AUTOMATED MAMMAL LOCALIZATION AND IDENTIFICATION IN CAMERA TRAP IMAGES FOR THE NORTHEASTERN UNITED STATES

Brannon Barr[1], Harold Underwood[1], Giorgos Mountrakis[1], and Lindi Quackenbush[1]

[1]SUNY-ESF

July 4, 2022

Abstract

1. Camera traps are popular for monitoring animal populations and communities, primarily because they eliminate physical handling of animals. However, image acquisition typically outpaces information extraction. Most deep-learning based animal classifiers do not localize animals, limiting their applicability. Existing networks that localize animals have relatively high training data and hardware requirements. 2. To reduce the hardware and training data requirements, we extended the the Machine Learning for Wildlife Image Classification network (MLWIC2) to a Faster R-CNN. MLWIC2 is currently the most accurate wildlife classification network, and also the shallowest at 18 layers. We compared our model's performance at object localization, species identification, and deployment speed to the performance of a generically pre-trained 50-layer Faster R-CNN to determine a) relative importance of task similarity in pre-training vs. backbone depth, b) whether additionally finetuning the backbones during training is advantageous c) whether the Faster R-CNN architecture benefits from incorporating the feature pyramid network (FPN) and cascading pyramid network (CPN) modules, and d) how backbone depth and the additional modules affect deployment speeds. 3. We found that the deeper network provides a slight advantage for classification accuracy, while the shallower network with higher task similarity produces a slight advantage for object localization. The additional modules provided dramatic gains for the 18 layer backbone for both classification and localization. On a NVIDIA 1080-ti gpu, the 18-layer backbone trains ~ 30% faster than the 50-layer backbone. In deployment the 18-layer backbone is 2.5x faster than the 50-layer backbone, and 9.4x faster than Megadetector. These results show that backbone network task similarity, paired with the FPN and CPN modules, can substitute for depth, which improves deployment speeds. Our model is suitable for modest hardware and for integration into more complex pipelines. These are important steps towards the automation of data acquisition from camera trap images.

Hosted file

Automated_Mammal_Detection_NE_US.docx available at https://authorea.com/users/493131/articles/575645-automated-mammal-localization-and-identification-in-camera-trap-images-for-the-northeastern-united-states

Hosted file

Figure 1.docx available at https://authorea.com/users/493131/articles/575645-automated-mammal-localization-and-identification-in-camera-trap-images-for-the-northeastern-united-states

Hosted file

Figure 2.docx available at https://authorea.com/users/493131/articles/575645-automated-mammal-localization-and-identification-in-camera-trap-images-for-the-northeastern-united-states

Hosted file

`Figure 3.docx` available at https://authorea.com/users/493131/articles/575645-automated-mammal-localization-and-identification-in-camera-trap-images-for-the-northeastern-united-states

Hosted file

`Figure 4.docx` available at https://authorea.com/users/493131/articles/575645-automated-mammal-localization-and-identification-in-camera-trap-images-for-the-northeastern-united-states

Hosted file

`Figure 5.docx` available at https://authorea.com/users/493131/articles/575645-automated-mammal-localization-and-identification-in-camera-trap-images-for-the-northeastern-united-states