Species and Gender Identification from Camera Trap Data for Roe Deer (Computer Vision)

Nhung Vu
Department of Computer Science
Stanford University
nhungvu@stanford.edu

1 Introduction

Roe deer (Capreolus capreolus) is a synanthropic species widespread across Europe. The species thrives in forested areas and agricultural land and population numbers have increased steadily. To limit population densities, roe deer is culled in most areas of its habitat. Estimating the population size for deer management plans can be challenging, and population sizes are frequently underestimated. Photo traps are a useful tool for hunters to determine population sizes and gender distribution, but the amount of data with large amounts of false positives generated by detection of other species or vegetation moving in the wind can be infeasible for manual review and analysis. It is desired to automate the detection of number and gender of individuals of the species Capreolus capreolus in noisy imagery of photo traps to aid deer management efforts. To this end, photo trap images are used as input to a deep learning architecture. The output of the system will be axis-aligned bounding box coordinates and gender labels (male, female) for individual roe deer in the image.

2 Related work

High accuracy is achieved by studies on the large Snapshot Serengeti dataset with over 1 million images using deep learning architectures [1,2], but this dataset does not include bounding box labels, so counting individuals of a species is treated as a classification task, and gender identification is not attempted. Studies on specific species without a dataset of this size rely on transfer learning of object detection networks, but with few samples per species detection accuracy remains lower [3,4]. Authors cope with smaller numbers of samples per class combining datasets of multiple species under more general labels such as "deer" [5] or even "animal" [6] or by simplifying the task to identify presence/absence rather than number of individuals and bounding boxes [4,6,7]. The best results have been achieved in a large scale study attempting to identify the French fauna. With a dataset of 24.000 roe deer images the authors achieve 92% identifying bounding boxes [8]. To my knowledge no attempt at gender classification for roe deer has been undertaken. This project aims to improve on prior results by contributing a new dataset of hand-labeled roe deer images with bounding boxes and gender labels and demonstrating the use of deep learning for gender identification.

3 Dataset and Features

I have a 15,000 image dataset from photo trap data from hunting grounds near Steyr, Austria collected between 2019 and 2022. 11,800 images have been annotated with bounding boxes and gender labels for roe deer by me in about 25 hours. Around 4,000 images of this set contain one or more individuals of the target species. The data comprises several different camera locations and a large variety of daytimes, lighting conditions, weather conditions and is evenly distributed across the seasons. This is important since roe deer changes its appearance, growing a distinct winter coat in September/October and shedding it again in April. Four different classes have been labelled: FemaleRoeDeer, MaleRoeDeer, JuvenileRoeDeer, UnknownGenderRoeDeer. The distribution of labels is 2352, 1208, 639 and 435 respectively for female, male, unknown and juvenile. Roe deer is labelled male if the reproductive organs or the antlers are visible in the image. Roe deer is labelled female if no antlers are visible during the period where antlers are expected, or if a distinct brush of hair over the female reproductive organs is visible in the winter coat. Roe deer is labelled as juvenile when the typical spotted juvenile coat is visible, or when size or head shape allow for the conclusion that the individual roe deer is juvenile. Generally, juveniles are only labelled as such between April and November, since after November the size difference with adults is not clear enough and the winter coat allows for identifying the gender. Low-cost non-professional hardware is used for image acquisition, representative for the hardware likely being used by hunters. Image quality is therefore often low. Images include many examples with foggy lenses, rain drops and motion blur. Image quality differs across the different camera models used in the image capture phase, especially for night-time photos. Predicting gender for deer is challenging because the antlers that easily identify males are shed in the beginning of November and newly grown antlers start to be visible only by end of January. In the remaining months visual cues for gender identification are much more subtle, as demonstrated in Figure 1.

