Machine Learning for Oregon Wildlife Camera Trap Images

Presenter: Charles Ison
Faculty Advisor: Yue Zhang

College of Engineering - Electrical Engineering and Computer Science

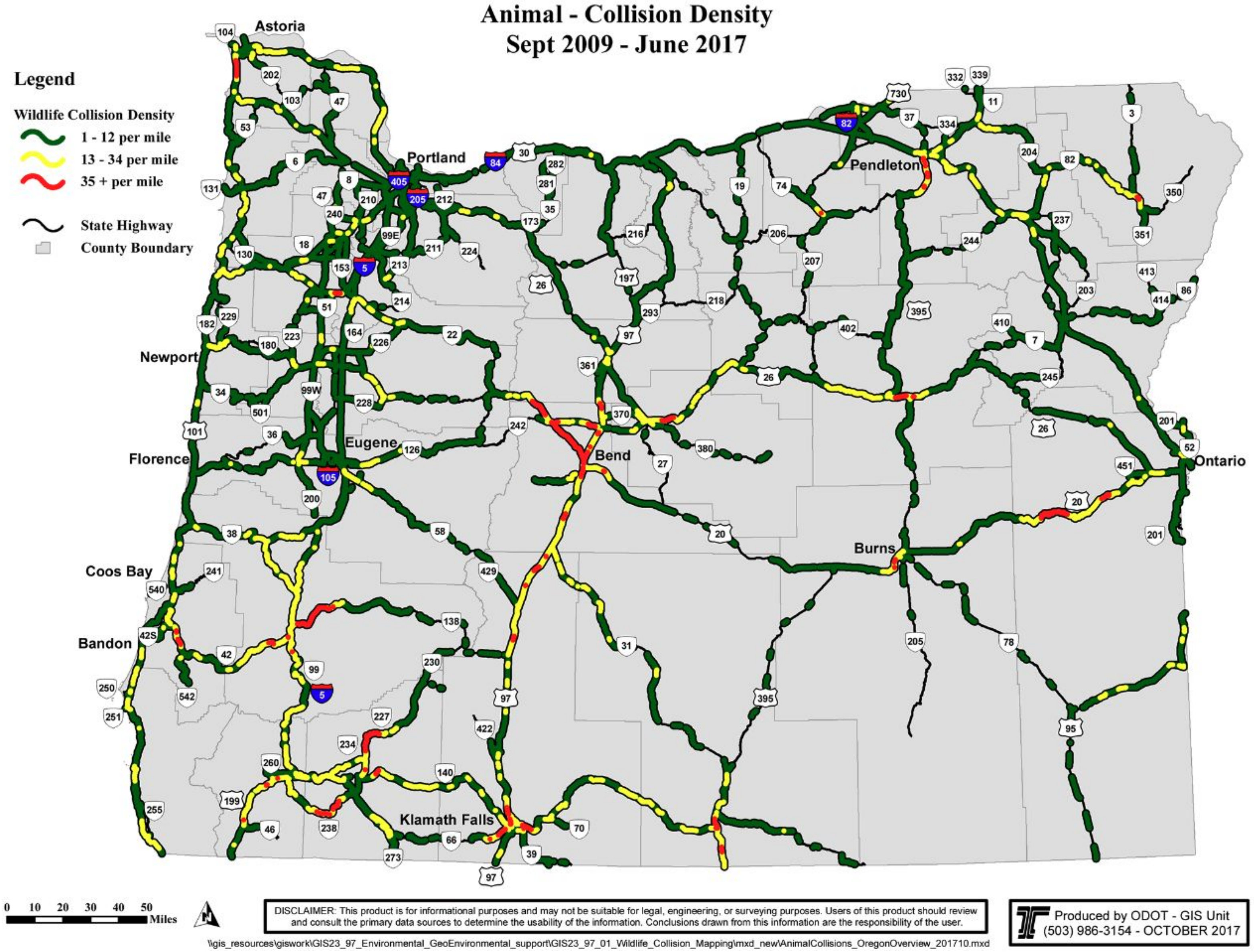
Oregon State University

Background

- A 2008 report to Congress estimated there are between 1 and 2 million collisions every year between vehicles and large animals in the United States[1]
- These accidents annually caused 26,000 human injuries, 200 human deaths, and cost $8.39 billion in monetary losses[1]
- Since these reports, road networks have continued to further expand, increasing the number of intersections between wildlife migration routes and major highways[2]

1. Huijser, M., Mcgowen, Patrick., J. F., A. H., A. K., A. C., D. S., AND Ament, R., (2008). *Wildlife-Vehicle Collision Reduction Study: Report to Congress*. Retrieved from https://www.fhwa.dot.gov/publications/research/safety/08034/08034.pdf
2. Coe, P.K., Nielson, R.M., Jackson, D.H., Cupples, J.B., Seidel, N.E., Johnson, B.K., Gregory, S.C., Bjornstrom, G.A., Larkins, A.N. and Speten, D.A. (2015), Identifying migration corridors of mule deer threatened by highway development. Wildl. Soc. Bull., 39: 256-267. https://doi.org/10.1002/wsb.544

Background - Oregon



Background - Mitigation Strategies

- Previously attempted mitigation strategies include[1]:
  - Wildlife fencing
  - Wildlife bridges
  - Wildlife underpasses
  - Wildlife crossing guards
  - Increased wildlife signage
  - Increased roadway lighting
  - Roadside animal detection systems

1. Huijser, M., Mcgowen, Patrick., J. F., A. H., A. K., A. C., D. S., AND Ament, R., (2008). *Wildlife-Vehicle Collision Reduction Study: Report to Congress*. Retrieved from https://www.fhwa.dot.gov/publications/research/safety/08034/08034.pdf

How can machine learning be leveraged to improve outcomes?

