Bornean Orangutan Nest Identification Using Computer Vision and Deep Learning Models to Improve Conservation Strategies

Donna Simon[1,2], Keeyen Pang[3], Rayner Bili[4], Song-Quan Ong[2]*, Henry Bernard[2]*

[1]WWF-Malaysia, Sabah Office, Centre Point Complex, Jalan Centre Point, Kota Kinabalu, Sabah, Malaysia.
[2]Institute for Tropical Biology and Conservation, Universiti Malaysia Sabah, Jalan UMS, Kota Kinabalu, Sabah, Malaysia.
[3]Intrajasa Sdn.Bhd, Lot 9, Harapan Baru Light Industrial Estate. Mile 8, Jalan Labuk, 90009 Sandakan, Sabah, Malaysia
[4]Sabah Forestry Department, Mile 6, Labuk Road, Sandakan, Sabah, Malaysia.

Corresponding Author:
Song-Quan Ong[1]
Jalan UMS, Kota Kinabalu, Sabah, Malaysia.
Email address: songquan.ong@ums.edu.my

Abstract

Background. Regular population surveys are crucial for the evaluation of conservation measures and the management of critically endangered species such as the Bornean orangutans. Uncrewed aerial vehicles (UAV) are useful for monitoring orangutans by capturing images of the canopy, including nests, to monitor their population. However, manually detecting and counting nests from UAV imagery is time-consuming and requires trained experts. Computer vision and deep learning (DL) for image classification offer an excellent alternative for orangutan nest identification.
Methods. This study investigated DL for nest recognition from UAV imagery. A binary dataset ("with nest" and "without nest") was created from UAV imagery from Sabah, Malaysian Borneo. The images were captured using a fixed-wing UAV with a complementary metal-oxide semiconductor camera. After image augmentation, 1624 images were used for the dataset and further split into 70% training, 15% testing and 15% validation for model performance evaluation, i.e. accuracy, precision, recall and F1-score. Four DL models (InceptionV3, MobileNetV2, VGG19 and Xception) were trained to learn from the labeled dataset and predict the presence of nests in new images.
Results. The results show that out of… (how many variants you had at the beginning??) InceptionV3 has the best model performance with more than 99% accuracy and precision, while VGG19 has the lowest performance. In addition, gradient-weighted class activation maps were used to interpret the results, allowing visualization of the regions used by InceptionV3 and VGG19

for classification. This study demonstrates the potential of integrating DL into orangutan conservation and suggests that future research should focus on automatic nest detection to improve UAV-based monitoring of orangutans.

Introduction

All three orangutan species (it is worth to know - please mention them) living on Borneo and Sumatra have been listed as 'Critically Endangered' on the International Union for Conservation of Nature (IUCN) Red List since 2016, due to significant population declines (Ancrenaz et al. 2023). These population declines are primarily driven by habitat loss, degradation, and fragmentation, along with retaliation killings due to conflicts with humans (Ancrenaz et al. 2023). In Sabah, Malaysian Borneo, several measures have been introduced to protect orangutans, including forest restoration in degraded areas (Mansourian et al. 2020), expanding totally protected areas to 30%, and committing to sustainable timber production (Simon et al. 2019). Additionally, the 10-year Sabah Orangutan Action Plan (2020-2029) was developed to ensure the species' long-term viability in the region (Sabah Wildlife Department, 2020). Continuous monitoring is crucial to assess population trends and evaluate the effectiveness of these conservation efforts (Piel et al. 2022).

Orangutans are primarily found in lowland tropical rainforests (how many elevations?) , where they spend most of their time in the forest canopy (Manduell et al. 2012). They construct new nests each day, with juveniles relying on their mothers to build them (Permana et al. 2024). These nests are used for both night-time sleeping and daytime resting (Casteren et al. 2012). Since observing orangutans directly is difficult due to the dense canopy and their elusive nature, researchers often monitor populations by counting nests, which serve as reliable indicators of their presence (Kuhl et al. 2008; Santika et al. 2019). Population estimates are derived from nest densities (nests per km²), which are converted into orangutan numbers using established statistical methods (Ancrenaz et al. 2005; Kuhl et al. 2008; Pandong et al. 2018).

Orangutan nests are distinct from those of other animals. Orangutans typically build their nests in the upper canopy, around 11-20 meters above the ground (Casteren et al. 2012), and the nests are about 100 cm wide to accommodate their large body size (Kamaruszaman et al. 2018). The nest's base is made from thick branches, with thinner branches twisted and bent but not fully broken. This partial break, known as a "greenstick fracture," is unique to orangutan nests (Casteren et al. 2012). Leaves are added to form a flat sleeping platform. Orangutan nests are usually oval and asymmetrical, with the long axis oriented towards the tree trunk (Biddle et al., 2014). While most nests are built in the upper canopy, they can also be found at branch ends or close to the main tree stem (Rayadin and Saitoh 2009).

