Math Industry Workshop — Natural Resources Canada / Ressources naturelles Canada

# OPTIMIZING MARKING TECHNIQUES FOR MARK-RECAPTURE STUDIES OF MOUNTAIN PINE BEETLES

J. BENESH, D. GOODSMAN, R. HAN, J. HOEPNER, H. HUANG, AND M. RAY

Industry mentor: Devin Goodsman, Entomologist, Northern Forestry Centre, Natural Resources Canada

Academic mentor:
Hui Huang, PIMS Postdoctoral Fellow, University of Calgary

Abstract. Whereas mark-recapture studies are sometimes used to estimate population size, mark-recapture studies initiated by Natural Resources Canada attempt to estimate movement of individuals in the population by marking them at source locations and recapturing them at various surrounding trap sites. A novel variation of traditional mark-recapture techniques developed by Natural Resources Canada involves coating trees with paper that fluoresces under black light such that the beetles are marked with paper dust as they emerge. Recaptured beetles are then photographed under black light. In this work, we classify images of the recaptured beetles as marked or unmarked using deep neural networks. In particular, we use transfer learning where existing top-performing classifiers are applied to our beetle image classification problem. We compare the performance of ResNet50 and EfficentNet base models when applied to images processed in a variety of ways.

## 1. Introduction

Since 1990, an outbreak of the mountain pine beetle (*Dendroctonus ponderosae*) has affected over 20 million hectares of forest in western Canada, making it the largest recorded insect outbreak in North American history. Mountain pine beetle adults disperse to attack and colonize trees in order to lay eggs beneath the outer bark. The process of attack and colonization disrupts nutrient flow and results in tree death [2, 5]. Although understanding beetle dispersal in this context is vital in making well-informed environmental decisions, tracking beetle relocation when adults disperse is challenging. Natural Resources Canada (NRCan) recently initiated a study designed to develop new and improved methods to quantify mountain pine beetle dispersal and to understand how far dispersing mountain pine beetles fly.

Mark-recapture studies typically involve the application of a harmless indicator to a small number of individuals, which are then released back into

This work was funded by PIMS and Mitacs.

the general population. The likelihood of recapturing a marked individual is thus inversely proportional to the size of the population, assuming nearly all of the marked individuals are still alive, and provided no significant immigration into or out of the population has occurred between the release and recapture dates. The goal of mark recapture studies initiated by NRCan is slightly different: Marked beetles are recaptured at various locations from release sites in order to better understand the dispersal process [6].

A recently developed NRCAN marking technique involves covering trees in paper that fluoresces under black light such that the beetles are coated in paper dust as they emerge, thereby allowing the marked beetles to naturally disperse without direct human intervention. Mountain pine beetles emerging from papered trees and control trees were later captured and photographed under black light.

Since manually classifying each image as marked or unmarked can be tedious and prone to error, it would be beneficial to automate the process using machine learning. The goal of this project is to develop a new algorithm to identify marked beetles by optimizing pre-existing image classification techniques. These techniques are discussed in further detail in Section 2, and their implementation is presented in Section 3. In Section 4, the project is summarized and potential improvements to the algorithm are addressed.

## 2. Deep Neural Networks

Deep learning has been well justified by its tremendous empirical success and state-of-the-art performance on various relevant real-life applications such as speech recognition [3], image recognition [4], language translation [10], and as a novel method for scientific computing [1]. It is an approach that enables the realization of complex tasks such as the ones mentioned above, by means of highly parameterized functions, called deep artificial neural networks $\mathcal{N} : \mathbb{R}^{d_0} \to \mathbb{R}^{d_L}$. A classical architecture is the one of feed-forward artificial neural networks of the type

$$\mathcal{N}(x) = \sigma\left(W_L^\top \sigma\left(W_{L-1}^\top \dots \sigma\left(W_1^\top x + b_1\right) \dots\right) + b_L\right), \tag{1}$$

where $L$ is depth of the network, the function $\sigma$ is a scalar activation function acting component-wise on vectors, for each layer $\ell = 1, \dots, L$, the matrix $W_\ell \in \mathbb{R}^{d_{\ell-1} \times d_\ell}$ represents a collection of weights, and the vector $b_\ell$ represents shifts/biases. The neural network (1) is then trained to minimize a given loss function (e.g, Mean Squared Error, Cross-Entropy, Kullback-Leibler divergence, or Wasserstein distances) over the parameters (weights and biases) of the network, usually measuring the misfit of input-output information over a given finite number of labeled training samples.