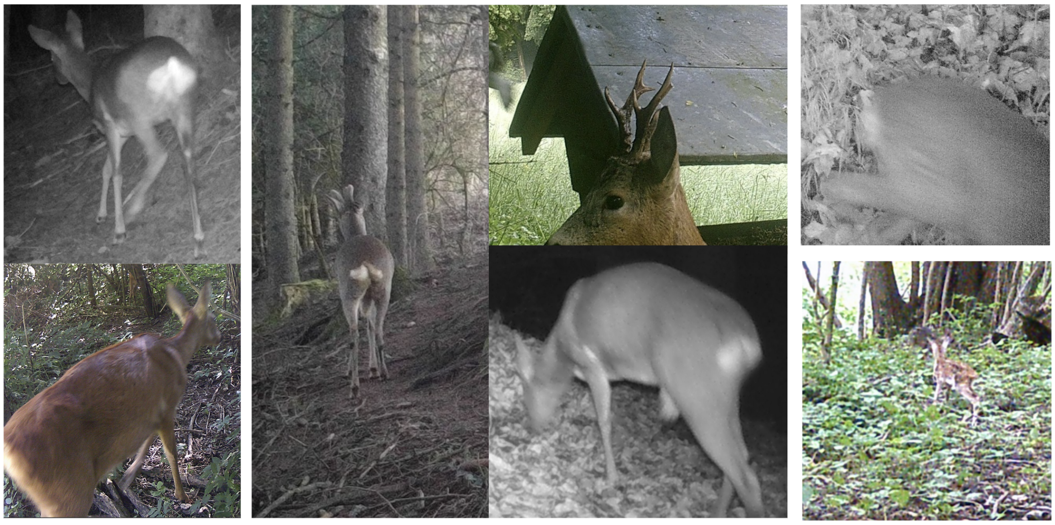


Figure 1: female roe deer in winter coat with brush of hair over female reproductive organ (top left). Female roe deer without antlers in summer coat (bottom left). Male roe deer with velvet antlers in winter coat (center left). Male roe deer with antlers in summer coat (center top right). Male roe deer with visible reproductive organ in winter coat without antlers (center bottom right). Roe deer of unknown gender since no features can be recognized (top right). Juvenile roe deer in spotted coat (bottom right).

4 Baseline

As a baseline the YOLOv7 algorithm was used [9]. Pretrained weights for YOLOv7 were available from the COCO image dataset [10]. The pretrained network was then transfer-learned using the hand-labelled roe deer dataset. A first attempt at establishing a baseline was undertaken with an inital dataset of 1,600 training images and 400 images in the dev set, a third of which containing deer. Due to memory constraints images were downsampled to 736x736 pixels and trained in batches of 16. Apart from necessary changes to account for the different number of classes, the parameterization for the COCO image dataset was kept. The network is not well able to recognize examples labelled as unknown gender and juveniles. Juveniles is the rarest of the three classes, and examples of unknown gender are frequently examples with low image quality that made labelling the gender impossible. This may explain the relatively higher share of false negatives for these two classes with 47% of unknown gender examples and 50% of juvenile examples not recognized. The baseline experiment was repeated with a larger number of labelled examples. Of the set of 11,800 images, 3,165 training images and 678 dev amd 678 test images were picked with each set containing 90% of roe deer images, resulting in a much higher number of examples per class for both test and training. This followed a recommendation by the YOLOv7 authors to include around 10% of background images. On more data the model exhibited reasonable performance with 0.664 mAP@0.5 and much reduced false negative rates for the difficult classes of juvenile and unknown gender.

