Towards automatic detection of wildlife trade using machine vision models

Kulkarni, Ritwik

2023-03

Towards automatic detection of wildlife trade using machine vision models

Ritwik Kulkarni [a,*], Enrico Di Minin [a,b,c]

[a] Helsinki Lab of Interdisciplinary Conservation Science, Department of Geosciences and Geography, FI-00014, University of Helsinki, Finland
[b] Helsinki Institute of Sustainability Science (HELSUS), FI-00014, University of Helsinki, Finland
[c] School of Life Sciences, University of KwaZulu-Natal, Durban 4041, South Africa

A B S T R A C T

Unsustainable trade in wildlife is one of the major threats affecting the global biodiversity crisis. An important part of the trade now occurs on digital marketplaces and social media. Automated methods to identify trade posts are needed as resources for conservation are limited. Here, we developed machine vision models based on Deep Neural Networks with the aim to automatically identify images of exotic pet animals for sale. We trained 24 neural-net models on a newly created dataset, spanning a combination of five different architectures, three methods of training and two types of datasets. Model generalisation improved after setting a portion of the training images to represent negative features. Models were evaluated on both within and out-of-distribution data to test wider model applicability. The top performing models achieved an f-score of over 0.95 on within-distribution evaluation and between 0.75 and 0.87 on the two out-of-distribution datasets (i.e., data acquired from a source unrelated to training data), therefore, showcasing the potential application of the model to help identify content related to the sale of threatened species on digital platforms. Notably, feature-visualisation indicated that models performed well in detecting the surrounding context in which an animal was located, therefore helping to automatically detect images of animals in non-natural environments. The proposed methods are an important step towards automatic detection of online wildlife trade using machine vision models and can also be adapted to study more broadly other types of online people-nature interactions. Future studies can use these findings to build robust machine-learning models.

1. Introduction

The unsustainable trade in wildlife is a major threat to the conservation of biodiversity globally (Maxwell et al., 2016) and is increasingly turning to online platforms. Wild populations are harvested at levels that threaten their persistence in order to supply wildlife markets with live animals, plants, and their derived products ('t Sas-Rolfes et al., 2019). Unsustainable trade in wildlife is also often linked to the introduction of exotic species that threaten the persistence of native biodiversity (Bellard et al., 2016). In addition, wildlife trade can help facilitate the spread of zoonotic diseases, such as SARS-CoV-2 (Bezerra-Santos et al., 2021; Morcatty et al., 2021). Online markets have created new opportunities to illegally trade wildlife at multiple scales, locally to globally (Lavorgna, 2014; Siriwat and Nijman, 2020). The internet provides cost-effective means for sellers to reach out to wildlife users (Siriwat and Nijman, 2020; Feddema et al., 2020; Sung and Fong, 2018). In turn, this can potentially increase demand for wildlife and their products and incentivize unsustainable hunting, collecting, and gathering of species in the wild. Innovative solutions are needed to cost-effectively monitor and help reduce the illegal global online wildlife trade.

Effective monitoring of global online wildlife trade requires automated content identification (Di Minin et al., 2019). In conservation science and practice, the use of machine learning or digital data analysis methods to monitor the online illegal wildlife trade is increasing but is still limited (Di Minin et al., 2018; Feddema et al., 2021; Qing et al., 2019; Microsoft, 2021). Compared to other tasks where machine learning methods are being used in conservation science, using machine learning to monitor the online trade in wildlife can be more challenging. Apart from the difficulty of obtaining high-quality labeled training datasets, the partly illegal and hidden nature of the trade, makes data discovery a challenging task (Di Minin et al., 2019). While natural language processing can be used in cases where text content is structured for analysis or to drive machine learning algorithms (Kulkarni and DiMinin, 2021; Stringham et al., 2021), online text content is often unstructured. For example, text scraped from a website is very noisy with no clear headings, fields, or relation between the text content. It

also contains software tags such as HTML script and/or non alphanumeric characters. This makes analysis of the content very challenging. Automatic image identification can help identify relevant information from online content when text processing is difficult (e.g., non-standard language use like colloquial or slang phrases and spelling errors, or the requirement of language agnostic tools due to multiple languages). Automatic image identification can help identify relevant information from online content when text processing is difficult (e.g., in the case of colloquial language and in presence of spelling errors, or when the requirement of language agnostic tools are high due to multiple languages). To the best of our knowledge this is the first study that explores the application of machine vision to identify in a generic way whether an animal is in captive or wild environment.