Opportunities for Machine Learning

- Can we use machine learning with camera trap image data to detect the presence of wildlife in real time?
  - Creates opportunities for systems that take action based on the detection of wildlife to try and improve outcomes
- Can we use machine learning with camera trap image data to count wildlife and build larger more comprehensive datasets
  - This information can be leveraged to make more informed mitigation decisions

Starting Point - Wildlife Detection

- Wildlife detection - present or not present?
- Most basic machine learning functionality required for a wildlife sensing and responding system
- Also lays the groundwork for wildlife counting
  - Counting is a much harder task, so without effective detection it is unlikely we will see good results for counting (more details later)

Wildlife Detection - Dataset

- Collected between May and September of 2022 by researchers at OSU Cascades
- Reconyx cameras were placed along wildlife underpasses location near US S. Highway 97 in central Oregon
- Three cameras were arranged at each underpass
    - One facing eastwards
    - One facing westwards
    - One decoy camera
- 3,995 images were manually labeled as wildlife present or no wildlife present by students in our research group

Wildlife Detection - Lighting Difficulties

Wildlife Detection - Weather Difficulties

Wildlife Counting - Human Difficulties

- Not only wildlife uses the underpasses!

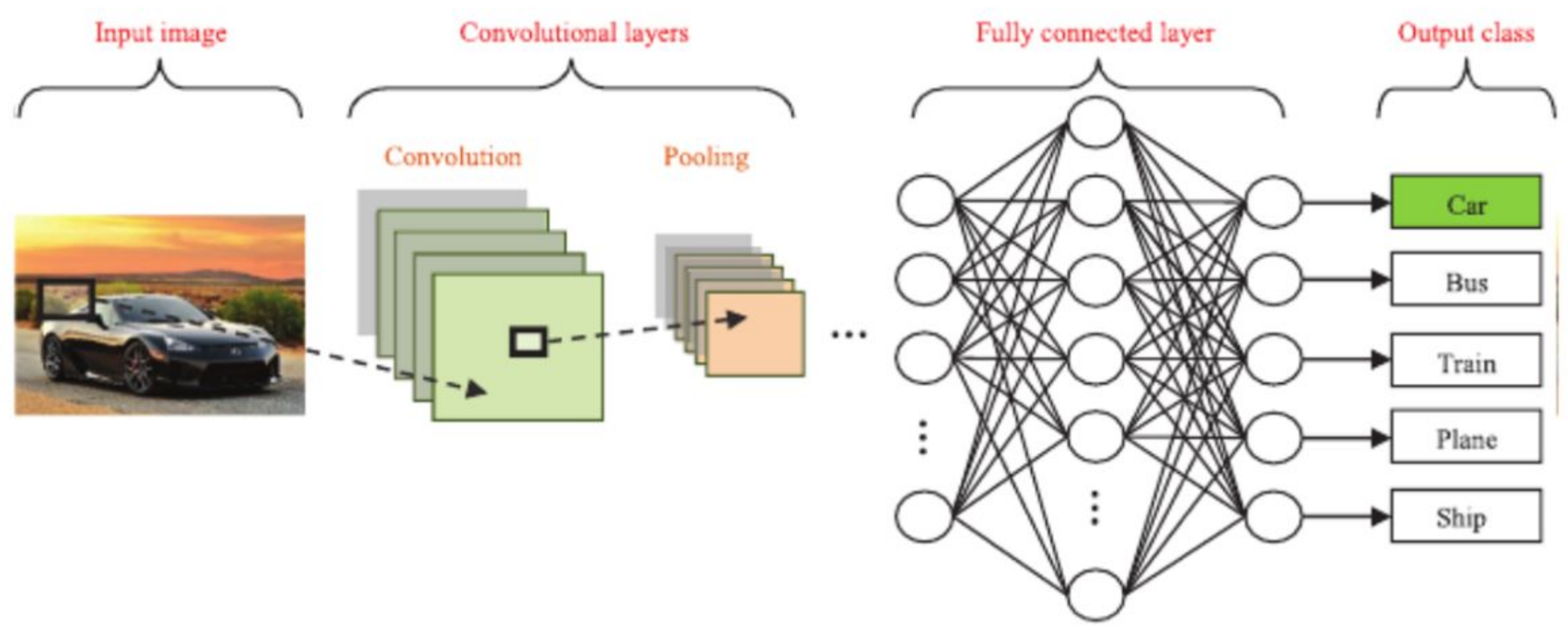
Wildlife Detection - Models

- We evaluated the performance of 9 separate neural networks for wildlife detection
  - Within these 9 models there are 2 main model categories:
    - Convolutional Neural Networks (CNNs)
    - Vision Transformers (ViTs)
- These two categories were chosen for previously demonstrating state of the art performance on computer vision tasks

Wildlife Detection - Convolutional Neural Networks (CNNs)

- CNNs have achieved state of the art performance on computer vision tasks for the past decade
- Well established in the literature
- Models ranged from simple and small to complex and large
- CNNs can be described as learning to recognize patterns and textures from images
- We tested 5 CNN variants of varying sizes:
  - AlexNet (oldest CNN tested)
  - VGG-11 (newer, smaller CNN)
  - VGG-19 (newer, larger CNN)
  - ResNet-50 (newest, smaller CNN)
  - ResNet-152 (newest, larger CNN)

Wildlife Detection - Convolutional Neural Networks (CNNs)

