Manaaki Whenua
Landcare Research

Artificial intelligence for vertically oriented cameras

Prepared for: Hawke’s Bay Regional Council

August 2022

Artificial intelligence for vertically oriented cameras

Contract Report: LC4190

Brent Martin and Al Glen

Manaaki Whenua – Landcare Research

Reviewed by:	*Approved for release by:*
Patrick Garvey Scientist – Wildlife Management and Conservation Ecology Manaaki Whenua – Landcare Research	Chris Jones Portfolio Leader – Wildlife Management and Conservation Ecology Manaaki Whenua – Landcare Research

Disclaimer

This report has been prepared by Manaaki Whenua – Landcare Research for the Hawke's Bay Regional Council. If used by other parties, no warranty or representation is given as to its accuracy and no liability is accepted for loss or damage arising directly or indirectly from reliance on the information in it.

Contents

Summary

Project and client

The Hawke's Bay Regional Council contracted Manaaki Whenua – Landcare Research to investigate the feasibility of detecting possums, and potentially other pest species, from vertical camera trap images using artificial intelligence.

Objectives

- Produce a prototype image recognition model that can distinguish possums from other animals in vertical camera trap images and reject empty images.
- Investigate the model's ability to identify other species of interest.
- Quantify the accuracy of the model.

Methods

Manaaki Whenua – Landcare Research has been developing an image recognition model to detect common animal species from camera trap images, taken both during the day and at night. The model automatically assigns a species label to a new image (e.g. stoat, cat, possum, etc.) indicating the most likely animal in the image. The model also returns a confidence score, which indicates how certain the model is that it has identified the species correctly. If the confidence scores for all species are low, it is likely the image is empty.

This model was extended to include detection of possums and other species with vertically oriented cameras using the following method.

- Obtain images of possums and other species, as well as empty images, from vertical camera trap images provided by Hawke's Bay Regional Council
- Sort the images into day and night images
- Sort a subset of the *night* images into species categories
- Close-crop a subset of images around the target animal and crop an empty area from each image (negative crop). Add the cropped vertical images to the original set of horizontal images previously used to train the model, giving an extended training set.
- Retrain the model using the extended training set and test it on an independent test set of vertical images.
- Measure and analyse the predictive performance for detecting possums and other target species. Predictive performance was assessed using two metrics:
  - a recall: also known as the true positive rate, this is the proportion of possum images correctly identified. It is calculated by dividing true positives by the total number of possum images.
  - b precision: the proportion of images labelled by the model as "possum" that truly contained a possum. It is calculated by dividing the true positives by overall positives.

- Measure the model’s performance at excluding empty images and images containing non-target species.

Results

- When retrained on a training set consisting of around 6,000 horizontal camera images infused with an additional 1,327 vertical camera images (including 124 possum images) and tested on 1,200 independent vertical images, the model correctly identified 78% of the images containing possums, with a precision of 63%.
- The false positive rate for detecting possums in empty images was less than 5%.
- Other species were detected with varying levels of recall and precision.
- Training a model on vertical camera images only resulted in much poorer performance.
- Performance of the model was limited by the small size of the training set; further training is likely to improve the model’s accuracy. However, even with this small additional training set, the model achieved comparable accuracy for possums on vertical images as was originally achieved for horizontal images, when trained using both horizontal and vertical images. Increasing the training set size for both horizontal and vertical images is recommended to increase the performance of possum detection overall.

Conclusions

- A deep learning image recognition model trained to detect target species in horizontal camera images can be enhanced to detect possums and some other species in vertical camera imagery by retraining it with a combination of horizontal and vertical images. This demonstrably improves classification performance on vertical camera images without negatively impacting horizontal image classification performance.
- Prediction accuracy can be achieved by adding as few as approximately 100 possum images; better results should be attainable from a larger training set. Performance may also be enhanced through careful selection of the training images.
- The model can be adjusted to favour precision over recall (or vice versa) according to the user’s needs. Favouring precision would reduce the amount of manual image processing required, and may be appropriate when estimating relative abundance, when a small number of missed detections might be acceptable. Favouring recall would require more manual processing but minimise the probability of missed detections. This may be appropriate in an eradication situation, when detecting every individual target animal is important.

Recommendations

This project has demonstrated the feasibility of automatically detecting possums in vertical camera trap images by developing a prototype detection model and testing it on a small set of representative images. This model lays the groundwork for a system that could be deployed for use in the field, but would benefit from further validation and enhancement to further improve performance and quantify accuracy:

- Further test the model using images from a source that is completely independent of the image sources used for this model.
- Further optimise the parameters used during training, including image augmentation.
- Increase the training set size, measure how the accuracy of the model improves as the size of the training set increases, and undertake a statistical analysis of the model's performance with respect to training set size.

1 Introduction

The Hawke's Bay Regional Council (HBRC) contracted Manaaki Whenua – Landcare Research to investigate the feasibility of detecting possums, and potentially other pest species, from vertical camera trap images using artificial intelligence. Manaaki Whenua have previously trained a model to detect possums and other species from camera trap images, but the cameras were all horizontally mounted. HBRC wanted to assess whether the model would work for vertically mounted cameras and, if so, what would be required.

2 Background

There is considerable interest in detecting species in camera trap images. *Computer vision* is a field of computer science that attempts to automatically label images according to their contents. Such algorithms are increasingly being used to automate the labelling of camera trap images. Recently, *deep learning* algorithms (a branch of artificial intelligence) have been developed which can be trained to label images with much higher accuracy than before. Nonetheless, detecting animals in camera trap images remains a difficult problem because of the highly variable nature and quality of imagery. Several studies have been conducted in New Zealand for various species, with mixed results. Whereas many common computer vision applications work on high-quality images, camera trap images may include animals that are far away from the camera, poorly lit or over-exposed, substantially obscured (or mainly out of frame), and of low resolution. Also, night images are black and white, and are often very low quality. Figure 1 shows examples.

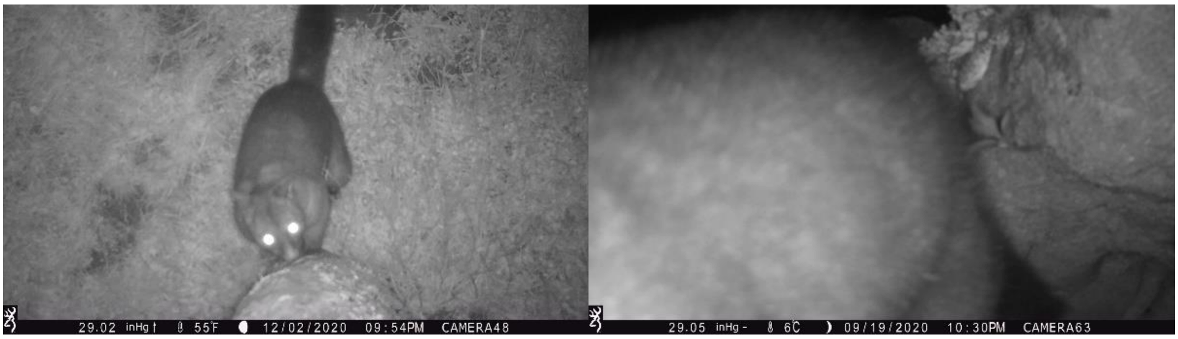

29.02 inHg 55°F 12/02/2020 09:54PM CAMERA48
29.05 inHg - 6°C 09/19/2020 10:30PM CAMERA63

Figure 1. Examples of high- and low-quality vertical possum images.

Deep learning-based image recognition models are typically trained on a large set of diverse images; the more diverse the images, the more general (and therefore robust) the prediction performance. For camera trap images, while a large number may be collected, they are typically in bursts of very similar images, and the images from each camera deployment share a common background. This makes learning from these images more difficult because strong features in the background may appear predictive of species if, for example, a particular camera photographs some species more commonly than others. The camera may also have other peculiarities, such as dirt on the lens, that the algorithm incorrectly focusses on when labelling the image. For this reason, deep learning computer vision models tend to perform poorly when trying to identify species in individual camera trap images. However, the fact that each animal visit is captured by a burst of more than one image can be beneficial when performing detection, because the predictions for each image in a burst can be combined to classify the burst overall (e.g. by selecting the species predicted with the highest individual score). This approach is commonly used.