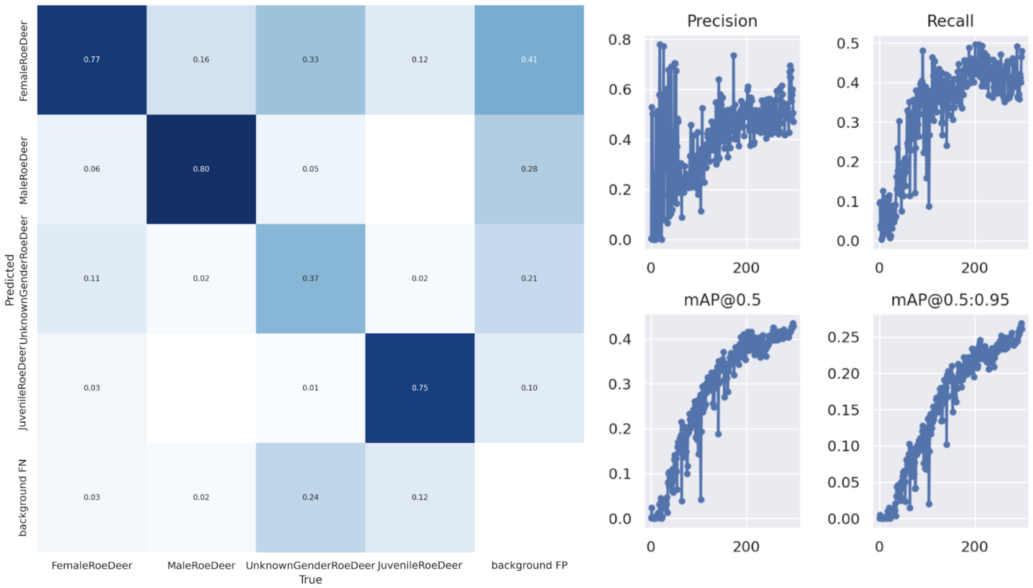


Figure 2: Result summary for YOLOv7 baseline

5 Error analysis

While the baseline experiment shows that is possible to train a network to detect roe deer gender, further accuracy increases are desired. Error analysis shows that there are 24% of unknown gender examples that are not detected. These examples include a large share of examples with extreme blur, deer being outside the range of the infrared flash in night time or deer being so far away as to become unrecognizable in the downsampled image. A large bayes error is to be expected for these examples, and gains there may not be of much practical significance since these examples will not help in analysing gender distribution even if classified correctly. Another frequent source of errors are multiple classifications of the same deer. It appears that for some hard examples multiple anchor boxes, usually two, are fitted to the same deer. When probabilities are distributed evenly both boxes survive the filter step. See an example for this in Figure 2. Another opportunity to improve the baseline is in the high number of individuals labelled as unknown gender that were detected as female. Of all individuals of unknown gender, 37% were correctly classified, 31% were misclassified as female but only 5% were misclassified as male. Analysis shows that there are several such examples where a deer with visible head but without antlers is classified female, but the image was taken between November and January and missing antlers are therefore not a reliable indicator of a female gender. See examples for deer misclassified as female between November and January in Appendix I.

6 Improvements to YOLOv7

A number of improvements where done within the original Yolov7 framework.

6.1 Class-agnostic non-max-suppression

Yolov7 runs non-max-suppression to merge or discard overlapping bounding boxes. By default non-max-suppression only considers bounding boxes with high IoU and the same class as duplicates.

If class labels are different, both will be kept. This does not work well for visually similar classes such as “RoeDeerFemale” and ”RoeDeerUnknownGender”. For this project a class-agnostic non-max suppression is performed. This means that non-max-suppression only considers IoU but not class, and duplicates are discarded on that basis. Since all classes in this detection problem are visually similar, this approach gives the desired result of a single label per class. It has to be noted that precision deteriorated to 0.64 mAP@0.5. This is because the YOLOv7 evaluation considers multiple predictions for the same labelled object and evaluates a correct match if any of the candidates carries the right class. For a practical application, class-agnostic non-max-suppression is still the more useful choice since single labels per object are desired.

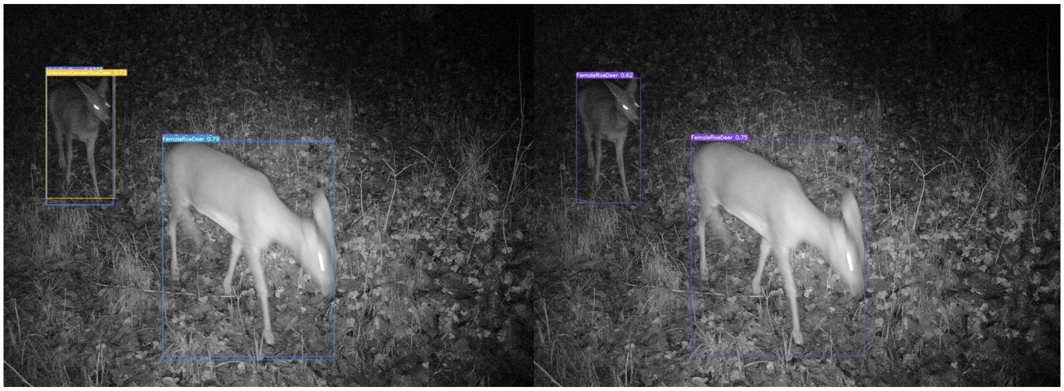


Figure 3: Example with non-max-suppression (left) and class-agnostic non-max-suppression (right)