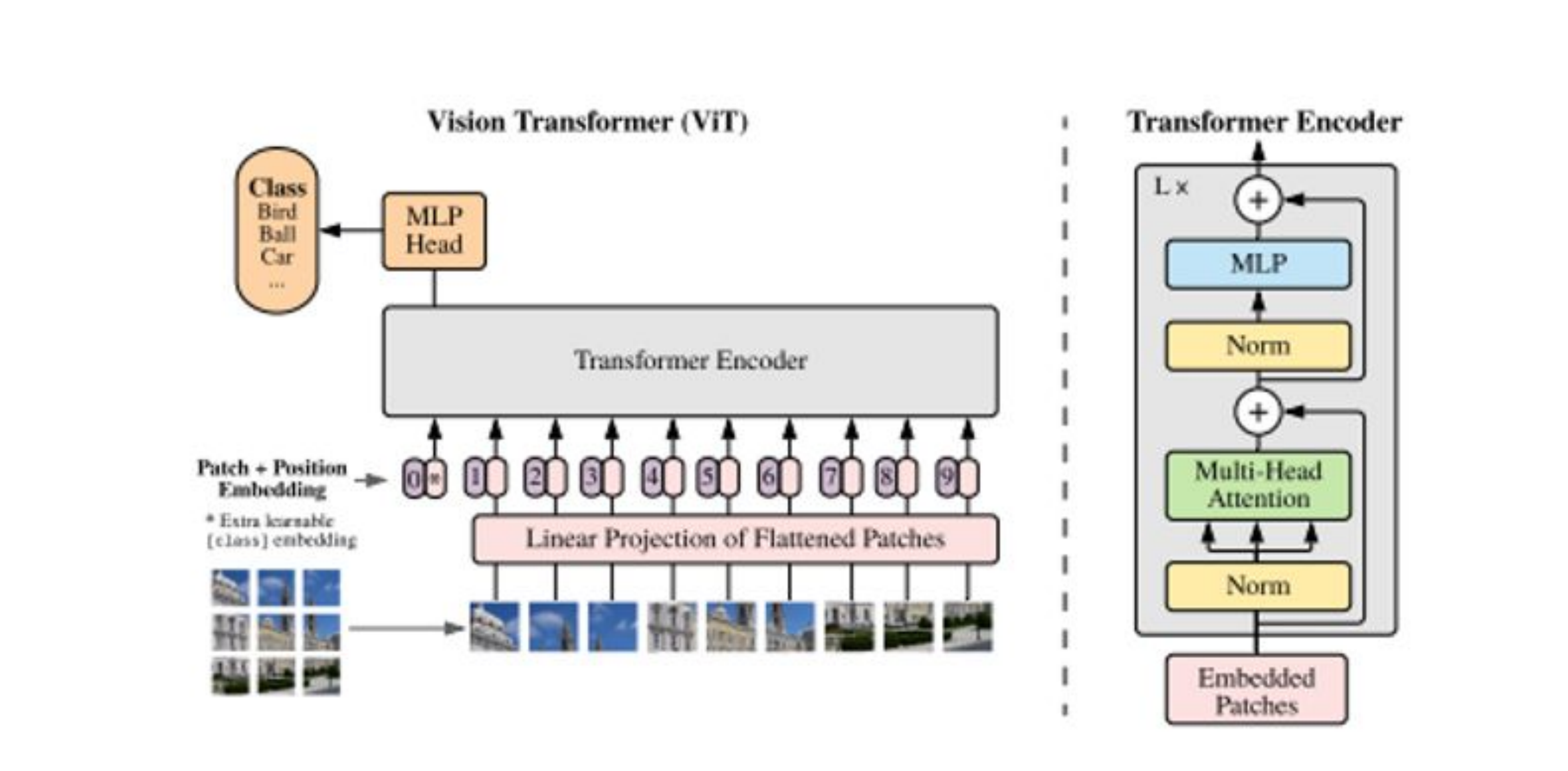


Rawat, W., & Wang, Z. (09 2017). Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review. *Neural Computation*, *29*(9), 2352–2449. doi:10.1162/neco_a_00990

Wildlife Detection - Vision Transformers (ViTs)

- ViTs are a newer model that has recently surpassed the performance of CNNs for specific computer vision tasks
- Typically requires a larger amount of image data for training than CNNs
- More computationally expensive and slower than CNNs
- We tested 4 ViT variants with varying size dimensions:
    - ViT-B-16
    - ViT-B-32
    - ViT-L-16
    - ViT-L-32

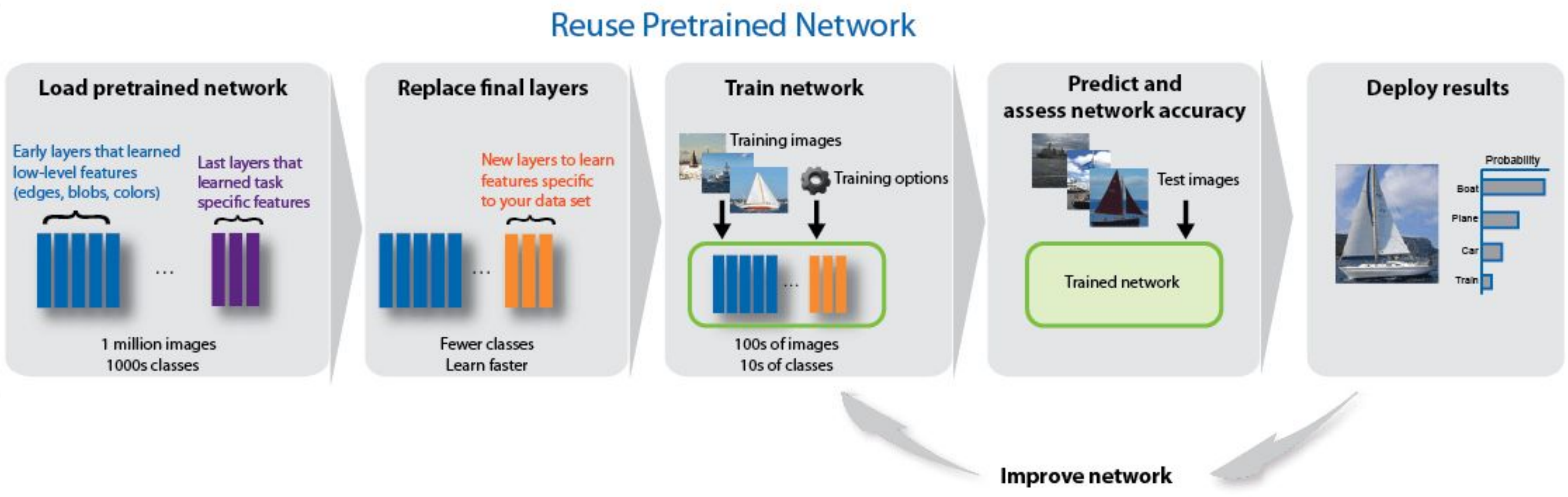
Wildlife Detection - Vision Transformers (ViTs)



Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., … Houlsby, N. (2020). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *CoRR*, *abs/2010.11929*. Retrieved from https://arxiv.org/abs/2010.11929