Vertical images differ from horizontal ones mainly in the poses the animals present, which fall into two categories: those where the animal is observed on the ground from above, and those where the animal is climbing up or down the object (e.g. tree) on which the camera is mounted, so are filmed from in front or behind. These poses are very different from those seen in horizontal imagery. A further complication is that the image background looks very different, and often contains a strong irrelevant feature, i.e. the tree, post etc that the camera is mounted on; this increases the possibility of the model incorrectly focussing on the mounting object instead of the animal.

For these reasons, we expected the existing model trained on horizontal images to perform poorly on images from vertically mounted cameras. In this research we explored whether these difficulties can be overcome such that possums (and potentially other species of interest) can be detected to a useful accuracy without requiring excessive effort to collect and process large amounts of imagery for training the model.

3 Objectives

The goal of this project was to train AI software to: (1) recognise possums in vertical camera trap imagery, (2) differentiate between images of possums, empty images, and images of other animals, and (3) quantify the model's level of accuracy.

4 Methods

We used a deep learning computer vision model to detect possums in camera trap images, with some refinements to overcome the difficulties we have previously described. The model labels animals in images as belonging to one of 12 categories: bird, cat, ferret, hedgehog, kiwi, lagomorph, livestock, mouse, possum, rat, stoat, and wallaby.

It is difficult to reliably detect animal species in camera trap imagery because of the large degree of variability in the imagery, in terms of both the quality and the range of conditions (camera setup, lighting, distance from camera to target) of the images. In particular, a small animal, such as a bird or mouse, may take up less than 1% of the image area, meaning that the number of pixels available for prediction is small and, more importantly, over 99% of the image is irrelevant to the prediction. Worse, the model may be biased to predict species based on the *background*.

To overcome this limitation, our model predicts the most likely species present in each of a set of 117 overlapping image crops of size 299×299, extracted from the images after first rescaling them to three different sizes (image heights) of 450, 600 and 900 pixels. During prediction, the model assigns to each crop a confidence score for each species based on how well the image crop matched the model for that species. For each target species, the crop with the highest confidence prediction is selected, with the species that achieved the highest confidence score overall being selected. A further step is then to decide whether to accept or reject the prediction: if the confidence score exceeds a threshold parameter, the prediction is accepted; otherwise, it is rejected, meaning either the image is empty or a target animal was not reliably determined.

The model is trained on tightly cropped examples of the target species, plus a second 'empty' crop from near the target animal; the purpose of the second crop is to minimise the likelihood of the model training on the background features. The model is trained via *transfer learning*: a model already trained on millions of generic images is retrained to detect animals in camera trap images. The advantage of this approach is that it requires a much smaller training set.

Figure 2 shows an example possum image and the two crops extracted for training.

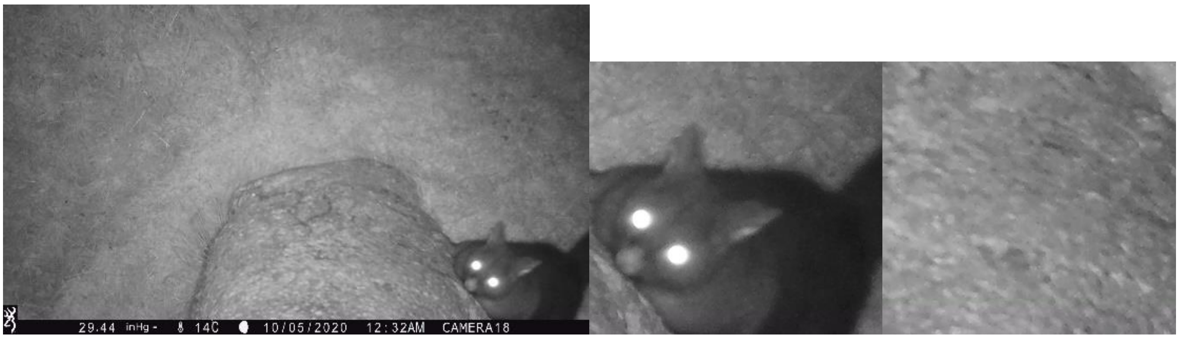


Figure 2. Example possum image (left) with corresponding possum (middle) and empty (right) crops.

4.1 Original model performance – horizontal images

The original model, trained on horizontal images only, was tested on a representative random sample of 100 day and night horizontal images per species, and a further 240 empty images, giving a total of 2,241 test images.

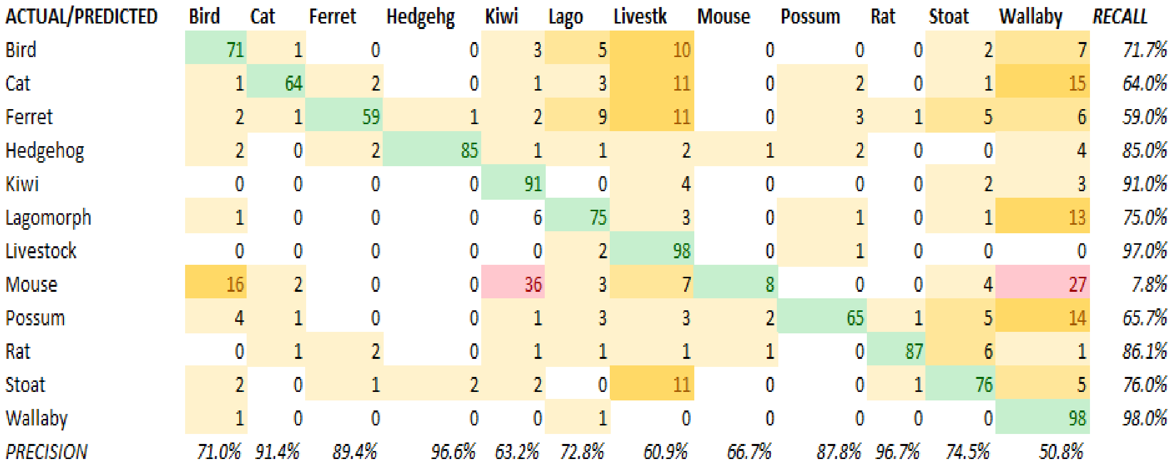
Predictive performance was assessed using two metrics:

1 recall: also known as the true positive rate, this is the proportion of images of a given species that were correctly identified.

2 precision: the proportion of images predicted to be a particular species that were correct.

Table 1 shows the confusion matrix for the model's predictions, for horizontal night images. Numbers on the diagonal (shaded green) are the number of correct predictions for each species, with the other numbers showing how images were mis-labelled. Both precision and recall vary between species, with livestock and wallabies having the highest recall, but hedgehogs having the highest precision.