6.2 Reduced image-augmentation

YOLOv7 by default uses a wide array of image augmentation techniques during training. Some of these techniques involving random cropping and translations may result in important gender identification features being lost. On resulting images the gender may not be identifiable anymore and the "UnknownGender" label may be more appropriate then the original gender label. The image augmentation procedure in YOLOv7 was reduced to only include techniques that are less likely to hide important gender features. Remaining augmentation techniques include small variations in brightness, variations in color space, small rotations, mosaics without cropping and flipping the inputs across a vertical axis.

6.3 Hyperparameter search

I conducted a random hyperparameter search for the single-stage Yolov7 detection. Parameters with the most significant impact on the results were the NMS parameters for IoU thresholds and confidence thresholds. Capping the ADAM learning rate, using multiscale detection or changes to the convolutional architecture did not result in significant changes.

My best experiment yielded a mean average precision at 0.5 recall (mAP@0.5) of 0.709. The improvements are therefore shown to compensate for the drop in KPIs due to the class-agnostic NMS and improve substantially on the baseline of 0.664 mAP@0.5. See Appendix II for precision/recall plots of the best single-stage detector.

7 Second-stage classification for gender

The best Yolov7 experiment still suffers from a relatively low accuracy for the difficult class of "unknown gender" with 0.352 mAP@0.5. One critical information that Yolov7 is missing is the month in which the image was taken. Knowing the month helps to decide whether a picture of a deer without antlers and without other visible gender traits is an image of a female deer (during the months in which males are likely carrying antlers) or an unknown gender (during November, December and January where no visible antlers are expected to be seen for males). To improve on this I experimented with a second stage classifier. This second stage classifier is a ResNet50V2 base network pretrained on ImageNet. After the base network global average pooling and dropout is applied. Then the month in which the image is taken is concatinated with the resulting feature map as a second input. A dense hidden layer and a softmax output layer are applied to this, yielding a prediction for the gender. The schematic for this architecture is shown in Appendix I. The second stage classifier turned out to be prone to overfitting and sensitive to various hyperparameter choices. Random search was done over the number of layers and neurons in the final connected layers, the ADAM learning rate cap, batch size, and the choice of base network architecture. Also different training procedures were tested, with the best results achieved by first training only the head to convergence at about 0.7 accuracy, then fine-tuning by allowing weight updates to the top layers of the base network. Image augmentation including flipping, rotations, small translations within the bounding box margins, brightness and contrast variations were applied. In the training set, underrepresented examples where upsampled. In the dev set the original distribution was kept. The training data and validation data where generated from the labels of the Yolov7 training and validation sets, keeping the same test and validation split. The best experiment yielded an accuracy of 0.78. Unfortunately, connecting this second stage classifier to Yolov7 yielded a lower mAP@0.5 of 0.652, versus a mAP@0.5 of 0.709 for my best single stage network. While the confusion matrix shows a significantly higher true positive rate of 0.49 (versus 0.37 from single stage) for the unknown gender class, a slightly lower true positive rate for the overrepresented female class cancelled out any gains. See Appendix IVprecision vs recall plots, and Appendix III for details on the architecture.