In this report we use Convolutional Neural Networks (CNNs) to solve our image classification problem. However, to train on a very large dataset, deep CNN models may take a significant amount of time. A way to bypass this process is to re-use the model parameters from pre-trained top performing

CNN models that were developed for standard computer vision benchmark datasets, such as the ImageNet image recognition tasks. This is the so-called transfer learning. One can see it as the deep learning version of "standing on the shoulder of giants". There are many top-performing models that are available for the basis for image recognition tasks, such as VGG (e.g. VGG19 [7]), GoogLeNet (e.g. InceptionV3 [8]), Residual Network (e.g. ResNet50 [4]) and EfficientNet (e.g. EfficientNetB0 [9]). In the following we are going to focus on the implementations on the Residual Network and the EfficientNet models, which are the state-of-the-art methods in imagine classification.

## 3. Exploration of the dataset

Our dataset consisted of 1057 images, each of which was labeled in 5 parts.

- The first component of each name is whether or not the tree segment from whence the beetle in the image emerged was **papered or not**. Papered bolts were considered marked whereas unpapered bolts can be considered unmarked.
- The second component of each name is the **color** of the paint that was applied to the outside of the trees from which beetles emerged: Possible values of the paint color include: transparent, green, pink, or control (no paint).
- The third component of each name is whether the marked and unmarked beetles were **mixed** in a jar together to test the persistence of the marking paper in a slightly more realistic context. We note that mixing could potentially lead to cross-contamination of unpapered beetles as a result of physical contact with marked beetles or with paper fragments that may have been shed from them.
- The fourth component is a **number** that is not unique to each beetle.
- The fifth and final component indicates whether beetles were photographed on their **dorsal (d) or ventral (v) sides**. We note that the tips of the abdomens and the mandibles on the ventral side are a location of higher concentrations of paper particles in some cases when beetles were marked.

A final component that was added to some of the images were comparison images which were labeled "light". These images were removed from the dataset as they were not imaged under blacklight and had a very distinct look to them (see Figure 1) in comparison to the rest of the images. This brought our total count down to 1013 images.

Finally, a series of images beginning with the word "Trap", which indicates that the beetles originated from a separate outdoor experiment in which standing trees that were infested with mountain pine beetles were papered. A number of Lindgren funnel traps were set up in the vicinity to capture beetles emerging from trees in the area. Most of the trapped beetles

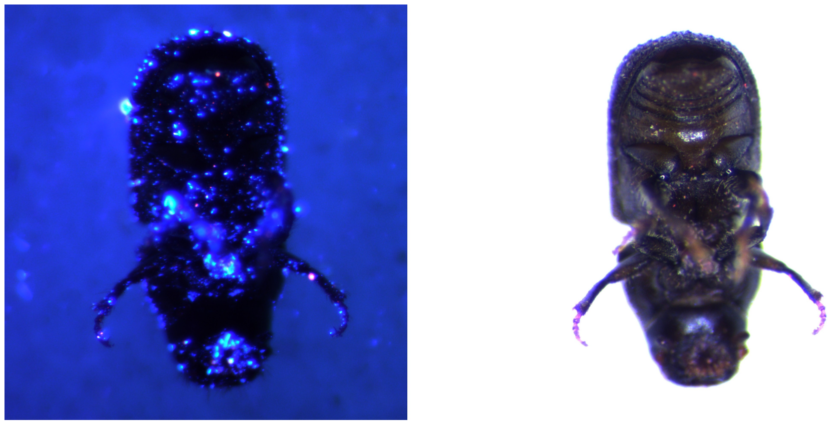


Figure 1. Blacklit vs. Lit Beetle

likely emerged from unmarked trees. For these beetle (and other insects), we do not know whether they emerged from papered trees. We chose to use this set of images as a validation set to see how well our machine learning algorithm of choice would be able to inform whether there are any marked beetles in the Trap set as well as the associated probabilities of being marked for each Trap insect potentially. From visual inspection, with the human classifier, we did not believe there were any marked beetles in this set.