Table 1. Prediction confusion matrix for horizontal night. Correct detections are shown in green; errors are coloured by magnitude (pink ≥25%, dark yellow ≥10%, mid yellow ≥5%, light yellow ≥1%). Row names (left) are actual species; column headings are the species predicted by the model. Integers are the number of images, e.g. 71 bird images were correctly classified, while 10 bird images were incorrectly classified as livestock



ACTUAL/PREDICTED	Bird	Cat	Ferret	Hedgehg	Kiwi	Lago	Livestk	Mouse	Possum	Rat	Stoat	Wallaby	*RECALL*
Bird	71	1	0	0	3	5	10	0	0	0	2	7	*71.7%*
Cat	1	64	2	0	1	3	11	0	2	0	1	15	*64.0%*
Ferret	2	1	59	1	2	9	11	0	3	1	5	6	*59.0%*
Hedgehog	2	0	2	85	1	1	2	1	2	0	0	4	*85.0%*
Kiwi	0	0	0	0	91	0	4	0	0	0	2	3	*91.0%*
Lagomorph	1	0	0	0	6	75	3	0	1	0	1	13	*75.0%*
Livestock	0	0	0	0	0	2	98	0	1	0	0	0	*97.0%*
Mouse	16	2	0	0	36	3	7	8	0	0	4	27	*7.8%*
Possum	4	1	0	0	1	3	3	2	65	1	5	14	*65.7%*
Rat	0	1	2	0	1	1	1	1	0	87	6	1	*86.1%*
Stoat	2	0	1	2	2	0	11	0	0	1	76	5	*76.0%*
Wallaby	1	0	0	0	0	1	0	0	0	0	0	98	*98.0%*
PRECISION	*71.0%*	*91.4%*	*89.4%*	*96.6%*	*63.2%*	*72.8%*	*60.9%*	*66.7%*	*87.8%*	*96.7%*	*74.5%*	*50.8%*	

The model achieved an overall accuracy of 73% for horizontal night images. These results compare favourably with prior research. Curran et al. (2022) achieved 75% accuracy using the Wellington Camera Trap (WCT) data set (Anton et al. 2018), but their model only partitioned images into three groups (bird, mammal, not interesting), which is a simpler task.

A more similar study was conducted by Shahinfar and Falzon (2021). Their model was commissioned by MWLR and tried to label a much larger image set comprising a combination of images provided by MWLR and the WCT images with one of 10 labels (bird, cat, hedgehog, kiwi, lagomorph, mustelid, nothing, others, possum, and rodent). Our model produced higher *recall* for six of the nine target species labels (i.e. ignoring the 'nothing' label) and had higher *precision* for all target species except bird. It should be stressed that the two studies are not directly comparable: as well as using different image sets, the target species are not quite the same.

Other related research has reported higher accuracies. For example, CameraTrapDetectoR (Tabak et al. 2022) reports precisions of 96% for high-level label detection (bird, human, mammal, vehicle) and 80% for species, but that study used a data set consisting of higher quality images.

4.2 Fine-tuning the model for vertical images

The original horizontal image model was then enhanced for prediction of both horizontal and vertical images by 'infusing' the original set of horizontal images with a small number of vertical night images. Infusion is a common method of generalising a deep learning model so that it transfers well to new image sets. The retrained model was tested on a hold-out set of vertical images, and also retested using the original horizontal test set to ensure overall performance has not degraded. The following process was used:

- Close-crop a subset of images around the target animal, and crop an empty area from each image (negative crop). Add the cropped vertical images to the original set of horizontal images previously used to train the model, giving an extended training set.
- Retrain the model using the extended training set and test it on an independent test set of vertical images.
- Measure and analyse the predictive performance for detecting possums and other target species. Predictive performance was assessed using two metrics:
  - recall: also known as the true positive rate, this is the proportion of possum images correctly identified.
  - precision: the proportion of images labelled by the model as "possum" that truly contained a possum.
- Measure the model's performance at excluding empty images and images containing non-target species.

Although we had a fairly large number of vertical night images (over 11,000), most of these were either empty or livestock. Further, the process of manually cropping training images is highly labour-intensive. For the initial retraining run, we therefore used a much smaller number of image crops: we trained the model on an image set that included 459 vertical images of target animals, including 126 vertical possum images, some of which were withheld to validate the model's performance during training. It should be noted that the images provided included bursts of very similar images so that, in practice, the number of substantially different images is much smaller. It is therefore acknowledged that the image set only represents a very small subset of the types of vertical images likely to be encountered, and is probably insufficient to train the model to a high level of accuracy.

4.3 Assumptions and limitations

This research relies on various assumptions that may limit the generality of the results. The main ones are:

1 the training image set is a sufficient representation of the images likely to be encountered.

2 the training and test images are sufficiently diverse to demonstrate the model's ability to generalise to novel images.

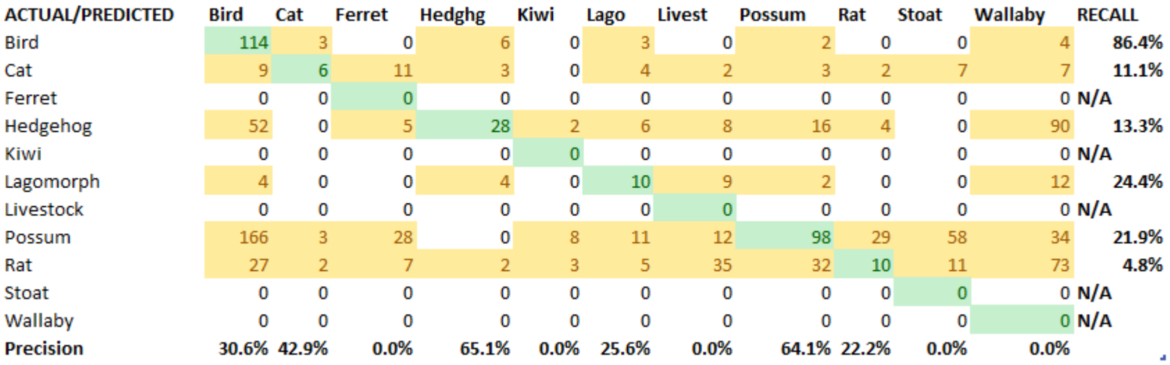
5 Results – detecting possums in vertical night images

5.1 Detecting possums using the current model

We first tested the ability of the current model (trained on horizontal images only) to reliably separate possums from other animals and empty vertical images. We anticipated that performance would be poor because vertical images look very different to the horizontal images used to train the model, including the presentation of the animals being very different.

The current model correctly classified only 11.5% of the 2,318 images containing animals, with 22% of possums being correctly classified. Table 2 shows the confusion matrix for the current model when run on the vertical images. Overall, just 24% of the species in the images were correctly identified. As can be seen from the confusion matrix, possums were most commonly misclassified as birds.

Table 2. Prediction confusion matrix for vertical night images



ACTUAL/PREDICTED	Bird	Cat	Ferret	Hedghg	Kiwi	Lago	Livest	Possum	Rat	Stoat	Wallaby	RECALL
Bird	114	3	0	6	0	3	0	2	0	0	4	86.4%
Cat	9	6	11	3	0	4	2	3	2	7	7	11.1%
Ferret	0	0	0	0	0	0	0	0	0	0	0	N/A
Hedgehog	52	0	5	28	2	6	8	16	4	0	90	13.3%
Kiwi	0	0	0	0	0	0	0	0	0	0	0	N/A
Lagomorph	4	0	0	4	0	10	9	2	0	0	12	24.4%
Livestock	0	0	0	0	0	0	0	0	0	0	0	N/A
Possum	166	3	28	0	8	11	12	98	29	58	34	21.9%
Rat	27	2	7	2	3	5	35	32	10	11	73	4.8%
Stoat	0	0	0	0	0	0	0	0	0	0	0	N/A
Wallaby	0	0	0	0	0	0	0	0	0	0	0	N/A
Precision	30.6%	42.9%	0.0%	65.1%	0.0%	25.6%	0.0%	64.1%	22.2%	0.0%	0.0%	

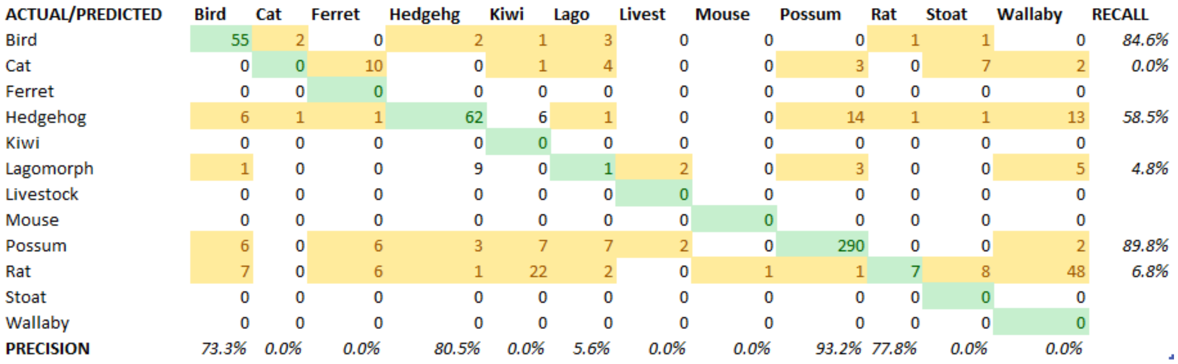
The model generates a confidence score for each prediction ranging from 0 to 1. Correct predictions generally have very high confidence. The confidence score can be used to discard predictions whose confidence falls below a threshold; this weeds out many of the incorrect predictions (raising the model's precision), and discards any false positive empty images, whose confidence scores should be very low.