8 Second-stage classification for whether gender is identifiable

A variation on the first idea is to use the second stage classifier to only check whether or not the gender is identifiable in the image. For this, the described second stage architecture was changed slightly to a binary classifier by switching the loss to binary crossentropy and shrinking the output layer to a single neuron. Labels of "female", "male" and "juvenile" were treated as true labels, "unknown gender" as false labels. Hyperparameter search again included learning rate, batch size, decision head architecture, augmentation and number of epochs in pretraining and finetuning. Since only 14% of examples in the training and validation data contained unknown gender examples I had to make sure the classifier would not achieve high accuracy by always predicting positive. To prevent this I balanced the examples in the training set by upsampling underrepresented classes to achieve completely balanced classes in the training set, but kept the original distribution in the dev set. The best classifier achieved an accuracy of 0.875 on the dev set, with a 0.99 accuracy on the training set, indicating that the classifier was not only predicting positive. The Yolov7 prediction is updated to "unknown gender" if the second stage classifies an image as "gender not identifiable" with high confidence. This algorithm beat the single-stage Yolov7 slightly with an mAP@0.5 of 0.713, due to a higher precision for unknown gender of 0.374 vs 0.352. See Appendix III for precision vs. recall plots of the final result.

9 Contribution

This is single-contributor project. My contribution includes preparing and labelling the dataset, preparing the Yolov7 baseline, improving the Yolov7 single stage detector by improvements to NMS, augmentation and hyper-parameter tuning, and implementing a second stage classifier that can run in end-to-end test and detection procedures. I have also prepared the report, video and presentations used for the status meetings.

10 Conclusion

To my knowledge I demonstrate for the first time how roe deer gender can be identified in camera trap images. For this I prepared a novel handlabelled dataset of 11.800 images containing 5.644 labelled deer. While a single stage Yolov7 detector gives good results, slight improvements for the difficult case of "unknown gender" can be achieved by a two-stage detector where the second stage takes the cropped bounding box and the month in which the image is taken as inputs and outputs whether gender is identifiable. I achieve an mAP@0.5 of 0.713 in my best experiment detecting the four classes "FemaleRoeDeer", "MaleRoeDeer", "UnknownGenderRoeDeer" and "JuvenileRoeDeer".

Code

Soure code is forked from the YOLOv7 repository [9] and modified, can be accessed on github [11]. See the README.md for additional information.

References

[1] Mohammad Sadegh Norouzzadeh, Anh Nguyen, Margaret Kosmala, Alexandra Swanson, Meredith S. Palmer, Craig Packer and Jeff Clune. "Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning". In: *Proceedings of the National Academy of Sciences, 115 (25) E5716-E5725* (2018).

[2] Alexander Gomez, Augusto Salazar and Francisco Vargas. "Towards Automatic Wild Animal Monitoring: Identification of Animal Species in Camera-trap Images using Very Deep Convolutional Neural Networks". In: *arXiv:1603.06169v2* (2016).

[3] Matthew T. Duggan, Melissa F. Groleau, Ethan P. Shealy, Lillian S. Self, Taylor E. Utter, Matthew M. Waller, Bryan C. Hall, Chris G. Stone, Layne L. Anderson, Timothy A. Mousseau. "An approach to rapid processing of camera trap images with minimal human input ". In: *Ecology and Evolution 17 (11)* (2021).

[4] Ryan Curry, Cameron Trotter and Andrew Stephen McGough. "Application of deep learning to camera trap data for ecologists in planning / engineering – Can captivity imagery train a model which generalises to the wild?". In: *arXiv:2111.12805* (2021).

[6] Mitchell Fennell, Christopher Beirne, Cole A. Burton. "Use of object detection in camera trap image identification: Assessing a method to rapidly and accurately classify human and animal detections for research and application in recreation ecology". In: *Global Ecology and Conservation 35 (June)* (2022).

[7] Olivier Gimenez, Maëlis Kervellec, Jean-Baptiste Fanjul, Anna Chaine, Lucile Marescot, Yoann Bollet and Christophe Duchamp. "Trade-off between deep learning for species identification and inference about predator-prey co-occurrence: Reproducible R workflow integrating models in computer vision and ecological statistics". Online: *https://oliviergimenez.github.io/computo-deeplearning-occupany-lynx* (2022).

[8] Rigoudy et al. "The DeepFaune initiative: a collaborative effort towards the automatic identification of the French fauna in camera-trap images". In: *https://www.biorxiv.org/content/10.1101/ 2022.03.15.484324v1.full.pdf* (2022).