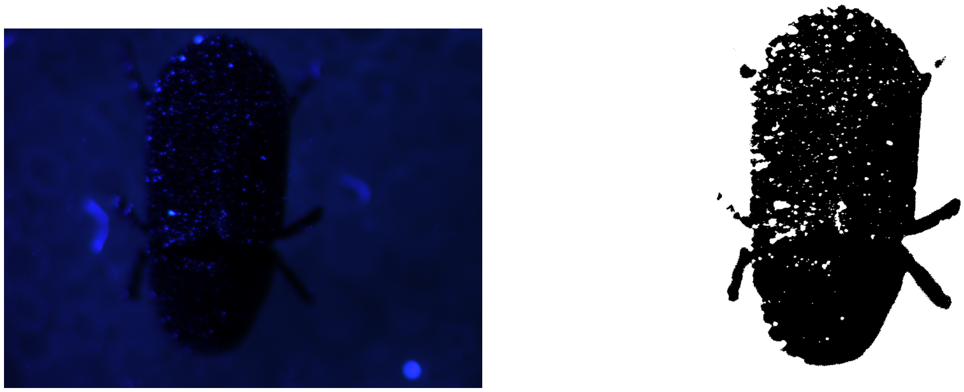
Using this labeling scheme we determined that we had 735 marked images, and 278 unmarked images, which is quite an unbalanced dataset. To mitigate this imbalance we chose to remove all images with the "Pink" tag, which fluoresced quite brightly under black light, and which formed the largest subset of our marked images. Upon removing these images, we were left with a total of 480 marked images, and 278 unmarked images, which gave us a much better balance to begin training our machine learning algorithms on.

We had two versions of this final dataset of 757 images - one with the original images and one with black and white images. The latter version was created using thresholding to convert each image into one with a binary colour scheme and then cropping around the beetle. We therefore suggestively refer to this data set as "Threshcropped". See Figure 2 to compare a sample from each data set.

## 4. Implementation and results

We ran our models with both the original and threshcropped datasets. We use the confusion matrix to visualize and test the performance of these models. The accuracy can be computed by the formula

$$\text{Accuracy} = \frac{\text{True positives} + \text{True negatives}}{\text{Total}}.$$



(A) Original Image (B) Threshcropped Image

FIGURE 2. Two versions of the dataset

The true positive rate is given by

$$\text{True positive rate} = \frac{\text{True positives}}{\text{Actual positives}}$$

We briefly describe each type of model below.

4.1. **Human Classifier.** A human classified the original images as marked or unmarked (see its confusion matrix in Figure 5 right). The prediction accuracy is

$$\text{Accuracy} = \frac{321 + 256}{755} \approx 76.4\%,$$

and the true positive rate is

$$\text{True positive rate} = \frac{321}{321 + 156} \approx 67.3\%.$$

Note here that there are two images missing because we have trouble loading them. We compare all other models' performances to this baseline.

4.2. **ResNet50 using original images.** *ResNet50* [4] is one of the most powerful deep neural networks which has won the ILSVRC 2015 competition because of its fabulous performance. It was proposed to solve the issue of vanishing/exploding gradient phenomenon. The idea is to use the "Residual Block" (see Figure 3) to skip connections and after-addition activations. If we denote by $\mathcal{F}(x) = \sigma(W^T x + b)$ a generic layer of the network, then the residual layer can be described as

$$x^{n+1} = x^n + \mathcal{F}(x^n). \tag{2}$$

Now let us present the transfer training result through ResNet50, which takes around 3 hours to run. First we refer to Figure 4 for the accuracy and loss evolution in terms of epochs on both training and validation set.