We also tested the model's ability to exclude empty images. For each species, empty images were excluded if their prediction's confidence score was below the model's threshold for the predicted species. Of 1,226 empty images, 864 were excluded (70.5%). However, using this same test, 31% of images containing target species were also excluded, including 12.5% of the images containing possums. This is a very poor result.

5.2 Detecting possums using a retrained model

We retrained the model using the original (horizontal) images training set, infused with 459 vertical images, as previously described. Table 3 shows the confusion matrix for the test set of vertical animal images.

Table 3. Confusion matrix for the retrained model



ACTUAL/PREDICTED	Bird	Cat	Ferret	Hedgehg	Kiwi	Lago	Livest	Mouse	Possum	Rat	Stoat	Wallaby	RECALL
Bird	55	2	0	2	1	3	0	0	0	1	1	0	*84.6%*
Cat	0	0	10	0	1	4	0	0	3	0	7	2	*0.0%*
Ferret	0	0	0	0	0	0	0	0	0	0	0	0	
Hedgehog	6	1	1	62	6	1	0	0	14	1	1	13	*58.5%*
Kiwi	0	0	0	0	0	0	0	0	0	0	0	0	
Lagomorph	1	0	0	9	0	1	2	0	3	0	0	5	*4.8%*
Livestock	0	0	0	0	0	0	0	0	0	0	0	0	
Mouse	0	0	0	0	0	0	0	0	0	0	0	0	
Possum	6	0	6	3	7	7	2	0	290	0	0	2	*89.8%*
Rat	7	0	6	1	22	2	0	1	1	7	8	48	*6.8%*
Stoat	0	0	0	0	0	0	0	0	0	0	0	0	
Wallaby	0	0	0	0	0	0	0	0	0	0	0	0	
PRECISION	*73.3%*	*0.0%*	*0.0%*	*80.5%*	*0.0%*	*5.6%*	*0.0%*	*0.0%*	*93.2%*	*77.8%*	*0.0%*	*0.0%*	

The new model correctly identified 64% of the images containing animals, including 90% of the possum images. The precision of the possum prediction (i.e. the number of images identified as possum that actually contained a possum) was also high, at 93%, meaning only 7% of animal images predicted to be a possum were not. As can be seen from the confusion matrix, most of the misclassified animals were hedgehogs. Of the 1,225 empty test images, 188 were predicted to contain possums (15%).

The model can be further tuned to achieve an appropriate balance of precision over recall by rejecting all predictions whose confidence falls below a threshold. Using the original model's possum prediction confidence threshold of 0.8 gives good precision (93.7%) and low empty image inclusion (7.3%) with 87% recall. Raising the threshold to 0.9 for vertical possum images increases precision to almost 96%, and false positives (included empty images) drop to 4.5%, for a small drop in recall to 83.6%. Further raising the threshold to 0.95 reduces recall to 78%, but raises precision of animal species identification to 98%, and reduces the false positive rate to 2.2%.

5.3 Analysis of errors

What is the model 'looking at'? We analysed the incorrectly labelled images to confirm that the model is using the appropriate features (rather than, for example, using the background as a signal). This confirmed the model appears to be making predictions based on the animals themselves.

Figure 3 shows the correctly labelled possum images that returned the *highest* confidence (images have been cropped for presentation purposes). All the images are readily identifiable as possums and present typical features, such as legs and tail.

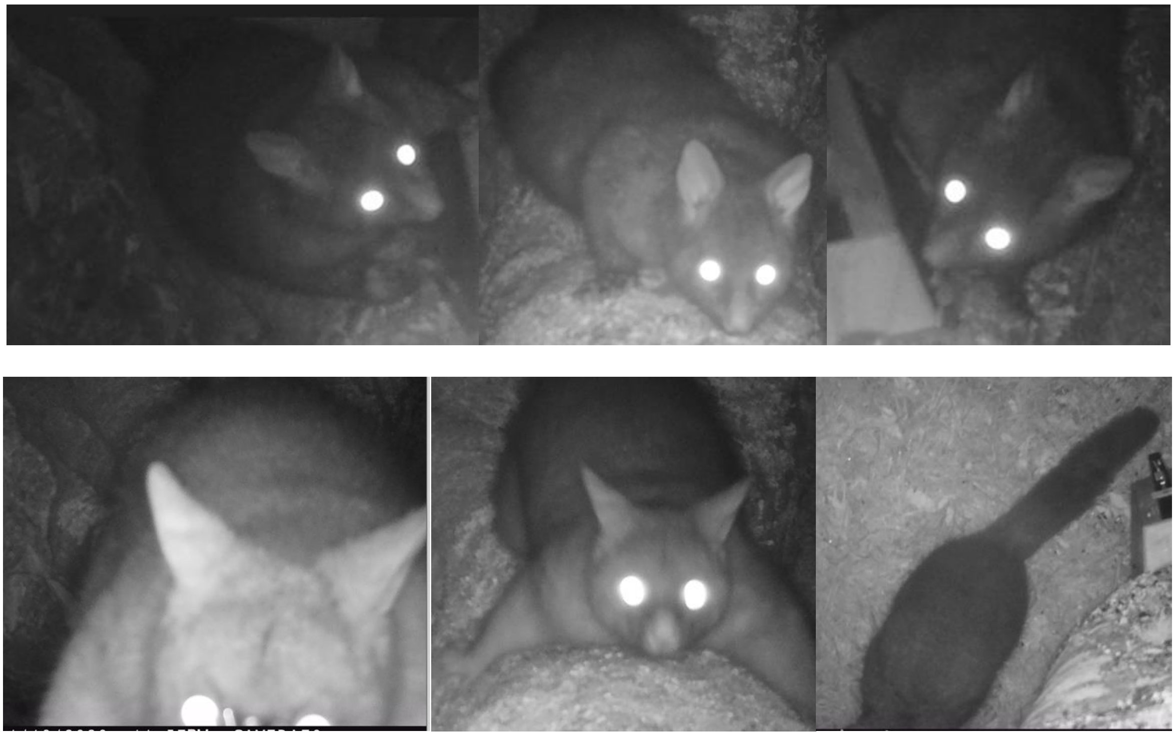


Figure 3. Possum images predicted with the highest confidence.

Most of the top-scoring images are of possums facing the camera, typically with eyes highlighted.

For all the lowest-scoring possum images, the possum is barely visible, including where only a very small part of the animal (such as the tail) is present. Figure 4 shows an example where only a portion of the possum's tail is present, which returned a low confidence score of 0.49.