Various methods are used to count orangutan nests, including ground-based nest surveys (Pandong et al. 2018; Santika et al. 2019), helicopter surveys (Ancrenaz et al. 2005; Payne 1988; Simon et al. 2019), and the latest technology involving uncrewed aerial vehicles (UAVs) or drones (Hanggito 2020; Milne et al. 2021; Wich et al. 2015). Among these methods, drones are becoming increasingly important as they are relatively inexpensive compared to helicopters and can capture images or time-lapse video from the forest canopy, allowing many hard-to-access areas to be studied (Wich and Koh 2018). In contrast to ground and helicopter surveys, where nests are detected through direct field observations, drone imagery requires careful examination of each image on a computer to identify nests. As nests decay, the fresh green foliage withers and turns brown, making them stand out more clearly against the surrounding green canopy in the images (Figure 1). During manual nest identification, each nest is marked or labelled and then counted across all images. This allows researchers to calculate nest density, which can be used to estimate the orangutan population size.

To classify the images, it is important to consider the canopy classification perspective. Although nests made of branches and leaves can be distinguished from healthy trees as they decay over time (Casteren et al. 2012), a key challenge in using drone imagery to explore orangutan nests is that labeling nests from large volumes of image data still relies heavily on human experts, making the process tedious and time-consuming (Milne et al. 2021; Wich et al. 2015). Therefore, there is a need for an alternative method to identify nests from drone imagery that is as effective as, if not more effective than, human expertise for nest detection.

Figure 1 Example images for a drone image with orangutan nests circled in red

The integration of artificial intelligence (AI) into nest detection is a possible alternative to improve the efficiency of counting nests in drone image surveys. Machine learning (ML), is a branch of AI that enables computers to learn from a diverse array of data, recognize patterns and make decisions with minimal human intervention (Chahal and Gulia 2019). This study uses supervised learning, a category of ML in which algorithms are trained on labelled datasets to predict outcomes and recognize patterns. In contrast to unsupervised learning, supervised learning algorithms are trained with labelled data to learn the relationship between the inputs, i.e., features such as colour, texture, and shape of objects in aerial images, and the outputs i.e., labels indicating the presence or absence of orangutan nests in those images (Wang et al. 2016). The term "annotation" used to label the presence of orangutan nests on aerial images, is similar to the term "data labelling" in supervised ML, where the images of the nests need to be labelled and used as training data for model development.

It is also important to note that the matrices used to evaluate the context of the ecological study and the ML model may be similar, e.g. accuracy and precision, but they differ in context. In an ecological study, accuracy is the difference between the sample estimates and the true population value (Hellmann and Fowler 1999). For example, the accuracy of species richness is the difference between the estimate of species richness based on sample data and the true species richness of the population or community being sampled. Whereas, precision is the difference between an estimate of species richness based on sample data and the significance of all possible estimates of species richness based on all possible samples of the same size from the sampled population or community (Hellmann and Fowler 1999). In calculations, accuracy is measured by the mean square error of the estimator and precision by the variance of the estimator. On the other hand, the ML model is evaluated by the true value [the actual number of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) from a test set of the prediction] (Lebovitz et al. 2021). In most cases, the result of the evaluation of the model can be expressed in a layout table, the so-called confusion matrix, in which the proportion of TP and TN can be calculated. Accuracy, for example, is the proportion of all classifications that were correct, whether positive or negative, and precision is the proportion of all positive classifications of the model that are actually positive. Since the model is calculated on the basis of ground truth, further evaluation matrices can be calculated, e.g. recall or True Positive Rate (TPR) and False Positive Rate (FPR), which is crucial for the evaluation of a model with an unbalanced data set. For example, where the number of instances in one class (e.g., positive cases) is significantly lower than that in another class (e.g., negative cases).

Another subset of ML, often used in computer vision and image classification, is known as deep learning (DL) or deep convolutional neural networks (DCNN). DL has been used extensively for aerial images classification. Pearse et al. (2021), for example, have shown how DL models can classify tree species by learning complex visual features of tree species from aerial images with an accuracy of 92%, a sensitivity of 91% and a specificity of 94%. For monitoring orangutan population, Davies et al. (2019) combined LiDAR and behavioural data to reveal relationships between tree canopy structure and nest choice of orangutans in disturbed forests.

DL models are well suited for image classification as the architecture uses multiple layers of neural networks consisting of perceptions to model complex data (e.g. images with different colour channels) by learning features from images and making predictions (Smith et al. 2018). Further details on how the DL model works can be found in Wang et al. (2016), Purwono et al. (2022), the protocol paper by Isawasan et al. (2023) and Madhavan & Jones (2024). In image processing, DL is widely used for image classification and object detection in ecological studies, such as species identification, animal behavior classification and species diversity estimation from camera traps, video and audio recordings (Christin et al. 2019). For orangutan studies, Guo et al. (2020) developed Tri-AI, an automatic recognition system that identifies 41 primates and four carnivores with 94% accuracy. In addition, Desai et al. (2023) developed an annotated database of apes in different poses which enables object recognition for behavioral studies of apes in zoos.