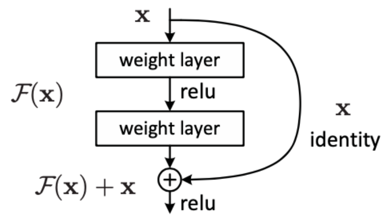



FIGURE 3. Residual Block, see [4]

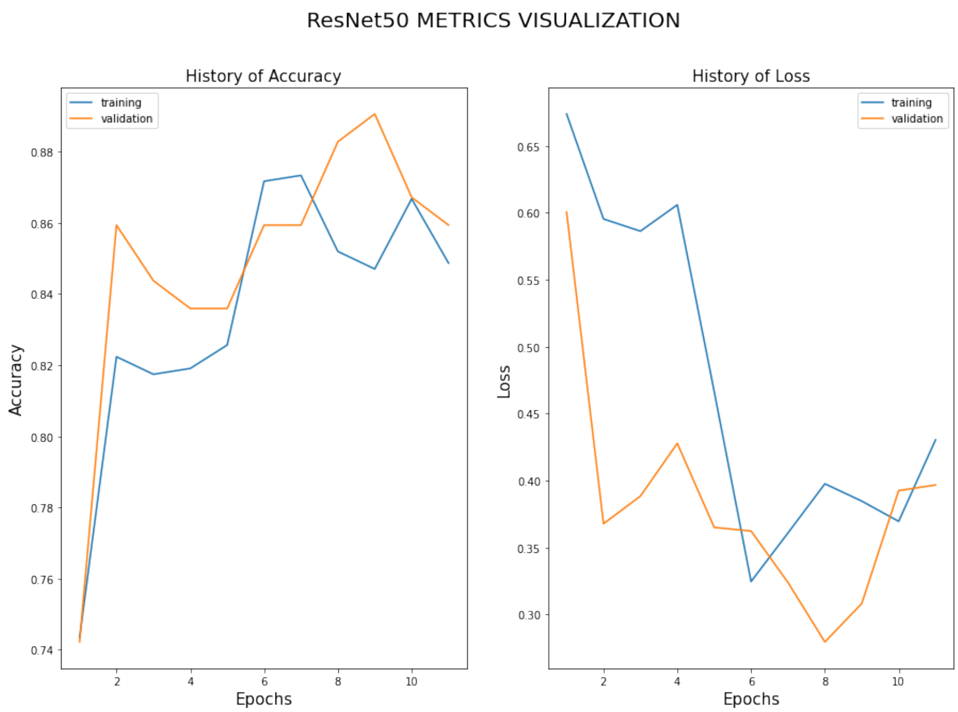



FIGURE 4. Accuracy and loss evolution in terms of epochs on training set (80%) and validation set (20%) from ResNet50.

It indicates that the accuracy can be up to 87.3% for the training set and 89.1% for the validation set.

Next we use the confusion matrix to test the performance of our classification model obtained from ResNet50. As it is showed in Figure 5 (left) that we have 431 true positives (marked beetles were being predicted marked), 48 false negatives (marked beetles were being predicted unmarked), 29 false positives (unmarked beetles were being predicted marked), and 249 true negatives (unmarked beetles were being predicted unmarked). The overall

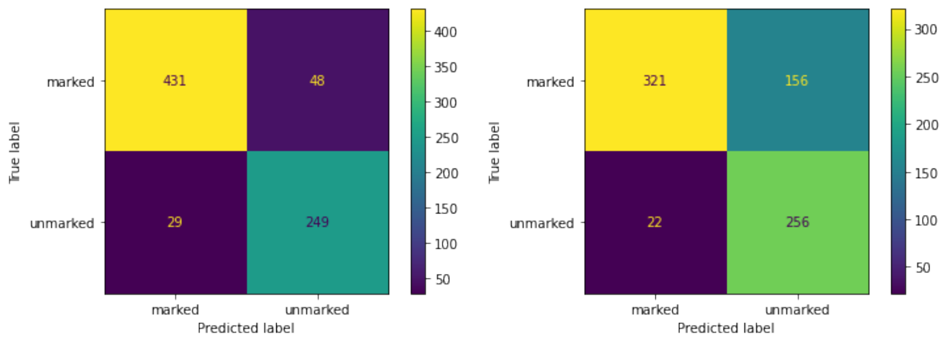



FIGURE 5. Left: Confusion matrix for the beetle classifier from ResNet50. Right: Confusion matrix for human classifier with the naked eye.

prediction accuracy is approximately 89.8% and the true positive rate is 90.0%.