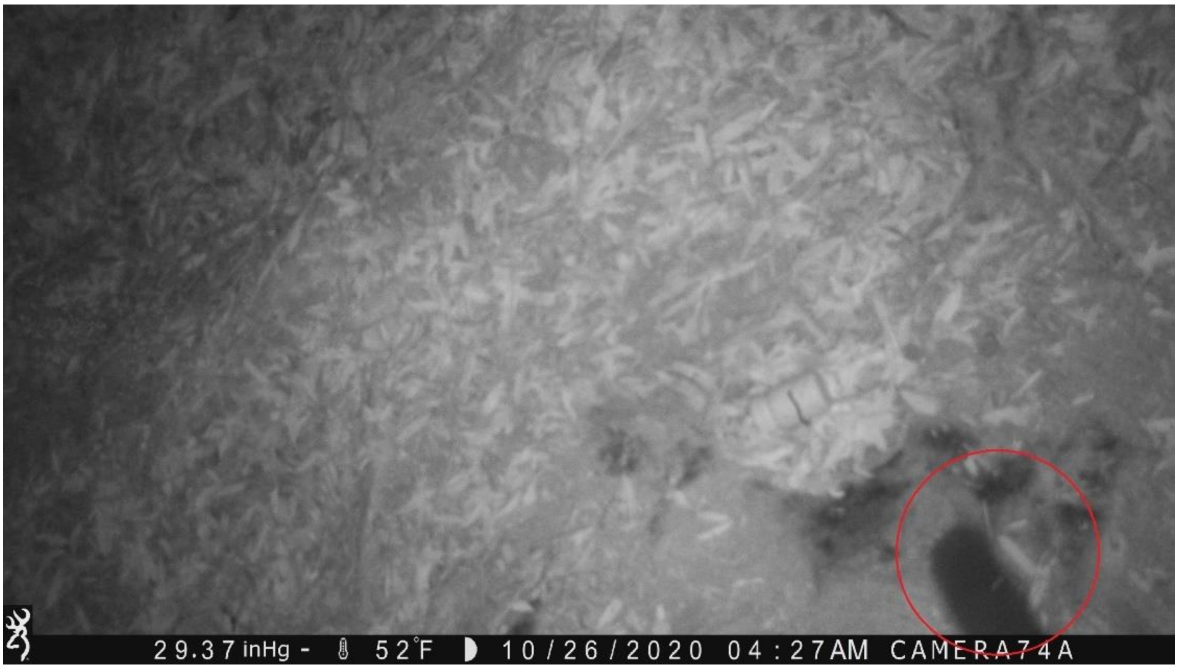
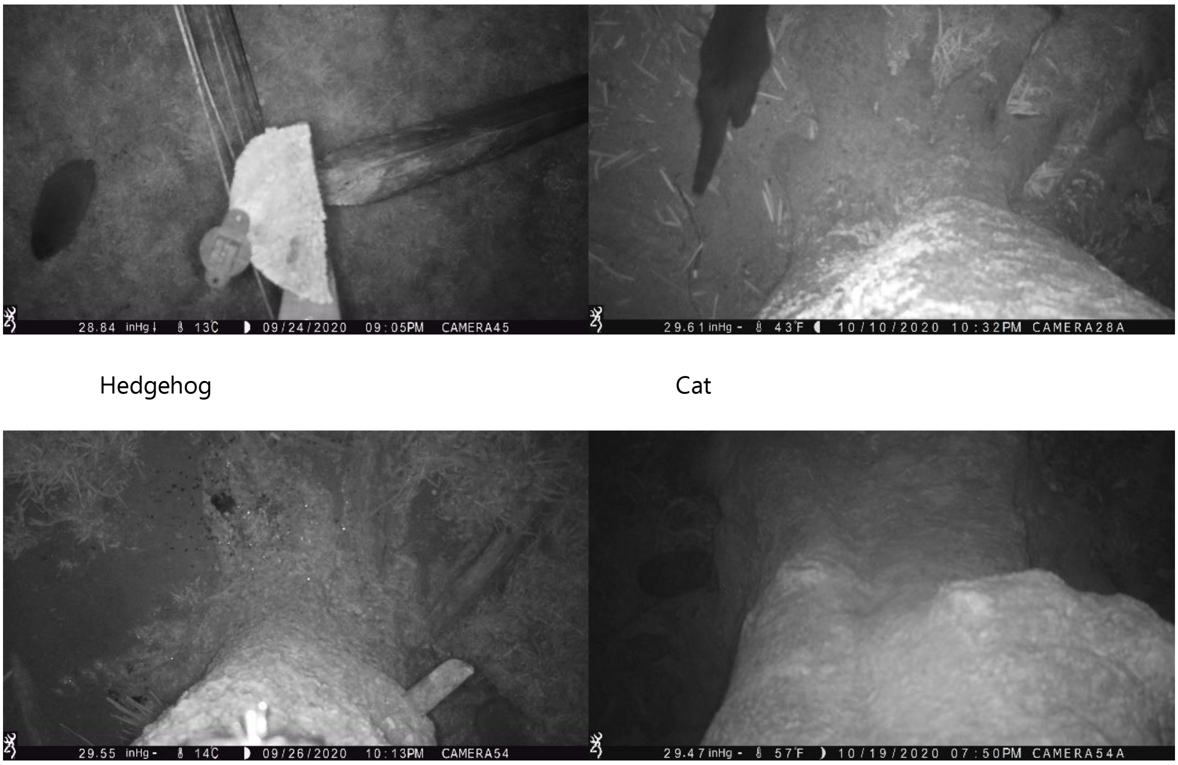


Figure 4. Example of possum image predicted with the lowest confidence.

Figure 5 shows four example images incorrectly labelled as possums with high confidence scores.


28.84 inHg 13C 09/24/2020 09:05PM CAMERA45
29.61 inHg 43F 10/10/2020 10:32PM CAMERA28A
29.55 inHg 14C 09/26/2020 10:13PM CAMERA54
29.47 inHg 57F 10/19/2020 07:50PM CAMERA54A

Hedgehog

Cat

Hedgehog

Lagomorph

Figure 5. Examples of images of other species that were misclassified as possums.

We observed that the incorrectly identified hedgehogs (which dominate the errors) tended to be images where the foreground is bright and the animal and ground is dark. In contrast, the correctly identified hedgehogs with the highest scores were more uniformly bright; Figure 6 shows the two correctly identified hedgehog images with the highest confidence score.

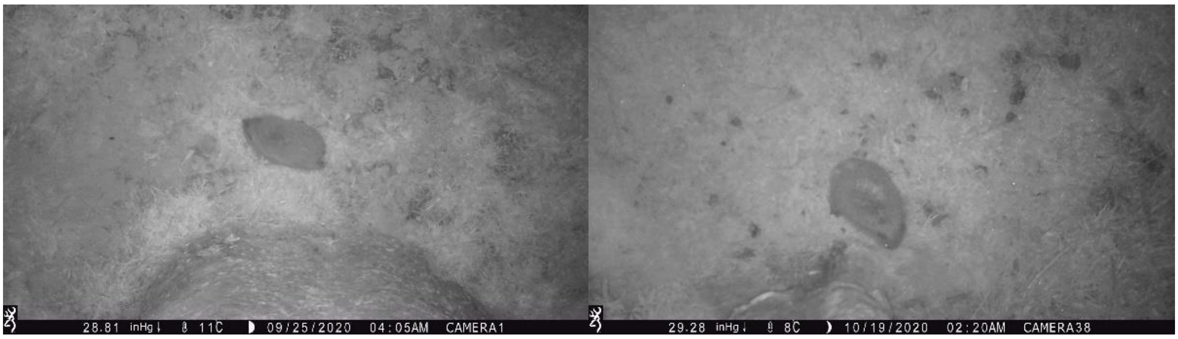

28.81 inHg 11C 09/25/2020 04:05AM CAMERA1
29.28 inHg 8C 10/19/2020 02:20AM CAMERA38

Figure 6. Examples of correctly classified hedgehog images.

The possum images that were misclassified as other species are also all on the ground with one exception; this image was too close to the camera and was misclassified as a kiwi. Figure 7 shows the misclassified possum images with the highest (possum) confidence scores.