Studies on orangutan recognition through computational methods to detect and count orangutan nests remain limited. Nest building, a unique daily behavior of orangutans for sleeping, offers valuable data for ecological monitoring, and by integrating DL techniques, it could enhance population monitoring efforts. Amran et al. (2023) initiated the study on the use of ML – Support Vector Machine (SVM) - in classifying the objects on the aerial images into branches, buildings and orangutan nests; Teguh et al. (2024) provided the most recent study (at the time of writing this manuscript) on orangutan detection using DL model, the You Only Look Once (YOLO) version 5 with 414 labelled orangutan nests and achieved a precision of 0.973 and a recall of 0.949. However, Teguh et al. (2024) applied an object detection algorithm and demonstrated the effectiveness of a DL model, but this raises additional questions. For instance, YOLO typically identifies and classifies objects in a single step, but alternative classification algorithms may offer improved performance. As biologists and ecologists, it is crucial not to treat these tools as black boxes. This study focuses on interpreting the outputs to gain insight into how DL models 'visualize' image patterns and identify the features utilized by neural network layers to classify tree canopy patterns as 'with nest' or 'without nest.' Understanding this process is essential for accurate ecological interpretation.

Therefore, this study aims to evaluate the effectiveness of different DL models in detecting orangutan nests from aerial images captured from two orangutan habitats in Sabah, Malaysia. More importantly, this study will visualize the model layers to understand how the features and characteristics of orangutan nests are 'learned' by the model. Specifically, the aim of this study is to create a labelled dataset of drone images containing the presence and absence of orangutan nests, and finally to develop and compare four DL models for detecting and predicting nest presence from drone images. Additionally, gradient-weighted class activation maps (Grad-CAM) were presented which can visualize the activation region used by the models to distinguish orangutan nests from the tree canopy.

Materials & Methods

Study site

These drone surveys were conducted in Sepilok Virgin Jungle Reserve (VJR) and Bukit Piton Forest Reserve (FR) in Sabah, Malaysia (Fig 2.). Both reserves are under the management of the Sabah Forestry Department and are known habitats for orangutans. It is estimated that there are about 200 (100-300) orangutans in Sepilok (Ancrenaz et al. 2005) and 176 (119-261) orangutans in Bukit Piton (Simon et al. 2019). The Sepilok VJR covers an area of approximately 40 km2 and is characterized by lowland dipterocarp and heath forests (Ball et al. 2023). The reserve has been designated as a protected area where logging is strictly prohibited to keep the forest canopy intact. In contrast, Bukit Piton FR, which consists mainly of dipterocarp lowland rainforest and is about 120 km2 in size, is severely degraded due to heavy logging and forest fires in the past. In 2008, a large-scale project was initiated to restore the forest for orangutans and the area was declared as a protected forest in 2012. Since then, the forest has slowly regenerated, with fast-growing tree species being used by the orangutans for nesting just three years after planting (Mansourian et al. 2020).

Fig 2. Location of Sepilok Forest Reserve and Bukit Piton Forest Reserve

Study duration

The Sepilok survey was conducted in July 2015 and covered an area of approximately 0.5 km2. A total of three flight missions were conducted to complete the survey with 1720 number of images. The Bukit Piton survey was conducted in January 2016 and covered an area of approximately 0.5 km2 resulting in 1911 images. A total of 4 missions were flown to survey the area in January 2016. Both surveys were conducted in the morning on a sunny day (Temperature, Relative Humidity?).

Equipment

This study utilizes UAV imagery captured by a fixed-wing drone assembled by ConservationDrones.org using an FX-71 frame. A Canon Power Shot S100 digital camera with RBG CMOS sensor (type number usu. with code, made in what country?) was installed in the drone. The drone was flown at least 100 meters from the highest point, which was determined using the Digital Elevation Model (DEM). The time-lapse recordings were made in 3-second intervals. The DL models compared in this study solve a classification problem in which the models process the entire image as a target object instead of recognizing different objects from one image (Sharma 2019). The datasets were created by combining images from both locations and having four human experts examine them for nests, annotate them and categorize them into two binary classes, i.e. images with nests and images without nests. This binary classification is needed to train the model and determine whether an image contains an orangutan nest or not. The field study and the use of the drone for aerial images were conducted in 2014 with the permission of the Sabah Forestry Department under reference number (JPHTN/PP 100-22/4/KLT.11(44)).

Pre-processing of the data and categorization

Using the image classification task, the entire images were classified either into “with nest” or “without nest”. For images with multiple nests, the images were pre-processed by cropping out the nest and labelling it as "with nest". The total number of aerial drone images from both Sepilok and Bukit Piton is 406 images, which were further classified into two classes, i.e., with nest (162 images) and without nest (244 images) (Table 1).

Nests from drone images have been identified by six orangutan field specialists, with more than two years of field experience in conducting ground and helicopter nest surveys. The identification of the orangutan nest at the same sites where drone images were captured is also consistent with the ground survey data which confirmed the presence of nests through direct observations. Then, the total number of images in each class was divided into three parts, also known as data splitting, with 70% of the total images used for training, 15% for validation and 15% for testing or a 70:15:15 ratio (Figure 2). The ratio of data splitting is based on the amount of data used for training and evaluation, and reducing the size of the training dataset tends to result in a poorly performing model. Therefore, an international standard of computer vision and DL competition (Fei-Fei et al. 2009) was referenced, along with insights from previous studies (Khan and Ullah 2022; Ong and Hamid 2022). Data splitting enables the machine to use the training set to obtain the weights and biases for classification. The validation set helped to better generalize the models to new, unseen data and prevent over-fitting while the testing set is to assess the model's performance. As the number of images was relatively small, each image was subjected to a rotation expansion of 0°, 90°, 180° and 270° and finally the number of images was increased by a factor of four (Ong et al. 2022; Chen et al. 2021), totaling to 1624 images used for the model development.