Wildlife Detection - Models

- Each model of both possible types takes a single image as input and then produces a classification of wildlife present or not present
- All model were implemented in direct alignment with their originally proposed architectures
  - No custom changes besides updating the final layer to produce the required two option output
- Pretrained weights were used for each model and then fine-tuned for wildlife detection



https://www.mathworks.com/help/deeplearning/ug/transfer-learning-using-pretrained-network.html

Wildlife Detection - Evaluation Metrics

$$Accuracy = \frac{True\ Positive + True\ Negative}{True\ Positive + True\ Negative + False\ Positive + False\ Negative}$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

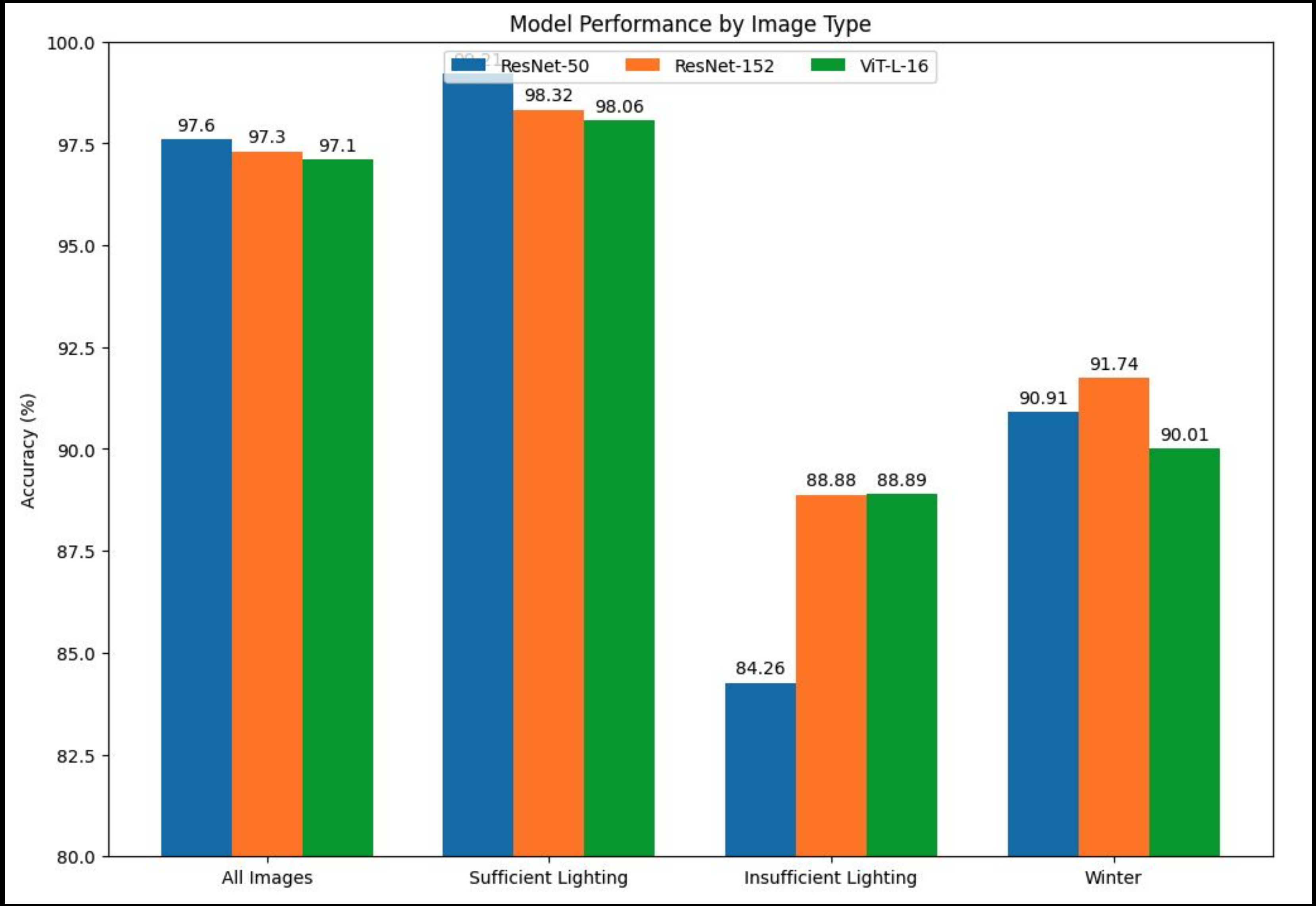
$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

*Recall is how often the model is right when wildlife is actually present and precision is how often the model is right when the model predicts wildlife is present

Wildlife Detection - Results

	Accuracy	Precision	Recall	F-score
VGG-11	96.50%	89.96%	97.29%	93.48%
VGG-19	96.50%	97.85%	88.37%	92.87%
ResNet-50	97.60%	96.06%	94.57%	95.31%
ResNet-152	97.30%	97.93%	91.47%	94.59%
AlexNet	95.40%	87.32%	96.12%	91.51%
ViT-B-16	96.10%	87.01%	97.36%	91.89%
ViT-B-32	94.89%	83.08%	97.36%	89.66%
ViT-L-16	97.10%	91.95%	95.59%	93.74%
ViT-L-32	92.09%	93.02%	70.48%	80.20%