4.3. **ResNet50 with preprocessing.** We use the sigmoid activation function at the output layer for the binary image classification. For each image, the model predicts the probability of being in the 'marked' class. Therefore, if the prediction of an image is greater than or equal to 0.5, we assign it the label 'marked'. Otherwise, we assign 'unmarked'. The validation accuracy is approximately 84.2% which is a little bit lower than the one without preprocessing, whose accuracy is 88.2%.

The confusion matrix summarizes the predictions made on the validation set in Figure 6.

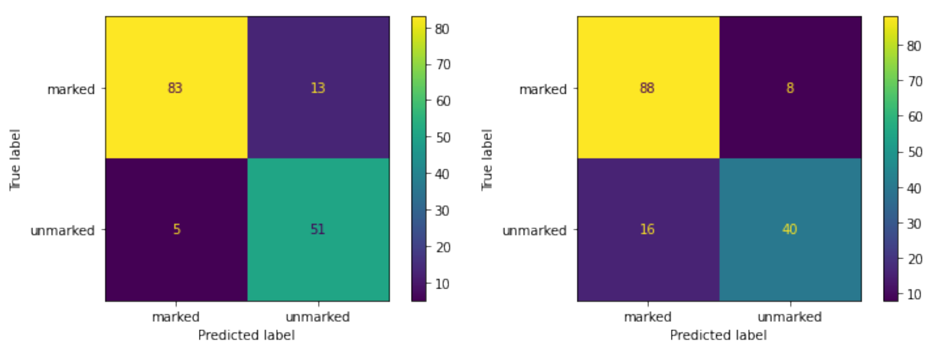



FIGURE 6. Confusion matrix for the validation set from ResNet50. Left: original images; Right: preprocessed

4.4. **EfficientNet.** *EfficientNet* was first introduced in [9], since then it has became one of the most efficient models that reaches state-of-the-art accuracy on both ImageNet and common image classification transfer learning