In this study, we investigate the application of automatic image classification using deep neural networks on a new dataset that we constructed and that contains images about the sales of live animals on the internet (see Fig. 1 for a schematic overview of model application). This study presents a novel attempt at building image classification models to help identify the trade of exotic live animals on the internet. It differs from other image identification studies in conservation science e. g., identifying the species from camera trap images (Miao et al., 2019; Tabak et al., 2018); or analysing land cover changes through satellite imagery (Bellis et al., 2008; Patrignani and Ochsner, 2015; Dube et al., 2014); or identifying domestic animals from wild on farmland (Chen et al., 2019). Our work is unique in terms of image classification for its application to monitoring online wildlife trade and its ability to deduce the context of the target object. The objective here is not to detect a pattern that identifies the animal, but rather to deduce if an animal is present in the image along with identifying the context (e.g., a cage) in which the animal is present. Image classification models can be sensitive to the background information and a change in context adversely affects model performance to detect species as shown in (Singh et al., 2020). With no existing datasets available for the task, we created a novel dataset using online resources such as digital market places and crowd sourced databases (see Section 2.2) and obtained images of live animals for sale along with images of animals in the wild. Rather than focusing on identifying individual species, the models aim to differentiate a post showing animals kept in captivity as opposed to posts where animals are found in the wild. This is a necessary condition for the practical applicability of the models in real world scenarios. Several different models were experimented which covered five deep learning architectures, two training datasets, three evaluation datasets, three types of training methods and different model sizes (see Section 2 for more details). The training objective of the models was to classify between images that have the presence of an animal in natural surroundings in contrast to animals in captive settings. Data were collected from a website[1] trading exotic animals, iNaturalist,[2] Google Image Search[3] and Flickr.[4] Thus, along with evaluating models in a standard manner by splitting training and testing data, we also evaluated the models with 'out of distribution' datasets to measure generalisability, which emphasises the need to go beyond the traditional data splits. Model performance was interpreted using feature visualisation to highlight areas of the image which impact model decision. The best model based on performance metrics and size efficiency was established for this task.

2. Methods

2.1. Overview

The three main stages of developing and evaluating machine learning models to determine whether an image is of a captive animal, are summarised below, with details following in the remainder of the section (see also Fig. 1).

1. Data Collection: We collected images from a website that advertises exotic pets for sale. After manually annotating the species present in the images, we collected images of the same species verified in the wild through iNaturalist (see Section 2.2). Further sets of images were collected from Flickr, a social media platform hosting images of the targeted species, to evaluate the performance of the model on data that models would see in case they were deployed in real-world settings.

2. Model Building: We constructed various well established architectures of image classification models that use Convolutional Neural Networks (CNN). All deep neural architectures essentially function as information transfer between cascading layers of the network, where each layer encodes different salient features of the input image. In repeated training cycles the layers adjust the connecting weights between each other in such a manner that the final layer produces the desired output to classify an image. We experimented with three training methods which included (i) using a model pretrained on a standard image dataset, and training only the final classification layer, (ii) training all the layers of the model without pre-training, and (iii) the combination of the above techniques (see Section 2.3).

3. Evaluation: Models were evaluated by comparing the predicted class for an image to the ground truth class. This is quantitatively measured using the metrics of precision, recall and f-score. Precision measures how accurate the prediction is in terms of the proportion of the correct predictions to incorrect ones. $Precision = \frac{TP}{TP+FP}$; where TP = True Positive, FP = False Positive, while Recall measures the coverage of the correct predictions as a proportion of the correct predictions to all the predictions that should have been correct. $Recall = \frac{TP}{TP+FN}$; where FN = False Negative. The geometric mean of these measures is termed as the F-score, $2\frac{Precision*Recall}{Precision+Recall}$. We go beyond the standard practice of evaluating the model only on a hold-out set of images (termed as within distribution test), and we also evaluate the model on images collected from a different source (out of distribution test) as a stronger test for the model (see Sections 2.2 and 3).

We trained a total of 24 deep neural network models which spanned a combination of five different architectures, three methods of training, two types of datasets and two sizes of network (large and small for some architectures). The architectures tested were AlexNet (Krizhevsky et al., 2012), DenseNet (Huang et al., 2017), ResNet (He et al., 2016), Squeezenet (Iandola et al., 2016) and VGG (Simonyan and Zisserman, 2014). We trained a small and large size model for DenseNet (DenseNet121 and DenseNet201), ResNet (ResNet18 and ResNet152) and VGG (VGG11 and VGG19) (the number after each model indicates the number of layers used in that architecture). These models have been highly successful on many image classification tasks and set benchmarks on datasets like Imagenet (Deng et al., 2009) and CIFAR (Krizhevsky and Hinton, 2009) (i.e., large annotated datasets with thousands of images and label categories used widely as standard benchmarks in machine vision development), while being implemented widely for other applications (Khan et al., 2020). Models were trained to discriminate between animals in captivity and the same species in the wild. Thus, the training images consisted of these two classes of data labeled accordingly. As no database was readily available for the task, we created ex novo two datasets for training: i) one that consisted of examples of wild animals (negative class) vs captive animals (positive class) and ii) another one where the negative class also included random images consisting of possible background objects (indicating non-wild environments such as cage, furniture, room of a house, etc.) (see Section 2.2 for details). The

---

[1] the website is available upon request from the authors, see Methods for more information.

[2] www.inaturalist.org.

[3] images.google.com.