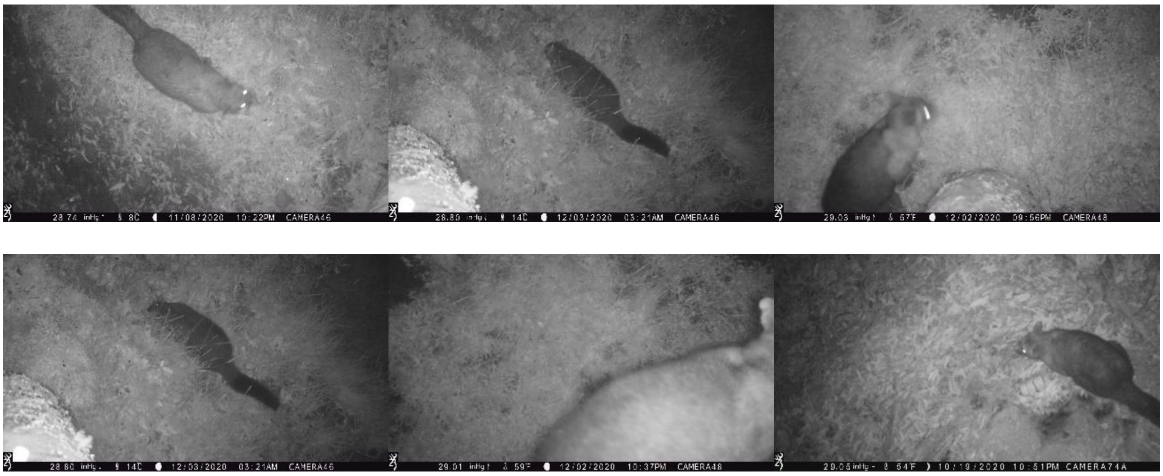

28.74 inHg 8C 11/08/2020 10:22PM CAMERA46
28.80 inHg 14C 12/03/2020 03:21AM CAMERA46
29.03 inHg 67F 12/02/2020 09:56PM CAMERA48
28.80 inHg 14C 12/03/2020 03:21AM CAMERA46
29.01 inHg 59F 12/02/2020 10:37PM CAMERA48
29.05 inHg 64F 10/19/2020 10:51PM CAMERA74A

Figure 7. Possum images misclassified as other species with the highest confidence.

Finally, Figure 8 shows the empty images that were most confidently mis-identified as containing possums (c>0.98). The other classes all had significantly lower confidence scores (i.e. they would have rejected these images), suggesting the possum classifier may be picking up excessive signal from the background.

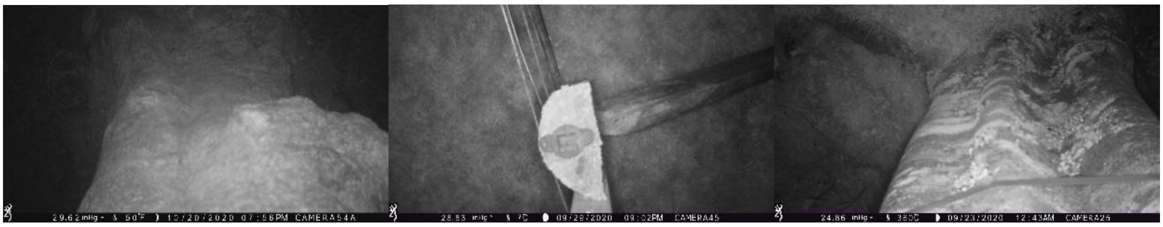

29.62 inHg 50F 10/20/2020 07:56PM CAMERA54A
28.83 inHg 7C 09/29/2020 09:02PM CAMERA45
24.86 inHg 36C 09/23/2020 12:43AM CAMERA25

Figure 8. Empty images most confidently classified as containing possums.

Overall, the model correctly predicts possums where the image quality is high and the possum presents identifying features, such as the face. It has more difficulty when image quality is poor or the animal is only partly visible (at the edge of the image or obscured). However, the model also tends to mis-classify other species and empty images as possums with high confidence, suggesting more careful image selection and cropping is needed.

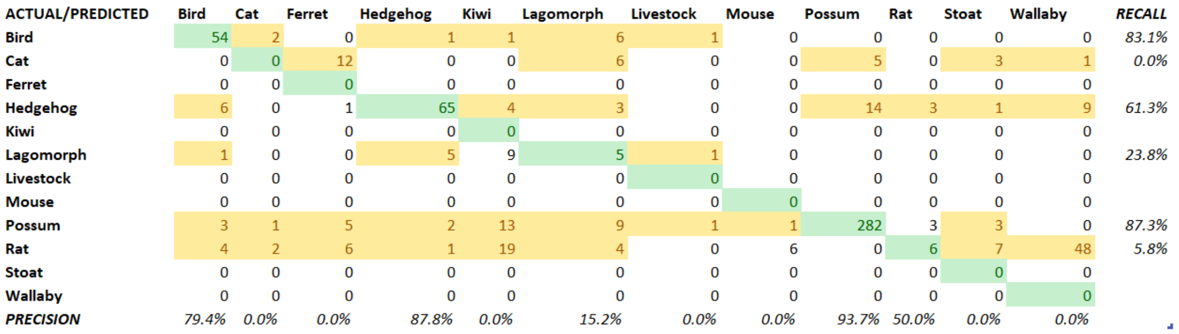
5.4 Fine-tuning the training set

We made changes to the training set to determine whether this might improve the model's performance when detecting possums:

- Remove or re-crop images with large amounts of background
- Remove images that look similar to other animals (particularly hedgehogs)
- Remove images that are impossible to identify manually

We removed the 15 possum crops that met one or more of the above criteria and retrained the network. Table 4 shows the confusion matrix for the refined model.

Table 4. Confusion matrix for the refined model



ACTUAL/PREDICTED	Bird	Cat	Ferret	Hedgehog	Kiwi	Lagomorph	Livestock	Mouse	Possum	Rat	Stoat	Wallaby	*RECALL*
Bird	54	2	0	1	1	6	1	0	0	0	0	0	*83.1%*
Cat	0	0	12	0	0	6	0	0	5	0	3	1	*0.0%*
Ferret	0	0	0	0	0	0	0	0	0	0	0	0	
Hedgehog	6	0	1	65	4	3	0	0	14	3	1	9	*61.3%*
Kiwi	0	0	0	0	0	0	0	0	0	0	0	0	
Lagomorph	1	0	0	5	9	5	1	0	0	0	0	0	*23.8%*
Livestock	0	0	0	0	0	0	0	0	0	0	0	0	
Mouse	0	0	0	0	0	0	0	0	0	0	0	0	
Possum	3	1	5	2	13	9	1	1	282	3	3	0	*87.3%*
Rat	4	2	6	1	19	4	0	6	0	6	7	48	*5.8%*
Stoat	0	0	0	0	0	0	0	0	0	0	0	0	
Wallaby	0	0	0	0	0	0	0	0	0	0	0	0	
PRECISION	*79.4%*	*0.0%*	*0.0%*	*87.8%*	*0.0%*	*15.2%*	*0.0%*	*0.0%*	*93.7%*	*50.0%*	*0.0%*	*0.0%*	

The retrained model returned the same raw overall accuracy of 64% as the previous version. For possums, recall was 87.3%, slightly lower than the previous result of 89.8%. However, precision rose slightly from 93.2% to 93.7%.

Table 5 compares the performance of the two models for possums at various confidence thresholds. Overall performance is roughly the same, with the new model improving precision but with lower recall for low confidence thresholds, and the reverse being true at higher confidence thresholds.

Table 5. Comparison of model precision, recall and false positive rate for possums at various confidence thresholds

	No threshold	C ≥ 0.8	C ≥ 0.9	C ≥ 0.95
First model recall	89.8	87.0	83.6	78.0
First model precision	93.2	93.7	95.7	98.4
First model false positives	15.3	6.7	4.5	2.2
Revised model recall	87.3 (-2.5)	86.3 (-0.7)	83.9 (+0.3)	81.1 (+3.1)
Revised model precision	93.7 (+0.5)	94.3 (+0.6)	95.4 (-0.3)	97.0 (-1.4)
Revise model false positives	12.2 (-3.1)	4.2 (-2.5)	3.2 (-1.3)	2.2 (-)

We anticipated a reduction in the number of hedgehogs being mis-classified as possums having removed a training image without a tail showing, but this did not occur. Instead, all *lagomorphs* mis-classified as possums now classified correctly.