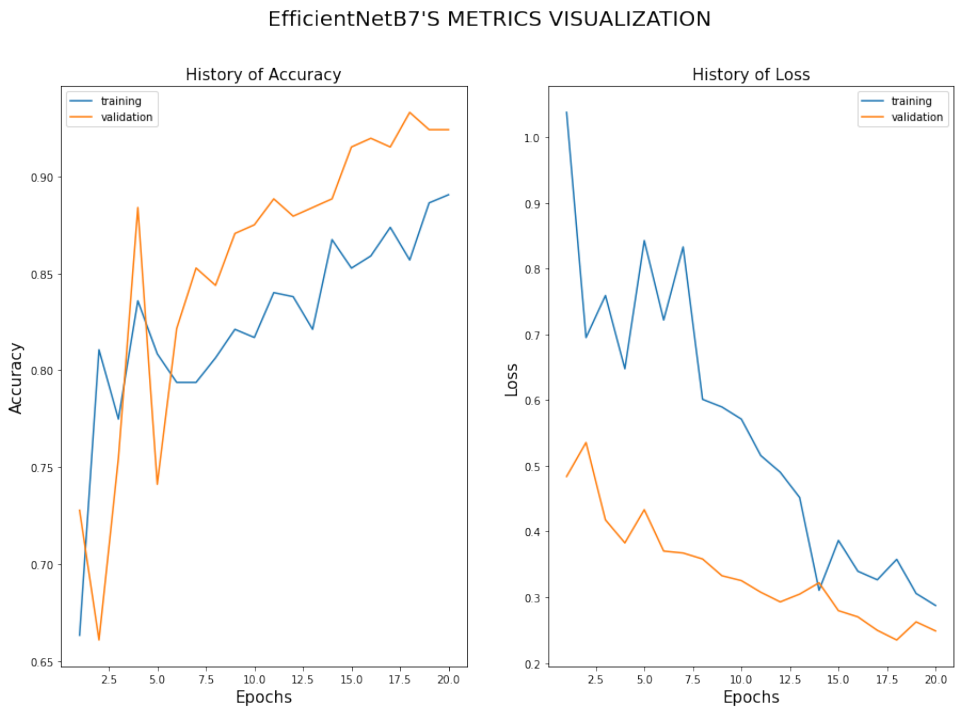



FIGURE 7. Accuracy and loss evolution in terms of epochs on training set (67%) and validation set (33%) from EfficientNetB7.

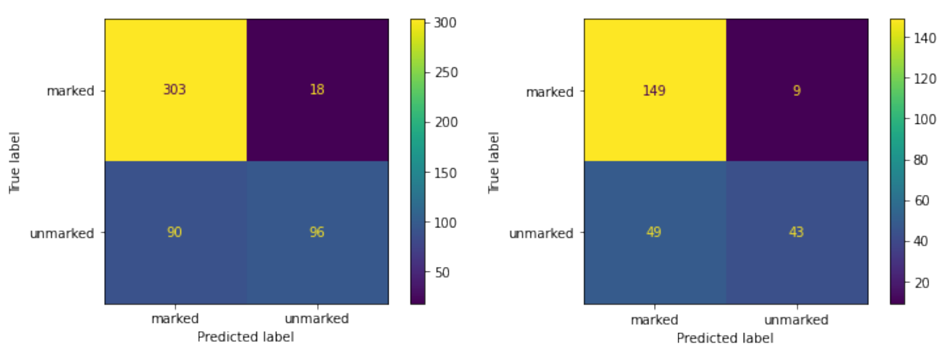



FIGURE 8. Left: Confusion matrix for the training set from EfficientNetB7. Right: Confusion matrix for the validation set from EfficientNetB7.

tasks. It proposes a compound scaling method to scale up CNNs in order to obtain better accuracy and efficiency. Unlike conventional approaches that arbitrarily scale network dimensions, such as width, depth and resolution, the EfficientNet uniformly scales each dimension with a fixed set of scaling coefficients. More specifically, it uses a compound coefficient $\varphi$ to uniformly scales network width, depth, and resolution in a principled way:

$$\begin{aligned} \text{depth: } & d = \alpha^{\phi} \\ \text{width: } & w = \beta^{\phi} \\ \text{resolution: } & r = \gamma^{\phi} \\ \text{s.t. } & \alpha \cdot \beta^2 \cdot \gamma^2 \approx 2 \\ & \alpha \geq 1, \beta \geq 1, \gamma \geq 1\,, \end{aligned}$$

with $\alpha, \beta$ and $\gamma$ to be determined by a grid search. Especially for our image classification problem, depending on the choice of the resolution of the input image, we can use a series of EfficientNet models from B0 to B7.

In the following we use the EfficientNetB7 with resolution $600 \times 600$ for our transfer learning, since our given images have very high resolution ($1944 \times 2592$). Compared to ResNet50, it takes at least 6 hours to run the EfficientNetB7. From the evolution of the accuracy and loss (see in Figure 7), one can see that the accuracy can be as high as 93.3%. However this is not consistent with the confusion matrices built from the training data and validation data by using the model we have trained. Indeed, it follows from Figure 8 that the prediction accuracy of the training set is

$$\frac{303 + 96}{507} \approx 78.7\%\,,$$

and the prediction accuracy of the validation set is

$$\frac{149 + 43}{250} \approx 76.8\%\,.$$

Especially the performance on predicting unmarked beetles is very poor. Different versions of EfficientNet were run on both the original and threshcropped images. We defer the reader to Table 1 in section 5 to see the performance metrics.

