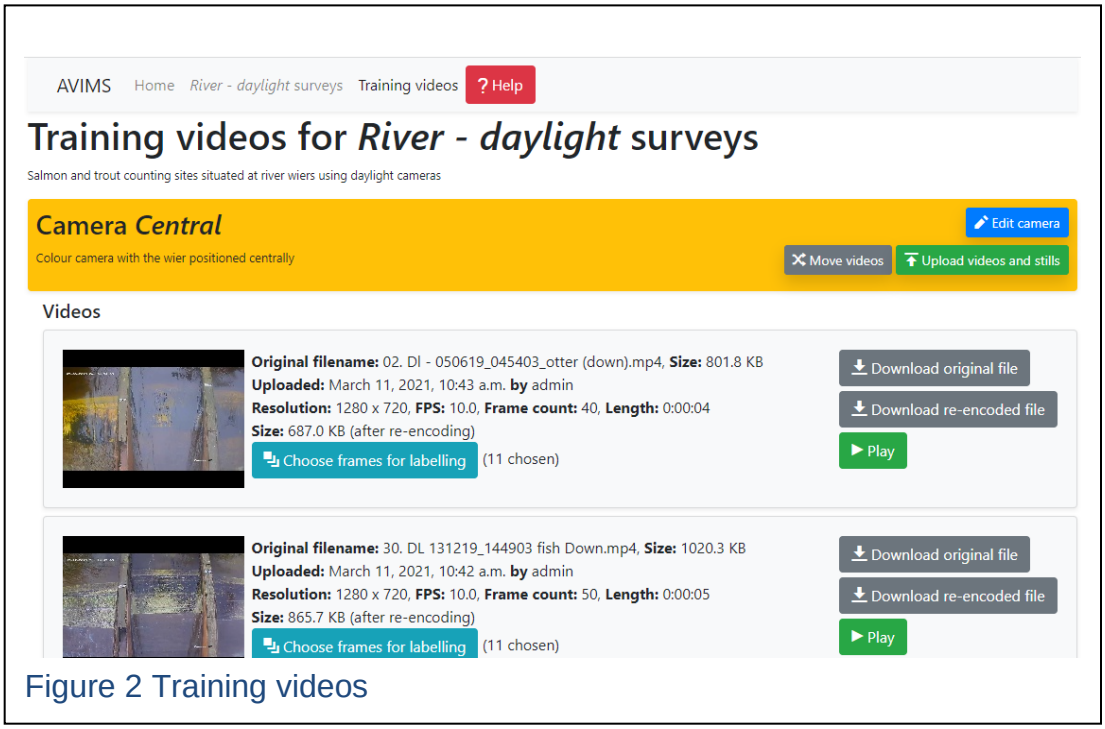

Figure 2 Training videos

Selecting an existing survey, or creating a new survey produces a *workflow* (Tasks numbered 1-6) with links to enable users to navigate the system and

allow them to build a model (Figure 1). The “**Upload training videos and images**” button provides access to existing video footage and allows new footage to be uploaded (Figure 2).

Using the “**Choose frames for labelling”** button we select video frames containing objects/creatures of interest (Figure 3) which will subsequently be labelled using the Django labelling tool. Note: If the scene background changes (i.e. a non-stationary background) it may be advantageous to add a small number of frames that contain no objects/creatures of interest.

Figure 3 Choose frames for labelling

The system is not optimised for video streaming so users may find it quicker to download the video and identify frames of interest by replaying the video using on their desktop video player.

A labelling schema (Workflow Task 2) identifies classes of objects/creatures of interest and denotes the overlay colour used by the Django labeller to denote them in the imagery (Figure 4).