While the refined training set did not have a large influence on mis-classifications, it did reduce the number of false positives, particularly at lower thresholds. At the original threshold of confidence ≥ 0.8, false positives reduce from 6.7% of the empty images to 4.2%. The new model therefore appears to give marginally improved performance, and was retained.

5.5 Detection of other species of interest

A secondary goal was to try to detect other species of interest. In this study we have only considered species already present in the predictive model; of these the following potential target species were present in the vertical imagery: birds, cats, hedgehogs, lagomorphs, possums, and rats. Table 6 summarises performance before (and after) thresholding.

Table 6. Comparison of precision and recall of the original (horizontal) model and refined vertical image model for other target species

	Bird (c≥0.67) 57 images	Cat (c≥0.95) 16 images	Hedgehog (c≥0.8) 95 images	Lagomorph (c≥0.98) 10 images	Rat (≥0) 91 images
Original model recall	71 (70)	64 (64)	85 (85)	75 (58)	86 (86)
Original model precision	71 (71)	91 (97)	97 (97)	73 (81)	97 (97)
New model recall	83.1 (76.9)	0 (0)	61.3 (57.7)	23.8 (14.3)	5.8 (5.8)
New model precision	79.4 (84.7)	0 (0)	87.8 (89.6)	15.2 (33)	50 (50)

Performance of the model on vertical cameras is generally poorer than that originally achieved for horizontal images, with the exception of birds. This is likely to be because the small training set covers only a small number of camera installations, reducing the variability of scene and lighting conditions. This is particularly true for cats and lagomorphs, which had very few training images. We also note some particular features influencing performance for the various species:

1 For the bird class, there is considerably less variation in the vertical images than in the horizontal images, which range from small birds with grass backdrops to wekas in forests and keas in alpine rock settings. This likely led to higher performance.

2 The model was particularly poor for cats, with the revised model failing to recognize any. The most likely explanation for this is that the cats in the training set were almost all dark or tabby, with just two lighter-coloured cats climbing toward the camera. In the test set, most of the cats appear light coloured. The training set was also small – just 16 images used during training.

3 Similarly for hedgehogs, the training images generally have light backgrounds, whereas a significant number of the images in the validation set have darker backgrounds, and the subject is poorly lit.

4 The rat class performed particularly poorly despite the relatively large number of images. This may be due to low variability in the background and lighting conditions, which failed to match the validation set. Also, there were a significant number of images where the subject was difficult, and sometimes impossible, to locate in the image. Finally, some of the images labelled rat may in fact be mice, which the original model also had difficulty identifying.

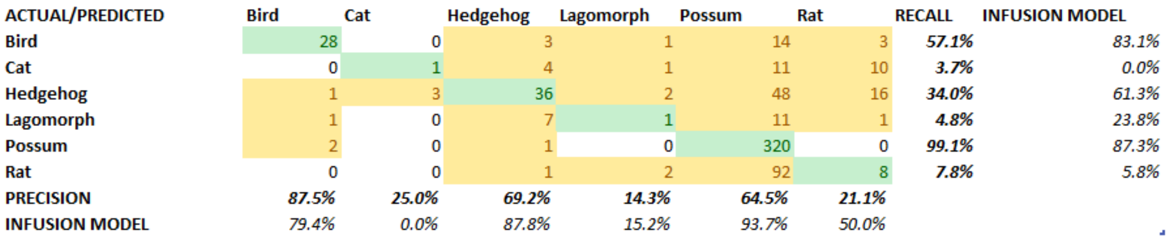
5 The lagomorph validation images contained 8 (out of 20) images where the subject was very poorly lit and difficult to see.

It is important to note that variation in camera setup during training is critical. For each image, a crop of an empty portion of the image is taken in addition to that of the animal itself; this is to teach the model to ignore the background. However, when a novel background is encountered, some extraneous features may match one of the animal classes more strongly than the animal itself.

5.6 Infusion versus separate models

We began by assuming that retraining our existing model by infusing its training set (of horizontal images) with vertical images would give superior performance to a model trained solely from the available vertical images. We tested this assumption by training a separate model using only vertical imagery and testing it on the same test set. Table 7 shows the confusion matrix for this model.

Table 7. Confusion matrix for model trained on vertical images only



ACTUAL/PREDICTED	Bird	Cat	Hedgehog	Lagomorph	Possum	Rat	RECALL	INFUSION MODEL
Bird	28	0	3	1	14	3	***57.1%***	*83.1%*
Cat	0	1	4	1	11	10	***3.7%***	*0.0%*
Hedgehog	1	3	36	2	48	16	***34.0%***	*61.3%*
Lagomorph	1	0	7	1	11	1	***4.8%***	*23.8%*
Possum	2	0	1	0	320	0	***99.1%***	*87.3%*
Rat	0	0	1	2	92	8	***7.8%***	*5.8%*
PRECISION	***87.5%***	***25.0%***	***69.2%***	***14.3%***	***64.5%***	***21.1%***		
INFUSION MODEL	*79.4%*	*0.0%*	*87.8%*	*15.2%*	*93.7%*	*50.0%*		

Recall improved dramatically for possums, but at a cost of reduced precision. Conversely, recall fell for birds, hedgehogs, and lagomorphs, with precision *increasing* for birds, falling for hedgehogs, and staying the same for lagomorphs.

This is a completely new model, so the original confidence thresholds are unlikely to apply. We therefore selected new thresholds using the same technique, by inspecting the predictions in descending order of confidence for each species, and selecting the threshold where precision noticeably degrades. Table 8 lists the results.

Table 8. Comparison of performance between the previous (infusion) model and the new model trained with vertical imagery only

Species	Threshold	Precision	Precision Infusion model	Recall	Recall Infusion model	False positive rate
Bird	0.9	92.0	79.4	46.9	83.1	2.8%
Cat	0	25.0	0.0	3.7	0.0	0.7%
Hedgehog	0.915	87.5	87.8	25.4	61.3	0.5%
Lagomorph	0.6	25	15.2	4.8	23.8	0.5%
Possum	0.93	92.0	94.3	72.8	86.3	20.4%
Rat	N/A	21.1	50	7.8	5.8	14.8%

Even after thresholding, the model trained on vertical images only is inferior to the model trained using horizontal images with infused images. For the possum detector, recall falls from 86% to 73%, and the false positive rate rises from 4.2% to 20.4%. This supports our assumption that better performance will be obtained by combining horizontal and vertical training images.

5.7 Further testing

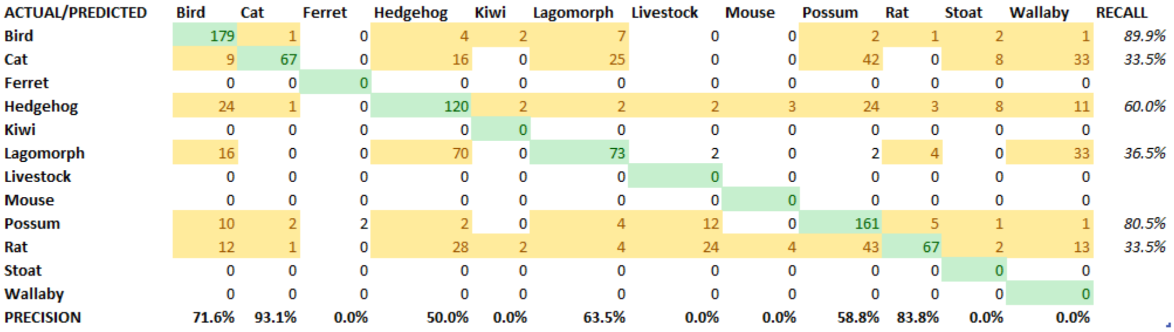
During their development, the models were tested on a single small set of images. To gain an appreciation of the model's generality, we sorted the remaining vertical images by species, giving 1,508 new test images for the target species:

- Bird: 202
- Cat: 24
- Hedgehog: 272
- Lagomorph: 114
- Possum: 203
- Rat: 693

Testing overall accuracy for all target species using this set is problematic because of the large bias towards rats, meaning the accuracy of this species dominates the outcome. We therefore balanced the sample by taking 200 samples (with replacement) of each class: for all species with more than 200 images, the extra images were randomly deleted, while for the species with insufficient images, the images were replicated in strict order (by file name) until the set contained sufficient images.