## 5. Conclusion

The following table summarizes the results from running different models. We saw that generally, we could classify beetle images much more accurately

| **Model** | **Dataset** | **Validation Accuracy** | **Training Accuracy** |
|---|---|---|---|
| Human Classifier | Original | 76.4% | N/A |
| Efficient B7 | Original | 93.3% | 89.1 |
| Effiecient B7 | Threshcrop | 78.3% | 85.6 |
| Efficient B5 | Original | 88.0% | 85.9 |
| Resnet 50 | Original | 89.0% | 87.3 |
| Resnet 50 | Threshcrop | 84.2% | 80.0 |

Table 1. Summary of model performance

and efficiently using machine learning tools compared to the human eye. The models ran faster with threshcropped images but with lower accuracy than the the model trained with original images. Overall, ResNet50 with original images performed the most effectively and consistently. We recommend training with preprocssed images when the computation resources are limited as the training can be faster with reasonable accuracy; otherwise, if computation resources allow and higher accuracy is desired, using the original images is suitable.

The major constraints we faced during the project were naturally those of time and resources. We were able to train three models with two datasets (original and preprocessed). More computational power such as cloud GPU should significantly improve the number of models we could train with ease and yield more information about how to make them better. Furturemore, training models with more computational power would enable us to gain insight into the optimal hyper-parameters such as the learning rate and the number of hidden layers.

Moving forward, one can try a number of things to improve predictions. This includes balancing out the training and validation datasets - removing the "Trap" and "mixed" images from the dataset and then choosing the 278 unmarked images and comparing them with a randomly chosen set of 278 beetles from the "marked" set. One can also make the training process more specialized - only train using the ventral sides. The reason for choosing the ventral side has to do with beetle behaviour and biology, which increases the likelihood of collecting paint particles on the mandibles and tips of the abdomens on the ventral side. Initial experiments in this direction looked promising. Finally, one can vary model parameters such as number of layers, image size, etc to experiment with further optimizations.

## References


[1] J. Berner, P. Grohs, and A. Jentzen. Analysis of the generalization error: Empirical risk minimization over deep artificial neural networks overcomes the curse of dimensionality in the numerical approximation of Black–Scholes partial differential equations. *SIAM Journal on Mathematics of Data Science*, 2(3):631–657, 2020.

[2] A. Dhar, L. Parrott, and C. D. Hawkins. Aftermath of mountain pine beetle outbreak in British Columbia: Stand dynamics, management response and ecosystem resilience. *Forests*, 7(8):171, 2016.

[3] A. Hannun, C. Case, J. Casper, B. Catanzaro, G. Diamos, E. Elsen, R. Prenger, S. Satheesh, S. Sengupta, A. Coates, et al. Deep speech: Scaling up end-to-end speech recognition. *arXiv preprint arXiv:1412.5567*, 2014.

[4] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[5] L. Safranyik. Mountain pine beetle: biology overview. In *Proceedings: Symposium on the Management of Lodgepole Pine to Minimize Losses to the Mountain Pine Beetle. USDA Forest Service, Intermountain Forest and Range Experiment Station, Gen. Tech. Rep. INT-262*, pages 9–12, 1989.

[6] L. Safranyik, D. Linton, R. Silversides, and L. McMullen. Dispersal of released mountain pine beetles under the canopy of a mature lodgepole pine stand. *Journal of Applied Entomology*, 113(1-5):441–450, 1992.

[7] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[8] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016.

[9] M. Tan and Q. Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International Conference on Machine Learning*, pages 6105–6114. PMLR, 2019.

[10] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008, 2017.

Joel Benesh, Department of Mathematics and Computer Science, University of Lethbridge
*E-mail address*: Joel.Benesh@uleth.ca

Devin Goodsman, Entomologist, Northern Forestry Centre, Natural Resources Canada
*E-mail address*: goodsman@ualberta.ca
*E-mail address*: devin.goodsman@canada.ca

Rachel Han, Department of Mathematics, University of British Columbia
*E-mail address*: 13hanr@gmail.com

Jules Hoepner, Department of Mathematics, University of Victoria
*E-mail address*: julesihoepner@gmail.com

Hui Huang, Department of Mathematics and Statistics, University of Calgary
*E-mail address*: hui.huang1@ucalgary.ca

Mishty Ray, Department of Mathematics and Statistics, University of Calgary
*E-mail address*: mraysamar@gmail.com