Wildlife Detection - Impact of Lighting and Snow



Next Steps - Wildlife Counting

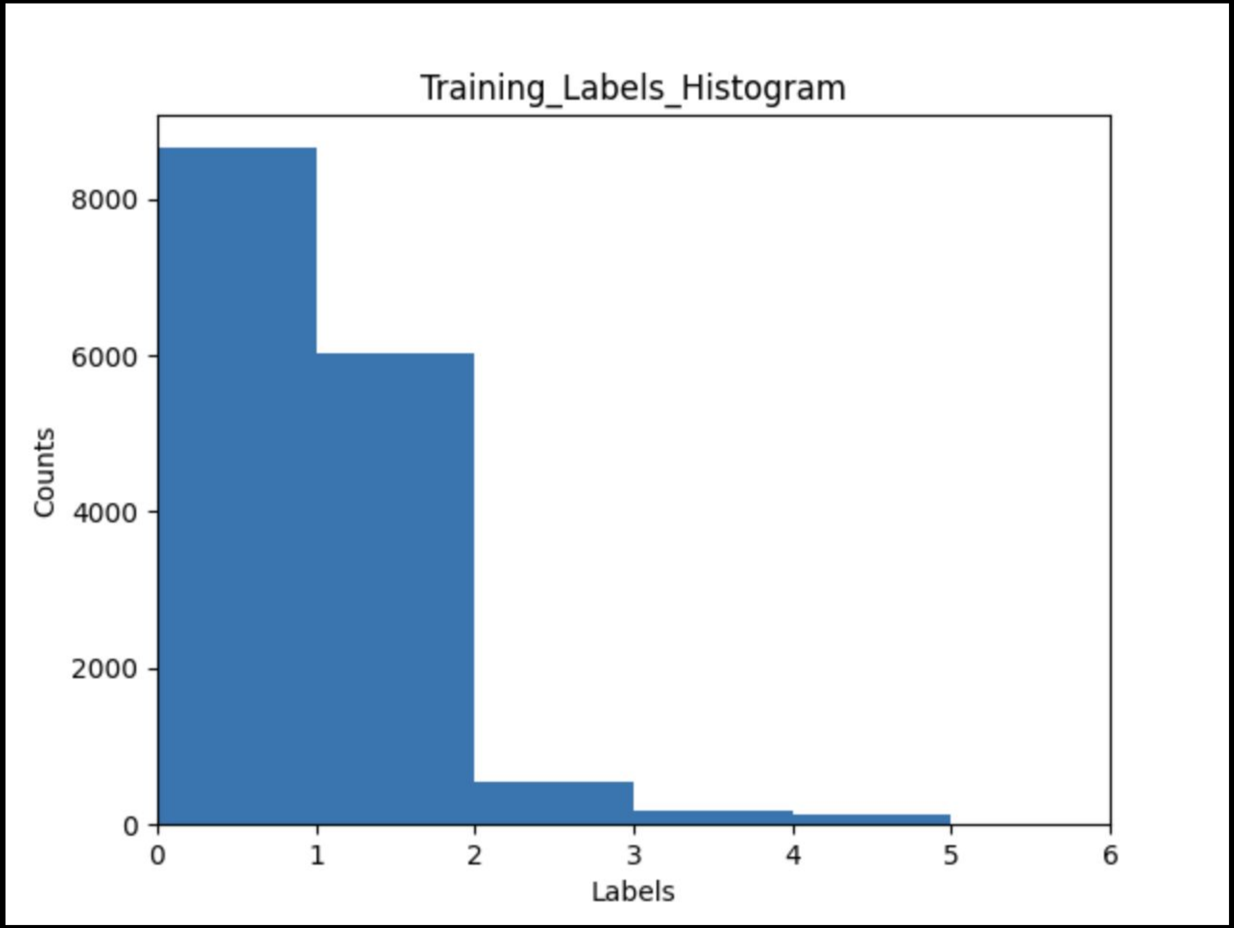
- Our wildlife detection results show the models are able to successfully detect the presence of wildlife
- A natural next question is if the models can be updated to count the wildlife present in an image?
- This would open the door for using the models to analyze and count wildlife in large collections of image data
  - These results could then be used to drive more informed mitigation strategies

Wildlife Counting - Difficulties

- It might seem like a trivial increase in complexity, but wildlife counting is a significantly harder problem than wildlife detection
- Frequently the wildlife is clumped together in a group so that is hard for the model to disentangle the animals and count them accordingly
- Often there is a batch of images taken for each crossing event where multiple animals are photographed multiple times
  - We do not want to double or triple count in these scenarios and need to come up with a mechanism to avoid overcounting

Wildlife Counting - Difficulties Continued

- There is substantial data imbalance



Wildlife Counting - Difficulties Continued

- Many images in the training datasets are very similar
  - This makes it possible for the models to memorize specific training images rather than the patterns and textures required for generalizing

Wildlife Counting - Data

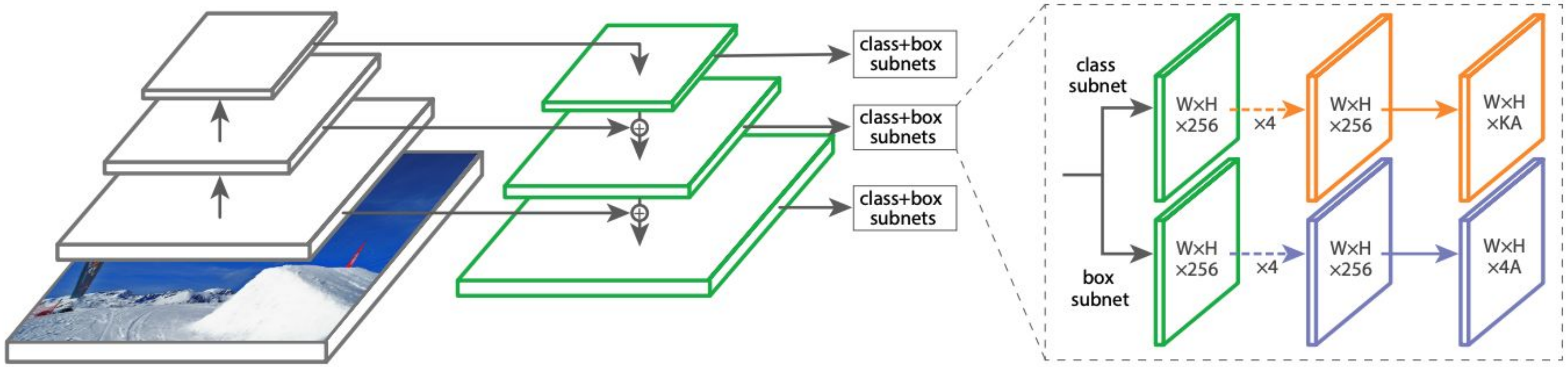
- Because of these difficulties, we moved away from simply assigning each image a numerical label indicating the total number of wildlife present
  - For example: 1, 2, 3, 4, . . .
- Instead each image with wildlife present is labeled using a bounding box that indicates the location of the wildlife in the image
  - Is more time intensive to label
- 19,767 images were labeled with bounding boxes by students in our research group
- This enables us to use a new category of neural network models called object detection models which specifically focus on the location of objects in an image