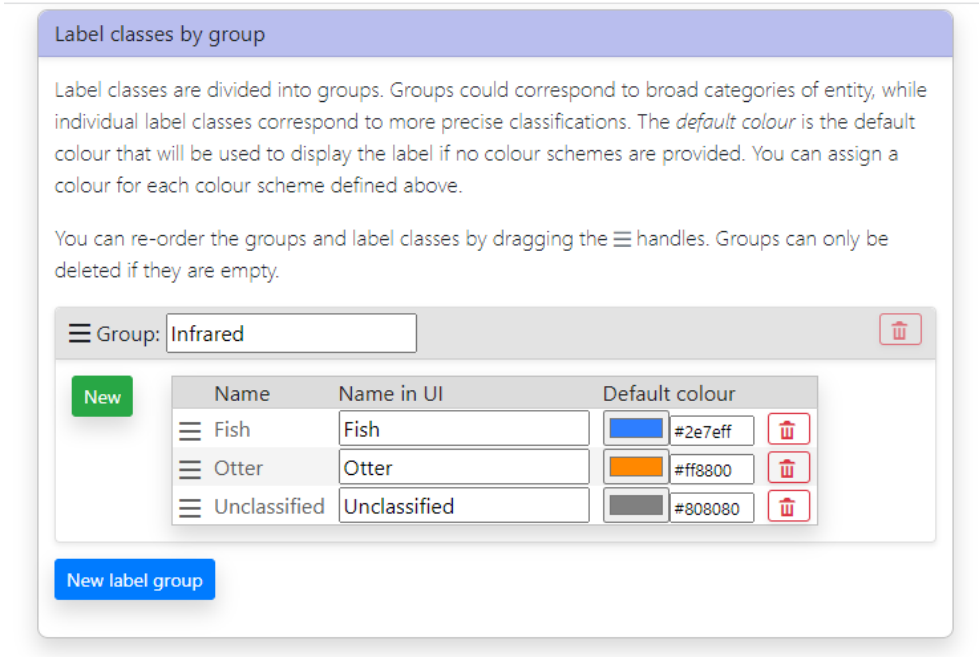

Figure 4 Labelling Schema

The Django labeller (Workflow Task 3) provides a variety of interactive tools to help us delineate objects/creatures of interest in the video frames we selected in Workflow Task 1 (Figure 5). It is important to label all the objects/creatures of interest in each frame because unlabelled pixels are considered as background. Hence, rigorous, consistent, careful labelling will improve the performance of the classifier.

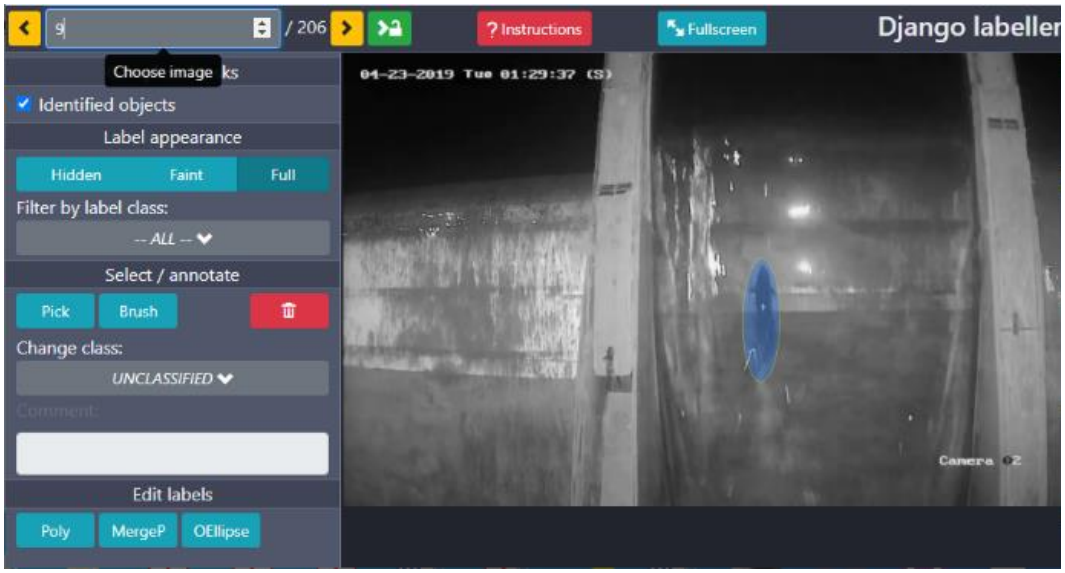

Figure 5 Django labeller

Successful training is usually achieved in stages. Start by building/testing a model by labelling a few (10's of frames) and then try to improve it by adding additional frames (creating a new dataset). Once the video frames have been labelled the remaining stages (Workflow Tasks 4-6) are relatively straight forward and automated.