Models development

Model build-up

To develop the DL models, the convolutional blocks of the pre-trained convolutional neural networks (CNNs) were unfrozen for retraining purposes (a process in which the weights and biases that the model learns from the ImageNet are unlocked for a customized task, i.e., orangutan nest classification). This was done for four DL architectures – InceptionV3, MobileNetV2, VGG19 and Xception – to optimize them for the specific task of identifying nests from aerial images, as described in Ong et al. (2022). The Keras DL Framework on an NVIDIA Tesla A100 Google Compute Engine (GPU) platform was used to train and evaluate the models. The models were trained with the Adaptive Moment Estimation (ADAM) optimizer, which improves the stability and efficiency of the training process and enables efficient learning (Shao 2024). Three learning rates (0.01, 0.001 and 0.0001) with 32 batches were analyzed. The training process was set to 50 epochs, meaning that the model performed 50 complete iterations through the training dataset (Wang et al. 2016). Increasing the number of epochs allows the model to refine its parameters and could improve its performance. After developing the models, the performance of these models were evaluated using the four metrics of accuracy, precision, recall and F1-score (Table 2) (Hosin and Sulaiman 2015; Kumar 2020). In addition, the mean accuracy (number of correct predictions/total number of images) was compared between the models to test the significance of the four DL models. The code that used for the model development was publicly available at github with the link https://github.com/songguan26/Bornean-Orangutan-Nest-

Activation map to distinguish orangutan nests from aerial images

To gain further insight into how the neural network in the DL models can recognize the orangutan nest, Grad-CAM was used to visualize the area used by the neural network to classify the orangutan nest with a variety of normal tree canopy backgrounds. In general, one layer at a time was retrieved to extract low- and high-level features. The code that used for the model development was publicly available at github with the link https://github.com/songguan26/Bornean-Orangutan-Nest-

Results

Model performance

Four DL models were attempted, and the images were trained, tested and validated for image classification tasks by classifying UAV images into "without nests" and images "with nests" categories. Fig. 3, shows the performance of the four models in predicting the images with presence or absence of nests. It can be seen that VGG19 performs lower than the other models. InceptionV3, MobileNetV2 and Xception were ranked first, second and third. The Shapiro-Wilk normality test was performed to assess the normality of the accuracy values for the models across three learning rates. The results are as follows: InceptionV3 ($W = 0.75$, $p = 0.0000009$), MobileNet ($W = 0.99$, $p = 0.99$), VGG19 ($W = 0.95$, $p=0.566$) and Xception ($W=0.89$, $p=0.37$). Based on these results, only InceptionV3 is not normally distributed ($p < 0.05$). Therefore, a non-parametric test, the Kruskal Wallis H-test, was used to compare the models based on their accuracy values across three LRs. The result of the Kruskal Wallis H-test shows no significant difference (i.e., at the threshold p-value <0.05) in the accuracy of the four models at three learning rates ($H(3) = 6.751$, $p = 0.087$). Additionally, as most of the models are normally distributed except InceptionV3 with a very small p-value, the model performance is presented in Fig.4 using the mean value to better represent the data.

To assess the generalization capabilities of the model — its ability to make accurate predictions on new data (Caro et al. 2022) the training validation accuracy (TVA) and training validation loss (TVL) of the models across three learning rates on the test set were evaluated and presented in Table 2. The new data was validation splits (15%, in section methodology) that were never used in the model development. Although the epochs were set at 50, the early-stopping-method was employed — to prevent overfitting and underfitting (Cai et al. 2022) causing the model computation to halt early once the validation accuracy did not improve (epochs indicated in X-axis). The results of TVA and TVL (**Table 3**) show that LR 0.001 generally achieves a balance between efficient training and robust generalization across the models. Whereas, LR 0.01 risks instability and overfitting, which occurs when the model fits the training data too closely and failed to generalize to new data (Charilaou and Battat 2022). Meanwhile, LR 0.0001 results in slow or failed convergence and underfitting is shown by the poor performance of VGG19 model, which is incapable of learning the patterns in the training data (Jabbar and Khan 2015).

In addition, the confusion matrix for each model is shown in **Table 4** to visualize how well the classification model works by showing the correct and incorrect predictions made by the model, in comparison with the actual answer. The confusion matrix in binary classification consists of four components i.e. True positives (TP) is when the model correctly predicts the positive class; True negatives (TN) is when the model correctly predicts the negative class; False positives (Type-1 error) is when the model incorrectly predicts the positive class and False negative (Type-2 error) when the model incorrectly predicts the negative class (Saito and Rehmsmeier 2015). InceptionV3 at LR 0.01, LR 0.001 and Xception at LR 0.0001 have made all correct predictions. Meanwhile, InceptionV3 and MobileNetV2 at LR 0.0001, Xception at LR 0.01 and LR 0.001, as well as VGG19 at all LR, have a Type-1 error in nest prediction. Whereas MobileNetV2 at LR 0.001 has a Type-2 error in nest prediction.