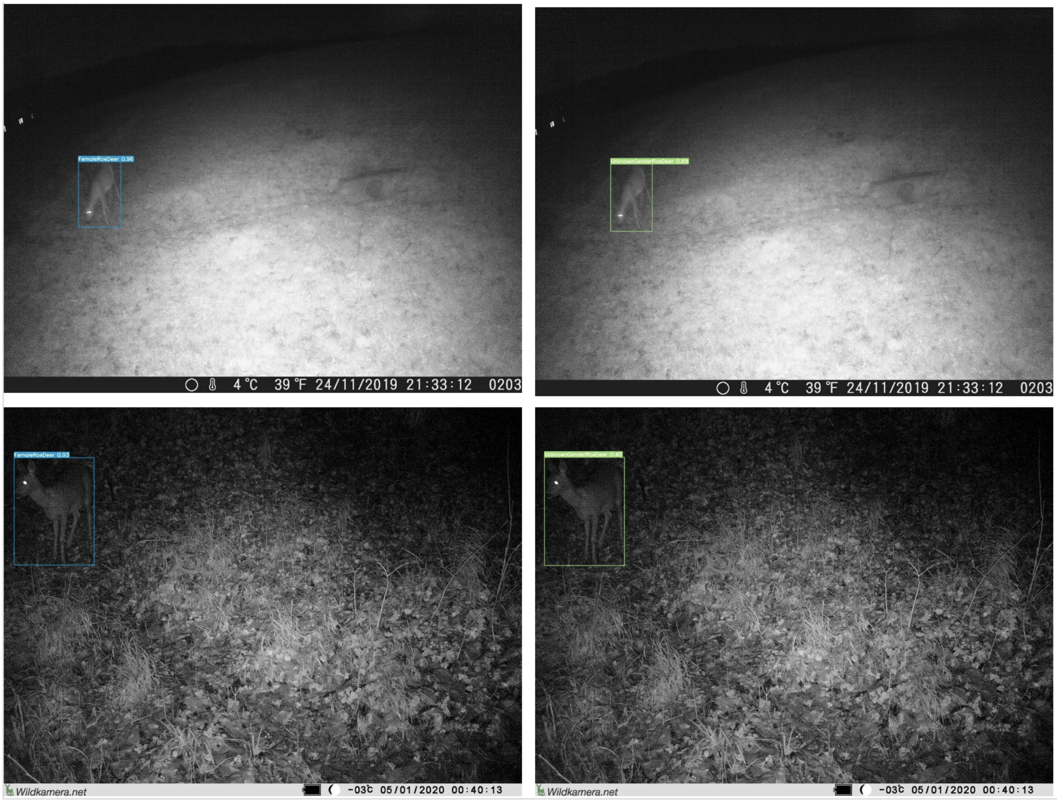
[9] Chien-Yao Wang, Alexey Bochkovskiy, Hong-Yuan Mark Liao. "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors". In: *https://arxiv.org/abs/2207.02696* (2022).

[10] Tsung-Yi Lin, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, Piotr Dollár. "Microsoft COCO: Common Objects in Context". In: *https://arxiv.org/abs/1405.0312* (2015).

[11] Nhung Vu. "Roe Deer Recognition". Source code repository: https://github.com/nhung-huyen-vu/roe-deer-recognition

Appendix I: Misclassified unknown gender examples

The following two images show deer without antlers between November and January. Since other features such as reproductive organs are not visible, gender would need to be inferred as "Unknown". The baseline detector classified them as "Female" (left) likely due to missing antlers. My final network correctly classifies them as "Unknown" (right)



FemaleRoeDeer 0.96
UnknownGenderRoeDeer 0.69
4 °C 39 °F 24/11/2019 21:33:12 0203
4 °C 39 °F 24/11/2019 21:33:12 0203
FemaleRoeDeer 0.93
UnknownGenderRoeDeer 0.40
Wildkamera.net
-03°C 05/01/2020 00:40:13
Wildkamera.net
-03°C 05/01/2020 00:40:13

Figure 4: Deer in winter without visible gender features misclassified as female by the baseline (left) and correctly classified as unknown by my final detector