The model achieved an overall raw accuracy on the balanced sample of 55.6%, which is somewhat lower than previously. Table 9 shows the confusion matrix.

Table 9. Confusion matrix for the vertical model when tested on the new test set



ACTUAL/PREDICTED	Bird	Cat	Ferret	Hedgehog	Kiwi	Lagomorph	Livestock	Mouse	Possum	Rat	Stoat	Wallaby	RECALL
Bird	179	1	0	4	2	7	0	0	2	1	2	1	*89.9%*
Cat	9	67	0	16	0	25	0	0	42	0	8	33	*33.5%*
Ferret	0	0	0	0	0	0	0	0	0	0	0	0	
Hedgehog	24	1	0	120	2	2	2	3	24	3	8	11	*60.0%*
Kiwi	0	0	0	0	0	0	0	0	0	0	0	0	
Lagomorph	16	0	0	70	0	73	2	0	2	4	0	33	*36.5%*
Livestock	0	0	0	0	0	0	0	0	0	0	0	0	
Mouse	0	0	0	0	0	0	0	0	0	0	0	0	
Possum	10	2	2	2	0	4	12	0	161	5	1	1	*80.5%*
Rat	12	1	0	28	2	4	24	4	43	67	2	13	*33.5%*
Stoat	0	0	0	0	0	0	0	0	0	0	0	0	
Wallaby	0	0	0	0	0	0	0	0	0	0	0	0	
PRECISION	71.6%	93.1%	0.0%	50.0%	0.0%	63.5%	0.0%	0.0%	58.8%	83.8%	0.0%	0.0%	

For this new test set the possum class produced a raw recall of 80.5% with a fairly low precision of 58.8%. Table 10 shows the final model's precision and recall on the new balanced test set before and after thresholding.

Table 10. Comparison of vertical and horizontal model performance on the larger test set

	Bird (c≥0.67)	Cat (c≥0.95)	Hedgehog (c≥0.8)	Lagomorph (c≥0.98)	Possum (c≥0.8)	Rat (c≥0.75)
Recall (no threshold)	89.9	33.5	60	36.5	**80.5**	33.5
Precision (no threshold)	71.6	93.1	50	63.5	**58.8**	83.8
Recall (thresholded)	77.4	29	51.5	20	**78.0**	22.5
Precision (thresholded)	73.3	100	56	80	**63.4**	90
Horizontal model recall (thresholded)	70	64	85	58	**65**	86
Horizontal model precision (thresholded)	71	97	97	81	**90**	97
False positive rate	1.6%	0.4%	0.8%	0.8%	**4.2%**	5.5%

While precision and recall for possum detection is somewhat lower than previously measured, performance improved for some other species, notably cats and rats. The original threshold for rats was 0, i.e. the original rat predictor did not benefit from thresholding the confidence because its precision was close to 100%. For the new model, precision fell somewhat, so we selected a new threshold based on the point where precision deteriorated markedly (c≥0.75). Note also that a significant number of the (training and test) images labelled as rats were likely mice, which the original model found extremely difficult to identify. Finally, the difference in performance between this trial and the previous smaller ones is likely due in large part to the balancing of the images per species. In particular, the model often confuses possums and cats, and the original test set had a much smaller number of cat images.

Based on this test set, classification performance of the model on vertical imagery is generally poorer than for horizontal imagery, with the exception being birds. For the main target class of possums, performance is somewhat similar.

5.8 Non-target classes

Finally, we tested the model on non-target classes (livestock, pig) to measure the false positive rate for these species. The model contains a classifier for livestock, but it has been trained on vertical images only, and there is none for pigs, so the latter will all be misclassified as some other species, potentially including the target species. For the vertical images, 'livestock' is a broad class including cattle, sheep, and goats.

The model performed poorly for both these classes, with 34% of livestock and 71% of pigs being misclassified as possums, reflecting that, of the images previously seen by the model during training, the vertical livestock and pigs looked more like possums than any other species. This highlights the need to ensure the model is adequately trained on images for all commonly observed species.

6 Conclusions

We conducted a small-scale trial to assess the ability of a deep learning image recognition model to accurately detect possums, and potentially other species, in vertical camera trap imagery, and concluded that an existing model trained using horizontal images could be fine-tuned to also work for vertically oriented cameras. The model may have utility in substantially reducing the number of images requiring manual inspection (especially false positives), but would require further training before being used as an automated detector.

The image set for this research was fairly small, being dominated by empty images, and images of non-target species (livestock). Given the limited size of the training set, we would expect that results could be considerably improved through further training, both by increasing the number and variety of training images, and by careful selection of the images used for training to maximise variability.

7 Recommendations

7.1 Further testing

We have developed a model for automatically detecting possums and other species of interest in camera trap images based on a moderately large set of images. The images were split into train and test folders based on file name; this should reduce the overlap between the two sets of images taken with the same camera setup and, potentially, at close time intervals. However, it is not infallible. To properly assess the model's performance, it should be tested on further, completely independent sets of test images taken at a wide variety of locations.

Recommendation 1: Further test the model using independent imagery.

7.2 Further tuning

During the course of this research, we experimented with a large number of setups and tuning parameters. However, these are all interdependent and so there are many more combinations that could be tried. Of these, image *augmentation* offers many potentially useful options. Augmentation is used to perturb each image randomly each time it is presented for training, to increase the robustness of the model by reducing the likelihood that the model is relying on highly specific aspects of the training images. We used a moderately aggressive augmentation scheme that randomly zoomed and cropped the images, as well as flipping them horizontally, but there are other options that could also be tried, such as varying the colour, brightness, and contrast, and rotating the images.

Recommendation 2: Explore further augmentation to see if it improves the robustness of the model.

7.3 Larger training set

The image set used to train the model was limited in size and contained a fairly small number of camera setups. Performance of the model would be expected to improve with a larger training set and careful selection of the training instances to maximise variability. We also noted that the images provided were all taken with the same type of camera; if images taken with other camera types are also anticipated, the model should be trained on images from a variety of camera brands and types.

Recommendation 3: Obtain more training images and retrain the model on a more comprehensive training set.

8 References

Anton V, Hartley S, Geldenhuis A, Wittmer HU 2018. Monitoring the mammalian fauna of urban areas using remote cameras and citizen science. Journal of Urban Ecology 4(1): doi: 10.1093/jue/juy002.

Curran B, Nekooei SM, Chen G 2022. Accurate New Zealand wildlife image classification – deep learning approach. In: Long G, Yu X, Wang S eds AI 2021: Advances in artificial intelligence. Cham: Springer International Publishing. Pp. 632–644.

Shahinfar S, Falzon G 2021. Assessing the performance of a convolutional neural network for cloud computing based recognition of New Zealand fauna in camera trap images. Armidale: University of New England.

Tabak MA, Falbel D, Hamzeh T, Brook RK, Goolsby JA, Zoromski LD, Boughton RK, Snow NP, Vercauteren KC, Miller RS 2022. CameraTrapDetectoR: automatically detect, classify, and count animals in camera trap images using artificial intelligence. bioRxiv https://doi.org/10.1101/2022.02.07.479461.