Identification and visualization of input features

Heatmaps illustrate which parts of an image the model considers important by highlighting them in warm colors such as yellow, orange and red. Due to the superior overall performance of InceptionV3, five convolutional layers of the InceptionV3 architecture covering the low- and high-level features were used to visualize how the neural network identified the orangutan nest. Table 5 shows some examples of the convolutional layers of InceptionV3 compared to the original image of a human. The most common 2D convolutional layer "Conv2d" (Khan 2019) is used to visualize the region used by the model for classification. The heatmaps derived from Conv2d_89 and Conv2d_90 highlighted the corners of the images and underlined subtle colors on the nest itself. In contrast, the nest was emphasized in the Conv2d_91 and Conv2d_92 heatmaps. In addition, the upper right corner of the image was emphasized in the heatmap derived from Conv2d_93. Based on the result, the neural network was able to identify the features of the nest – edge, shape and texture – reflected in the different intensities of warm color. As mentioned by LeCun et al. (2015), there were blocks of low and high feature extraction in InceptionV3. Fig. 5 shows an example of the original image used to extract the feature for classification.

Discussion

The increasing use of drones to monitor orangutan populations since when?? This would be interesting to know ..could be an excellent alternative to improve the monitoring and protection of orangutan populations. However, the enormous amount of data generated by UAV imagery, which needs to be identified and annotated by trained experts, poses a major time and labor-intensive challenge. What about the supporting facilities (equipment) as well as internet stability, where this is sometimes challenging for some developing countries? Therefore, this study was conducted with the aim of evaluating the feasibility of using computer vision and DL to classify orangutan nests from UAV imagery.

This study is focused on image classification rather than on object detection (Sharma 2019). Specifically, it supports the second stage of the two-stage object recognition algorithm which in this case involves identifying the orangutan nest. The concept of two-stage detection consists of the first stage of detecting the object of interest (usually with the YOLO or SSD algorithm) and the second stage of a classifier by a DL algorithm (the DL models investigated in this study). Although many data scientists or ML engineers have proposed only the YOLO algorithm, which can solve both localization (detecting the position of the object of interest on an image) and classification in one step, detecting and classifying an orangutan nest on aerial images of tree canopies is a great challenge in reality (due to the very similar patterns of tree canopies) and requires a large number of aerial images as training data.

The result of this orangutan nest recognition study is consistent with that of Chen et al. (2014), who integrated various AI methods, including ML, optimization algorithms and adaptive decision-making systems, to develop intelligent systems capable of performing complex orangutan nest detection tasks from UAV imagery. In addition, the current study on the use of DL architectures with feature extraction from the images has continued the study of Amran et al. (2023) who used hand-crafted feature extraction and multi-class classification with Support Vector Machines (SVM) for orangutan nest in Borneo. Although Teguh et al. (2024) attempted to use YOLOv5 and achieved a precision of 0.973 and a recall of 0.949 when recognizing the orangutan nest from the drone images, this study has shown that orangutan nest recognition can achieve higher accuracy and precision when using lower computational power (and focusing only on the classification task). In addition, this study has shown that unlike YOLO (single-stage recognition algorithm), the use of transfer learning (transferring weights and bias in the classification of ImageNet images to another classification task) also helps to overcome the problem of data scarcity associated with the lack of sufficient training examples. While counting nests from the ground is easier than locating and counting individual orangutans, drone surveys capture only a fraction of nests in aerial views. Nests under the canopy in dense forests are often missed, and fresh green nests or those in advanced decay stages are harder to detect in drone images. As a result, this may cause insufficient training data for model training. Please highlight the challenges when using the UAV imagery vs. with manual observation incl. counting??

So far, this study was the first to compare four state-of-the-art pre-trained DL models - InceptionV3, MobileNetV2, VGG19 and Xception. The data was further augmented and the hyperparameters were refined by training for nest recognition from UAV imagery, resulting in high accuracies (>96%). The model performance result is in line with Ong and Hamid (2022) and Ong et al. (2022), where InceptionV3 is the best model for this task, while VGG19 performs the worst. When comparing between the three learning rates, the learning rate (LR) of 0.001 achieved the optimal performance, with fewer problems related to overfitting and underfitting. InceptionV3 with LR 0.001 performed well and delivered all correct predictions.

It is worth noting that VGG19 performs the worst in this study, in contrast to other studies which showed that VGG19 performs better than InceptionV3 and MobileNet. A look at the layouts of VGG19 (Table 6) compared to InceptionV3 (Table 5) shows that VGG19 is not able to recognize the features of the orangutan nest, which could be the main reason for the poor performance. Nevertheless, there are previous studies that also show that VGG19 performs worse. This emphasizes the need to compare DL models for a specific task.