Workflow Task 4 splits the training set into two parts comprising a set of images/labels used to train the CNN model and a smaller set that is 'held out' and used to test the model (Figure 6).

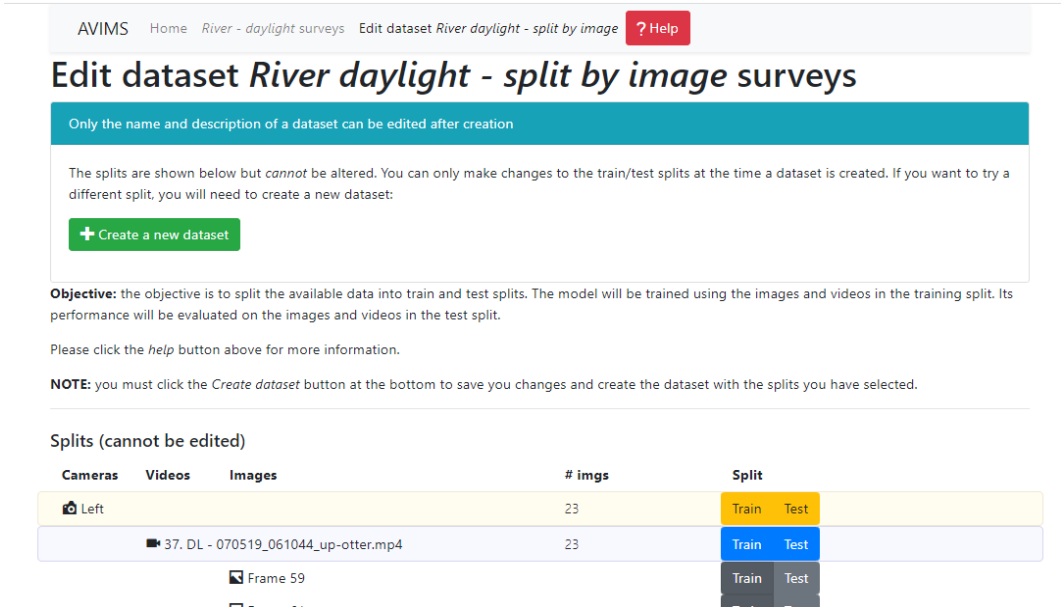

Figure 6 Train/Test Split

Frames are allocated to training or testing (randomly) when the dataset is created (but the choice can be overridden).

Once we have created the train/test split we can train an AI model (Workflow Task 5). This task is computationally heavy and requires a GPU worker. Jobs are queued and when completed a report is generated. The Training report (Figure 7) provides a log which is mainly used for debugging and details the performance (on the hold-out test set). Overall performance is summarised by a mean Average Precision (mAP) score.