Appendix II: Precision, recall and mAP@0.5 for best single-stage classification

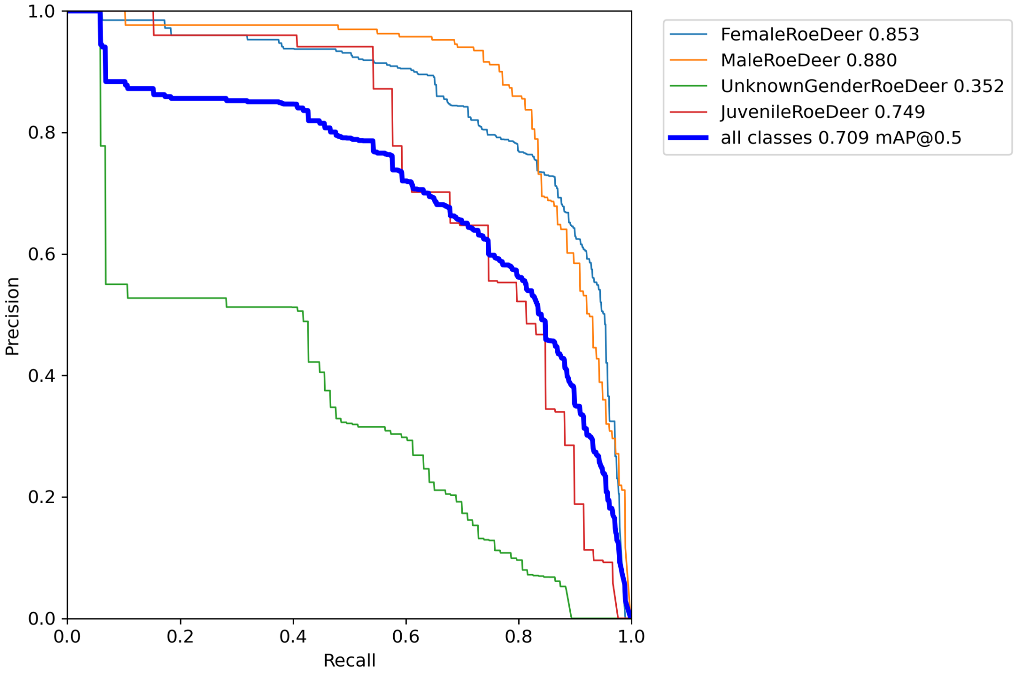


Figure 5: Best single stage result with a mAP@0.5 of 0.709

Appendix III: Architecture for two-stage classification

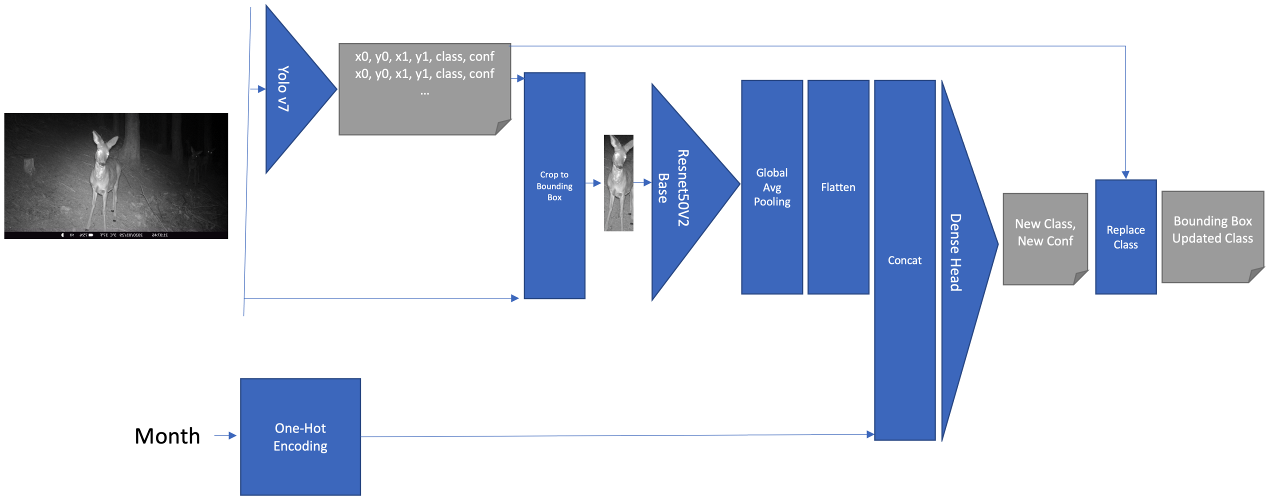


Figure 6: Schematic architecture for two-stage classification

Appendix IV: Precision, recall and mAP@0.5 for two-stage gender classification

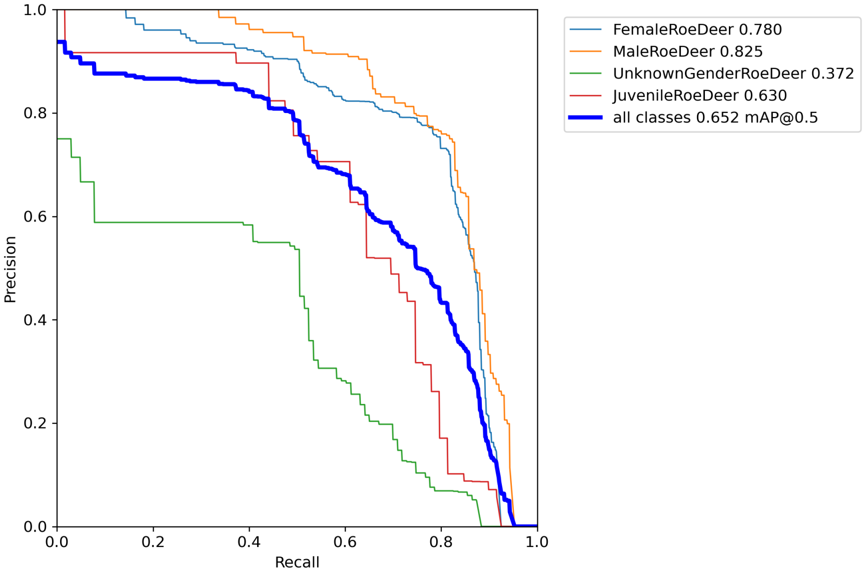