To interpret the result of the computer vision system for the orangutan nest, the layers of the architecture with Grad-CAM were visualized, which to our knowledge is also the first report. Using Grad-CAM, the region of biases and weights defined by the perceptron within the DL architecture was able to highlight the shape and texture of the orangutan nest, which was later used in the classification block for classification. Considering the similarity of the present study to the task of classifying the canopy of a forest, this study result was compatible with that of Nezami et al. (2020), who used a multilayer perceptron (MLP) to classify tree species using aerial images generated from RGB and hyperspectral (HS) images and achieved an accuracy of 99.6% with the best 3D CNN classifier. Moreover, the result of this study in classifying tree canopy with and without orangutan nests is consistent with that of Huang et al. (2023), who used ResNet, ConvNeXt, ViT and Swin Transformer and achieved at least 96% accuracy in classifying tree species from aerial images.

However, there are still many aspects that require further investigation and improvement. One of these is the quality of aerial images. As mentioned by Huang et al. (2023), the degradation of image quality and aerial images at different altitudes needs to be explored further. The key question for future study is to determine what altitude achieves the ideal balance between drone flight feasibility and image quality. In this study, for example, a fixed-wing drone with a Canon Power Shot S100 RGB CMOS sensor was used, which was flown at the highest point of the treetops at an altitude of 100 meters. The image quality could be improved by using a multi-rotor UAV with better camera control. Image quality could also be improved by flying at a lower altitude where the camera is closer to the canopy and can capture more detail. However, this depends on the feasibility of the flight, where many factors determine the closest distance between the drone and the tree canopy, such as the availability of the crash sensor. With better image quality, further exploration can be conducted, such as classifying the nest decay stage of nests and increasing the ability to detect fresh green nests. Additionally, there is a need to augment both the quantity and diversity of aerial imagery to increase the robustness and subsequent generalization of the model. The diversity of the data could also include false positives and negatives in the training data to further improve the generalization of the model. Another important consideration is the deployment of the model to ensure its practical applicability. In the field for detecting and counting the number of orangutan nests. Additionally, building a model by local or regional dataset was always facing a challenge in generalizing good results for other similar datasets (e.g. by using the DL model in this study to predict aerial images from Indonesia).

Many future studies will aim to improve the model, software and hardware. However, it is vital to ensure that these improvements consistently contribute to orangutan conservation. Streamlining orangutan survey and monitoring processes to be more cost and time efficient, alongside leveraging computer vision and DL models for automatic annotation of orangutan nests from aerial images, could significantly advance orangutan monitoring efforts.

Conclusions

The present study encourages further development of DL models for the automatic detection of orangutan nests from aerial UAV images. Further research and refinement in this area could lead to more accurate and efficient methods for identifying nests. Nevertheless, additional data sets, especially from different forest types used by orangutans, such as forest patches within plantations, timber plantations, logged and unlogged forests, are crucial to improve the generalization of the model in the field. In the future, other remote sensing data such as through partnerships with other agencies could be incorporated to obtain more imagery and make significant improvements in this area.

Ethics statement

The drone was deployed in the primary protection forest where no residents lived, and only images of the canopy were collected, so there was no risk to people's privacy. The field study and the use of the drone for aerial photography were conducted in 2014 with permission from the Sabah Forestry Department under reference number (JPHTN/PP 100-22/4/KLT.11(44)).

Acknowledgements

This research was funded in part by WWF-UK, WWF-Malaysia and ETIKA. The drone survey activities were funded by AREAS. We would like to express our thanks to Wiveca Soripin, Middle Kapis, Tinrus Tindok and William Joseph from the WWF- Orangutan team for their expertise in nest annotation. Additionally, we thank the Sabah Forestry Department for permitting and supporting the survey activities.

References

Amran AA, On CK, Hung LP, Rossdy M, Simon D, See CS (2023) Bornean orangutan nests classification using Multiclass SVM. In 2023 IEEE Symposium on Computers & Informatics (ISCI):1-6

Ancrenaz M, Gimenez O, Ambu L, Ancrenaz K, Andau P, Goossens B, Payne J, Sawang, A, Tuuga A, Lackman-Ancrenaz I (2005) Aerial surveys give new estimates for orangutans in Sabah, Malaysia. PLoS Biology 3(1). https://doi.org/10.1371/journal.pbio.0030003

Ancrenaz M, Gumal M, Marshall A J, Meijaard E, Wich S, Husson S (2023) *Pongo pygmaeus* (amended version of 2016 assessment). The IUCN Red List of Threatened Species 2023: e.T17975A247631797

Ball J, Hickman S, Jackson T, Jing K X, Hirst J, Jay W M, and Coomes, D. A. (2023). Accurate delineation of individual tree crowns in tropical forests from aerial RGB imagery using mask r-cnn. Remote Sensing in Ecology and Conservation, 9(5), 641-655. https://doi.org/10.1002/rse2.332

Biddle L, Deeming D, Goodman A (2014) Morphology and biomechanics of the nests of the common black bird *Turdus merula*. Bird Study 62(1):87-95. https://doi.org/10.1080/00063657.2014.988119

Casteren AV, Sellers WI, Thorpe SK, Coward S, Crompton RH, Myatt JP, Ennos AR (2012) Nest-building orangutans demonstrate engineering know-how to produce safe, comfortable beds. Proceedings of the National Academy of Sciences, 109(18): 6873-6877

Cai Y, Wang Z, Yao L, Lin T, Zhang J (2022) Ensemble dilated convolutional neural network and its application in rotating machinery fault diagnosis. Computational Intelligence and Neuroscience 2022: 1-14. https://doi.org/10.1155/2022/6316140