Wildlife Counting - Models

- Similar to previously, we evaluated the performance of CNNs and ViTs for wildlife counting
- We also tested a different class of neural networks call object detection models:
  - Faster R-CNN
  - Single Shot MultiBox Detector
  - RetinaNet

Wildlife Counting - Object Detection Models

- Object detection models leverage CNNs to detect where objects of interest exist within an image and then predict what type of object exists in that location
- Object detection models are also well established in the literature and widely used for computer vision tasks that need to be location aware
- Similar to previously mentioned wildlife detection work:
  - All model were implemented in direct alignment with their originally proposed architectures
  - Pretrained weights were used for each model



(a) ResNet (b) feature pyramid net (c) class subnet (top) (d) box subnet (bottom)

Lin, Tsung-Yi & Goyal, Priyal & Girshick, Ross & He, Kaiming & Dollar, Piotr. (2018). Focal Loss for Dense Object Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence. PP. 1-1. 10.1109/TPAMI.2018.2858826.

Wildlife Counting - Batching Strategy

- Each model produces a single prediction for every image that is passed as input
- A strategy is required to consolidate many images and corresponding predictions for a single crossing event
  - There needs to be a single count prediction that accurately reflects the number of animals present

Wildlife Counting - Batching Strategy

- We propose assigning a batch of photos to a single crossing event if the motion sensor trigger time is no more than 60 seconds apart for each photo
- We then take the maximum prediction from each batch as the final prediction
- Intuition is that all wildlife present will most likely be captured in a single image at least once

Wildlife Counting - Results Metrics

- Wildlife counting is a regression task instead of a classification task like detection
  - Because of this we need to use different metrics to evaluate model performance

$$\text{Mean Absolute Error (MAE)} = \frac{1}{n}\sum |\hat{y} - y|$$

$$\mathrm{R}^2 = 1 - \frac{\mathrm{RSS}}{\mathrm{TSS}}$$

$$\mathrm{RSS} = \text{sum of squares of residuals}$$

$$\mathrm{TSS} = \text{total sum of squares}$$

- R^2 can be described as the proportion of the variation in the dependent variable that is predictable from the independent variable
  - Values close to 1.0 indicate the model is performing well

Wildlife Counting - Validation Results

	ResNet34	ResNet50	ResNet152	RetinaNet
Validation Batch MAE	0.104	0.109	0.099	0.087
Validation Batch Accuracy	0.942	0.945	0.943	0.924
Validation Batch R-Squared	0.843	0.855	0.855	0.756