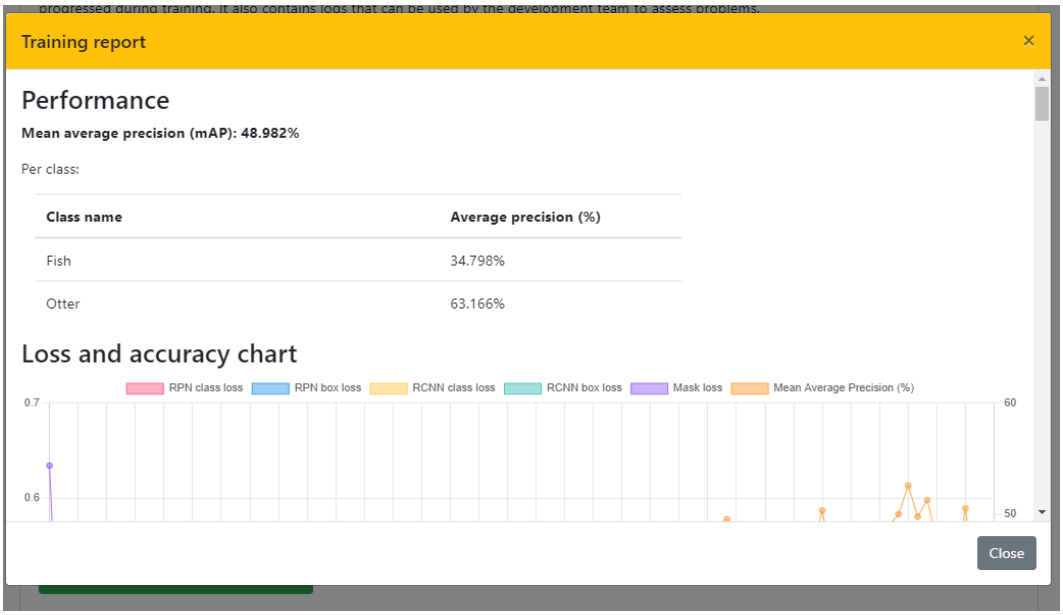

Figure 7 Training Report

We can also review results in the Django labeller (Workflow Task 5 "**view predictions**" button). This presents frames in the test set and overlays predicted class labels using the labelling schema defined previously (Figure 8).

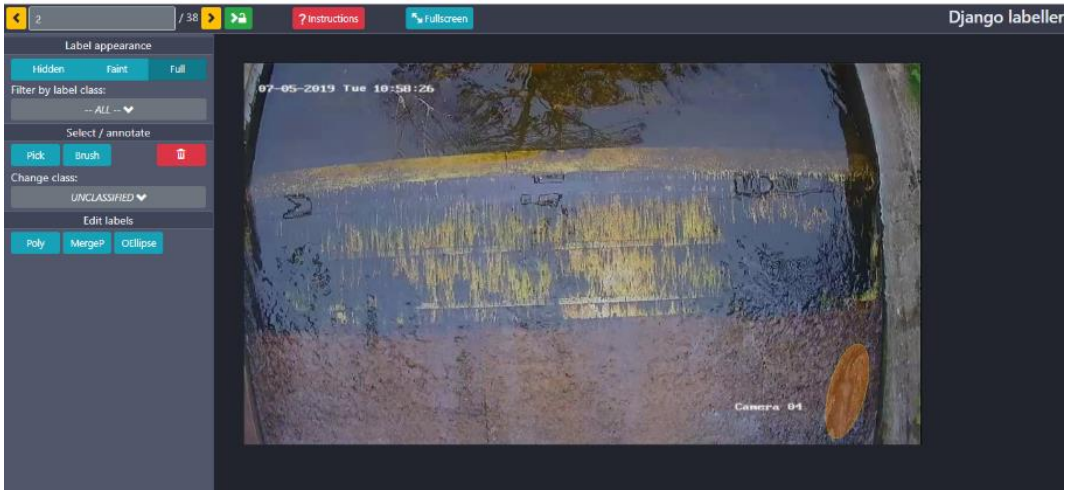

Figure 8 Predicted class labels

Finally, Workflow Task 6 allows video footage to be uploaded and analysed by the model. This footage might have been used for training (e.g. the user might be interested in predictions from all the video frames, not just those selected for testing). However, results more realistic of an operational scenario will be

obtained from previously 'unseen' footage (i.e. not used to train/test the model). This task is also computationally heavy and requires a GPU worker. Jobs are queued and when completed an analysis report with the counts of detected entities is available for viewing from the GUI or for download in either the .csv or the .json format. The analysed video containing the detected entities visually overlayed over the original footage can be also replayed or downloaded. Further, this workflow task allows the uploaded footage to be re-purposed for training (will require further labelling) with a view of improving the next iteration of the machine learning model.

References

Blowers, S., Evans, J. and McNally, K. 2020. Automated Identification of Fish and Other Aquatic Life in Underwater Video. In: Scottish Marine and Freshwater Science, Vol 11 No 18. 62pp. DOI: 10.7489/12333-1

Contact

Email: craig.robinson@gov.scot

© Crown Copyright 2023

Marine Directorate of the Scottish Government
Marine Laboratory
375 Victoria Road
Aberdeen
AB11 9DB

Copies of this report are available from the Scottish Government Publications website at https://www.gov.scot/publications/