Chahal A, Gulia P (2019), Machine learning and deep learning. International Journal of Innovative Technology and Exploring Engineering. 8(12), 2778-3075

Charilaou P, Battat R (2022) Machine learning models and over-fitting considerations. World Journal of Gastroenterology, 28(5), 605–607. https://doi.org/10.3748/wjg.v28.i5.605

Chen H, Guo S, Hao Y, Fang Y, Fang Z, Wu W, Li S (2021) Auxiliary diagnosis for Covid-19 with deep transfer learning. Journal of Digital Imaging 34(2): 231-241. https://doi.org/10.1007/s10278-021-00431-8

Chen Y, Shioi H, Montesinos CF, Koh LP, Wich S, Krause A (2014). Active detection via adaptive submodularity. In Proceedings of the 31st International Conference on Machine Learning: 55–63

Christin S, Hervet É, Lecomte N (2019) Applications for deep learning in ecology. Methods in Ecology and Evolution 10(10): 1632–1644. https://doi.org/10.1111/2041-210X.13256

Davies AB, Oram F, Ancrenaz M, Asner GP (2019) Combining behavioral and LiDAR data to reveal relationships between canopy structure and orangutan nest site selection in disturbed forests. Biological conservation 232: 97-107.

Desai N, Bala P, Richardson R, Raper J, Zimmermann J, Hayden B (2023) Open Ape Pose, a database of annotated ape photographs for pose estimation. Elife: 12

Fei-Fei L, Deng J, Li K (2009) ImageNet: Constructing a large-scale image database. Journal of vision, 9(8): 1037-1037

Guo S, Xu P, Miao Q, Shao G, Chapman CA, Chen X, Li B (2020) Automatic identification of individual primates with deep learning techniques. Iscience 23(8)

Hanggito MS (2020) Development of an unmanned aerial vehicle-based orangutan population assessment and monitoring method for the multifunctional landscape of East Kalimantan, Indonesia Open Access Theses and Dissertations https://scholarworks.utep.edu/open_etd/3166

Hellmann JJ, Fowler GW (1999) Bias, precision, and accuracy of four measures of species richness. Ecological applications, *9*(3), 824-834

Huang Y, Wen X, Gao Y, Zhang Y, Lin G (2023) Tree Species Classification in UAV Remote Sensing Images Based on Super-Resolution Reconstruction and Deep Learning. Remote Sensing, 15(11): 2942.

Isawasan P, Abdullah ZI, Ong SQ, Salleh KA (2023). A protocol for developing a classification system of mosquitoes using transfer learning. *MethodsX*, *10*, 101947.

Jabbar HK, Khan ZR (2015) Methods to avoid over-fitting and under-fittting in supervised machine learning (Comparative study). Computer Science, Communication and Instrumentation Devices.

Kamaruszaman SA, Nik Fadzly, Mutalib AH, Muslim AM, Atmoko SSU, Mansor M, Mansor A, Rupert N, Zakaria R, Hashim ZH, Sah ASR, Jamsari FF, Azman NM (2018). Measuring Orangutan nest structure using Unmanned Aerial Vehicle (UAV) and Image J. BioRxiv. https://doi.org/10.1101/365338

Khan MK, Ullah MO (2022) Deep transfer learning inspired automatic insect pest recognition. In Proceedings of the 3rd International Conference on Computational Sciences and Technologies. Jamshoro, Pakistan: Mehran University of Engineering and Technology : 17-19.
Khan AA (2019) What is Keras Conv2D https://medium.com/@arif_ali/what-is-keras-conv2d-f234e48fde6 Accessed 3 June 2024
Kühl H, Maisels F, Ancrenaz M, Williamson EA (2008) Best practice guidelines for surveys and monitoring of great ape populations occasional paper of the IUCN Species Survival Commission 36 (36). www.iucn.org/themes/ssc.
Lebovitz S, Levina N, Lifshitz-Assaf H (2021) Is AI ground truth really true? The dangers of training and evaluating AI tools based on experts 'know-what." *MIS quarterly* 45, no. 3
LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553): 436-444
Li J, Xu Q, Shah N, Mackey TK (2019) A machine learning approach for the detection and characterization of illicit drug dealers on instagram: model evaluation study. Journal of medical Internet research, 21(6), e13803.
Madhavan S, Jones TM (2024). Deep learning architectures. The rise of artificial intelligence. IBM Developer. https://developer.ibm.com/articles/cc-machine-learning-deep-learning-architectures/ Accessed on 18 September 2024
Manduell K L, Harrison M E, and Thorpe, S. K. S. (2012). Forest structure and support availability influence orangutan locomotion in Sumatra and Borneo. American Journal of Primatology, 74(12), 1128-1142. https://doi.org/10.1002/ajp.22072
Mansourian S, Vallauri D, France W (2020) Lessons learnt from 12 years restoring the orangutan's habitat: the Bukit Piton Forest Reserve in the Malaysian State of Sabah. https://www.researchgate.net/publication/343609942
Milne S, Martin JGA, Reynolds G, Vairappan CS, Slade EM, Brodie JF, Wich SA, Williamson N, and Burslem DFRP (2021). Drivers of Bornean orangutan distribution across a multiple-use tropical landscape. Remote Sensing, 13(3), 1–16. https://doi.org/10.3390/rs13030458
Nezami S, Khoramshahi E, Nevalainen O, Pölönen I, Honkavaara E (2020) Tree species classification of drone hyperspectral and RGB imagery with deep learning convolutional neural networks. Remote Sensing 12(7):1070
Ong SQ, Ahmad H, Majid AHA. (2021). Development of a deep learning model from breeding substrate images: a novel method for estimating the abundance of house fly (*Musca domestica* L.) larvae. Pest management science, 77(12), 5347-5355.
Ong SQ, Nair G, Yusof UK, Ahmad H (2022) Community-based mosquito surveillance: an automatic mosquito-on-human-skin recognition system with a deep learning algorithm. Pest Management Science 78(10):4092–104. https://doi.org/10.1002/ps.7028 PMID: 35650172