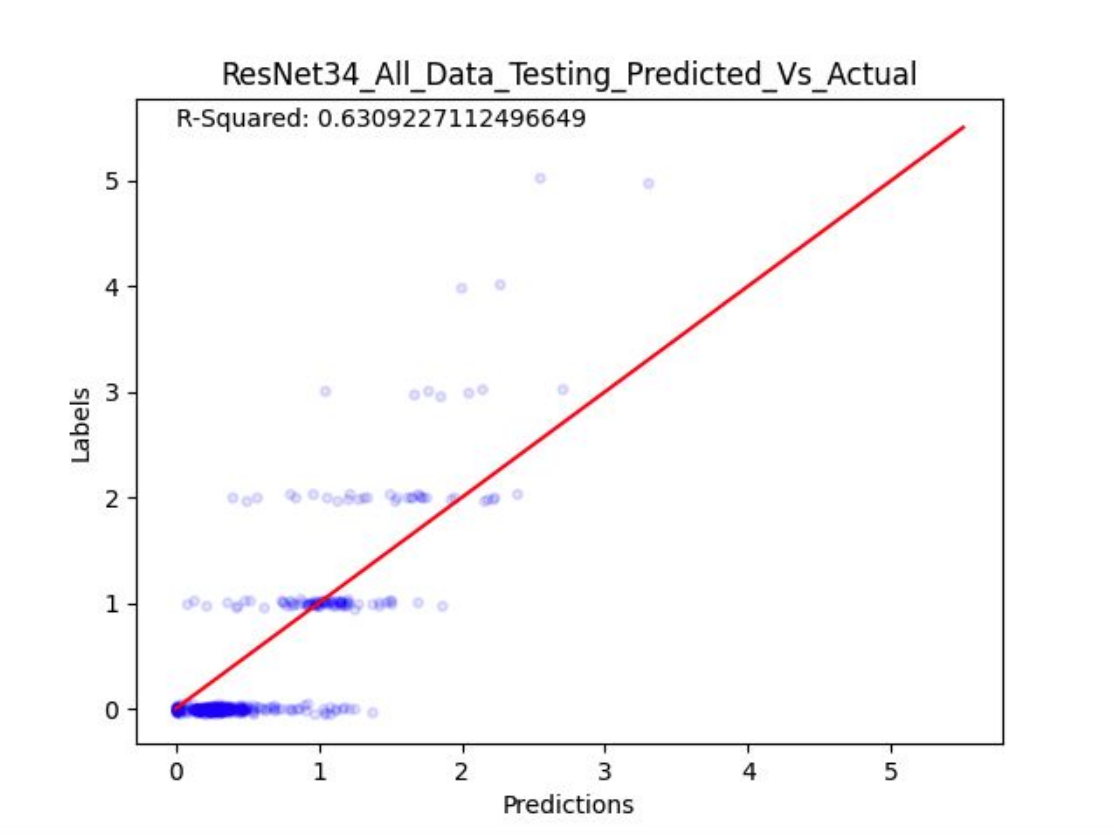
Wildlife Counting - Testing Results

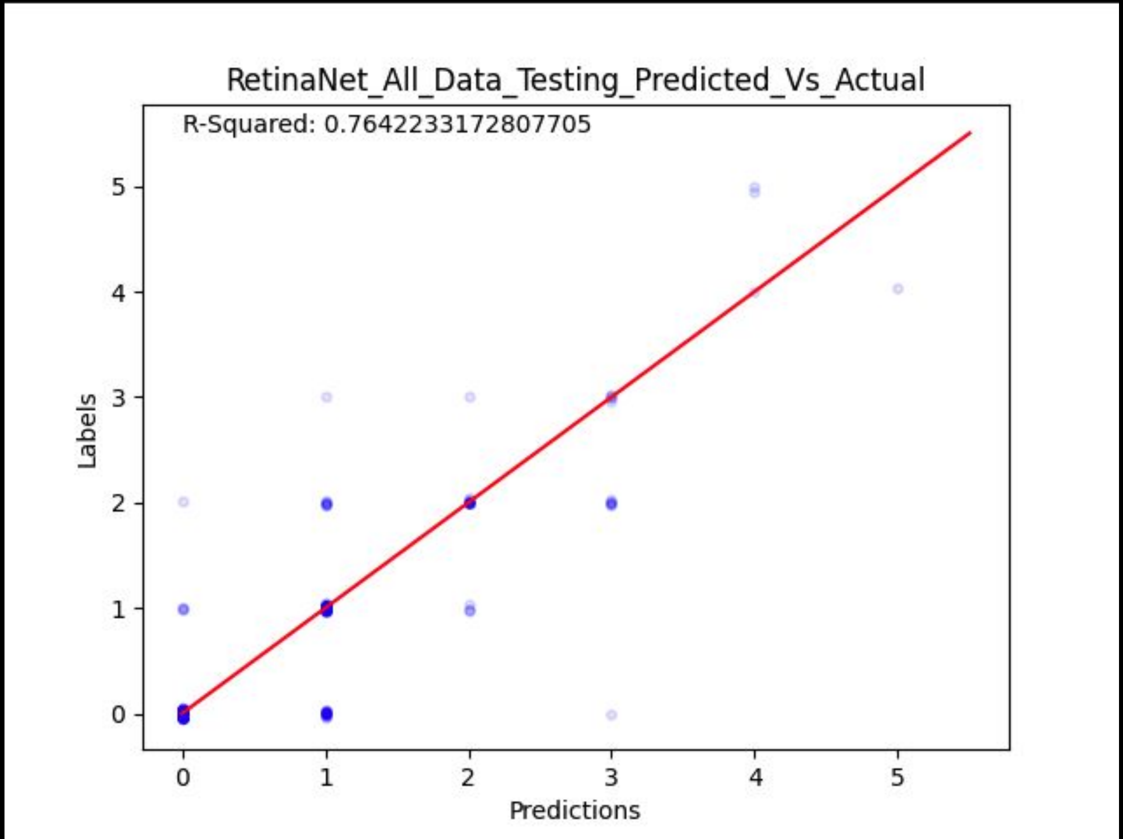
	ResNet34	ResNet50	ResNet152	RetinaNet
Testing Batch MAE	0.354	0.367	0.359	0.132
Testing Batch Accuracy	0.813	0.798	0.751	0.878
Testing Batch R-Squared	0.631	0.599	0.625	0.764

Wildlife Counting - RetinaNet Testing Comparison

	Cottonwood Eastface	NGilchrist Eastface	Fence Ends	Idaho	All Data
Mean Absolute Error (MAE)	0.000	0.000	0.138	0.174	0.132
Accuracy	1.000	1.000	0.870	0.849	0.878
R-Squared	1.000	1.000	0.750	0.680	0.764

Wildlife Counting - Testing Results





Wildlife Counting - State of the Art Comparison

- Counting Using Deep Learning Regression Gives Value to Ecological Surveys[1]
  - Attempted to count seals on beaches using aerial images
  - Found ResNet-18 was their best performing model
  - R^2 = 0.77 and MAE = 8.14
- Automatically Identifying, Counting, and Describing Wild Animals in Camera-Trap Images with Deep Learning[2]
  - Attempted to count wildlife from the Snapshot Serengeti Dataset
  - Found ResNet-152 was their best performing model
  - Reported 62.8% counting accuracy

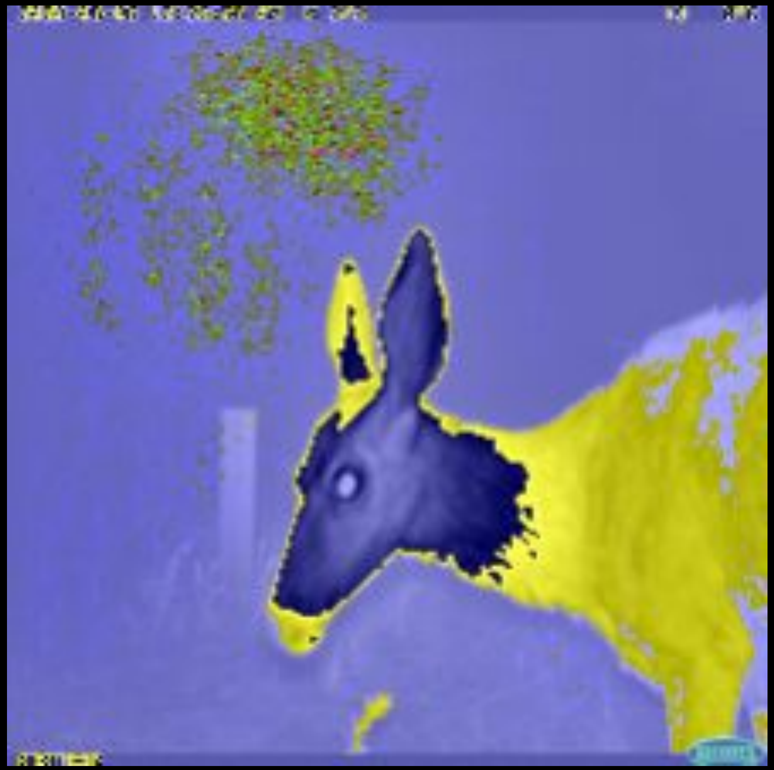
1. Hoekendijk, J.P.A., Kellenberger, B., Aarts, G. *et al.* Counting using deep learning regression gives value to ecological surveys. *Sci Rep* 11, 23209 (2021). https://doi.org/10.1038/s41598-021-02387-9
2. Norouzzadeh, Mohammad Sadegh & Nguyen, Anh & Kosmala, Margaret & Swanson, Ali & Packer, Craig & Clune, Jeff. (2017). Automatically identifying wild animals in camera trap images with deep learning. Proceedings of the National Academy of Sciences. 115. 10.1073/pnas.1719367115.