Figure 7: precision vs recall for two-stage gender classification with second stage predicting the gender. Note the improvement in mAP for the unknown class and the deterioration in the female and male classes.

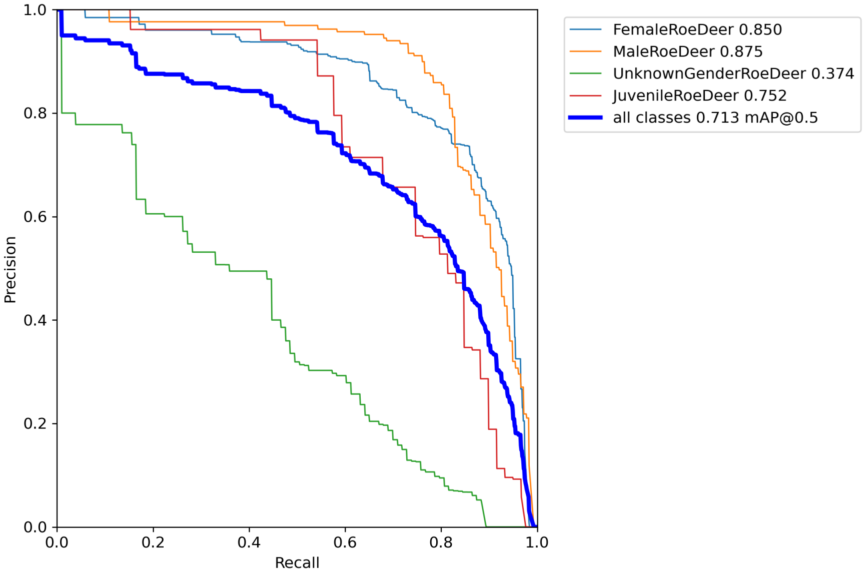


Figure 8: precision vs recall for two-stage gender classification with second stage predicting whether gender is identifiable. Note the improvement in mAP for the unknown class nearly identical performance in male and female classes compared to single-stage results.