Ong SQ, Hamid SA (2022) Next generation insect taxonomic classification by comparing different deep learning algorithms. Plos One 17(12), e0279094. https://doi.org/10.1371/journal.pone.0279094

Pandong J, Gumal M, Alen L, Sidu A, Ng S, Koh LP (2018) Population estimates of Bornean orangutans using Bayesian analysis at the greater Batang Ai-Lanjak-Entimau landscape in Sarawak, Malaysia. Scientific Reports 10:1–11. https://doi.org/10.1038/s41598-018-33872-3

Payne J (1988). Orang-utan Conservation in Sabah (Report No. 3754). WWF-Malaysia, Kuala Lumpur. 274 pp.

Pearse GD, Watt MS, Soewarto J,Tan AY (2021). Deep learning and phenology enhance large-scale tree species classification in aerial imagery during a biosecurity response. Remote Sensing, 13(9): 1789.

Permana AL, Permana JJ, Nellissen L, Prasetyo D, Wich SA, Schaik CPV, Schuppli C (2024) The ontogeny of nest-building behaviour in Sumatran orang-utans, Pongo abelii. Animal Behaviour, 211:53-67 .ISSN 0003-3472, https://doi.org/10.1016/j.anbehav.2024.02.018.

Piel AK, Crunchant A, Knot IE, Chalmers C, Fergus P, Mulero-Pazmany M, Wich SA(2022) Non-invasive technologies for primate conservation in the 21st century. Int J Primatol 43:133–167.https://doi.org/10.1007/s10764-021-00245-z

Purwono P, Ma'arif A, Rahmaniar W, Fathurrahman HIK, Frisky AZK, ul Haq QM (2022) Understanding of convolutional neural network (cnn): A review. International Journal of Robotics and Control Systems, *2*(4), 739-748

Rayadin Y, Saitoh T (2009) Individual variation in nest size and nest site features of the Bornean orangutans (*Pongo pygmaeus*). American Journal of Primatology 71(5): 393-399. https://doi.org/10.1002/ajp.20666

Riedler B, Millesi E, Pratje P (2010) Adaptation to forest life during the reintroduction process of immature *Pongo abelii*. International Journal of Primatology, 31(4): 647-663 https://doi.org/10.1007/s10764-010-9418-2

Saito T, Rehmsmeier M (2015) The precision-recall plot is more informative than the ROC Plot When evaluating binary classifiers on imbalanced datasets. PLOS ONE 10(3): e0118432. https://doi.org/10.1371/journal.pone.0118432

Simon D, Davies G, Ancrenaz M (2019) Changes to Sabah's orangutan population in recent times: 2002–2017. PLoS ONE 14(7): 1–14

Santika T, Wilson K, Meijaard E, Ancrenaz M (2019) The power of mixed survey methodologies for detecting decline of the Bornean orangutan. https://doi.org/10.1101/775064

Sharma P (2019). Image classification vs. object detection vs. image segmentation. Analytics Vidhya. Available via Online. https://medium.com/analytics-vidhya/image-classification-vs-object-detection-vs-image-segmentation-f36db85fe81. Accessed 14 October 2024

Sabah Wildlife Department (2020) Orangutan action plan for Sabah 2020-2029. Kota Kinabalu, Sabah, Malaysia.

Smith J, Legg P, Matovic M, Kinsey K (2018) Predicting user confidence during visual decision making. Acm Transactions on Interactive Intelligent Systems 8(2), 1-30. https://doi.org/10.1145/3185524

Teguh R, Maleh IMD, Sahay AS, Pratama MP, Simon O. Object detection of the Bornean orangutan nests using drone and YOLOv5. Int J Artif Intell ISSN, 2252(8938), 1641.

Wang H, Lei Z, Zhang X, Zhou B, Peng J (2016). Machine learning basics. Deep learning, 98-164.

Wich S, Dellatore D, Houghton M, Ardi R, Koh L P (2015) A preliminary assessment of using conservation drones for Sumatran orangutan (*Pongo abelii*) distribution and density. Journal of Unmanned Vehicle Systems, 4(1): 45–52. https://doi.org/10.1139/juvs-2015-0015

Wich SA, Koh LP (2018) Conservation Drones: Mapping and Monitoring Biodiversity. Oxford University Press.