Wildlife Counting - Results Reflections

- Oversampling to overcome data imbalance had minimal improvements in the models ability to generalize
- Increasing our batch size as large as our hardware will allow improved model performance
- Huber loss outperforms mean squared error as loss for CNNs
- RetinaNet generalizing the best makes sense because it uses focal loss
    - Intuitively this could be described as the model focusing more on hard examples during training

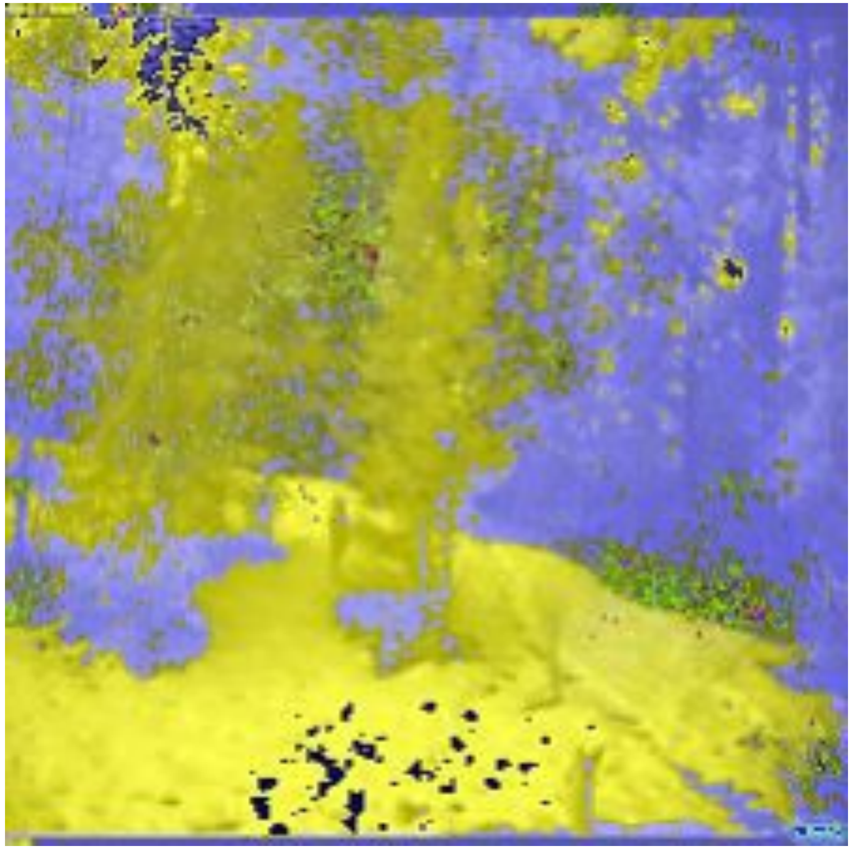
Wildlife Counting - Saliency Visualizations

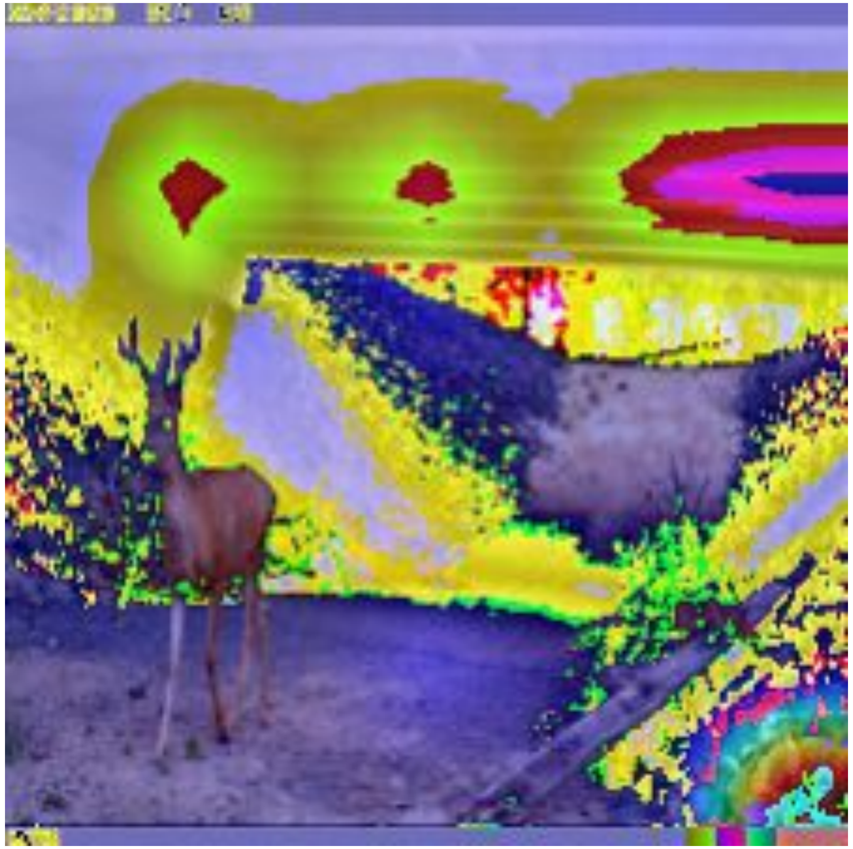
- Intuitively could be described as visualizations showing which part of the image most contribute to the models output
- Here are examples of correctly counted image where it is clear the shape of the wildlife is contributing to the prediction



Wildlife Counting - Saliency Visualization

- When the model is failing to perform as expected, saliency visualizations provide one mechanism to "audit" the model's performance
- Here are examples where the snow and background bridge appear to be driving the model's output rather than the desired wildlife





Wildlife Counting - Next Steps

- Communicate results and consider next steps with ODOT
- We could continue to increase the size and diversity of our training data which should continue to offer marginal model performance improvements
  - We increased our training data size from 3,995 to 6,947 to 19,767 images and saw improvements for all models and in particular their ability to generalize to new datasets

Thanks for listening!
Any questions?

Oregon State University