[4] www.kaggle.com/hsankesara/flickr-image-dataset.
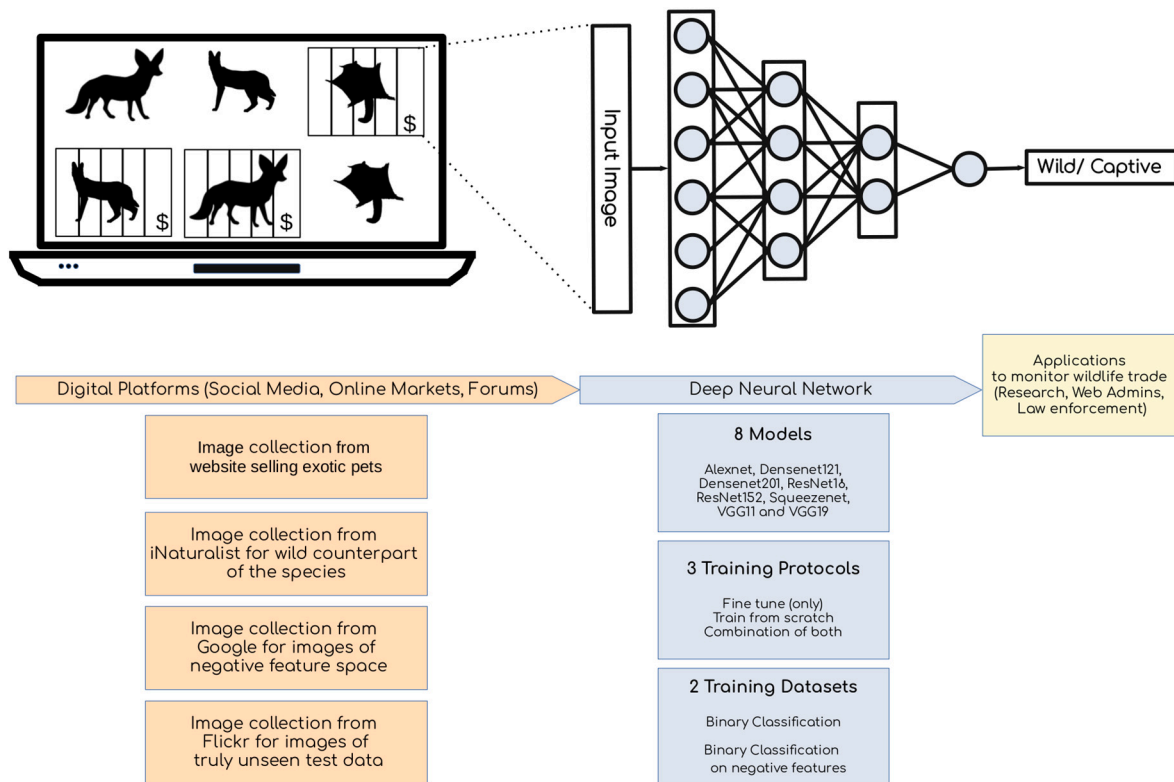
Fig. 1. Schematic flow diagram depicting the process from collecting images using four different sources, to training and evaluate the model. A combination of eight model designs and three training methods were studied on two types of training datasets.

second database was created to investigate the impact of model weights being focused solely on spurious image features as opposed to evaluating the entire context of 'animal in captivity'. Three types of training regimens were used to investigate the effect on model performance (Section 2.3): (i) Transfer learning, which entailed using the model pre-trained on the Imagenet data (Deng et al., 2009) and only training the last layer of the network as a binary classifier (Bengio, 2012); (ii) Training from scratch, where the entire model was trained from a blank state for the binary classification task; and (iii) Combination training, where the model was initially trained as in the transfer learning objective and then later the entire model was trained similar to training from scratch. Training types (ii) and (iii) were restricted to four models consisting of DenseNet and VGG that we found to have best results in type (i). Model performance was determined from accuracy and f-score metrics on three different test sets, namely one within distribution test data and two other test data sets which are out of distribution (see Section 2.2). Evaluating based on out of distribution data helps understand if the models can be applied in a general scenario that usually involves features not present in the training data. As the decision-making of neural networks is not easily interpretable (Huang et al., 2020), we used feature visualisation (Olah et al., 2017) for the top two best performing architectures (see Section 2.4) to gain additional insights on how the model was classifying the images. Feature visualisation enabled us to highlight parts of the image that have the strongest response in calculating the activation of neural units responsible for the classification. Therefore, we can determine on an abstract level of human cognition, whether the areas of interest (e.g., a cage for captive classification or forest for wild classification) are being identified by the model during decision making.

2.2. Data preparation

The target images for the positive class needed to be of live mammals (not domestic pets) for sale. We obtained these images from a peer to peer sales website where live mammals are traded and where sellers and buyers interact. As the website contains personal and identifiable information about the sellers, we refrain from publicly mentioning the website and adhere to data privacy regulations (the website was made available to referees during the peer-review of this manuscript). In order to assess the validity of the collected images, we compared them to other images obtained from general social media sites (Twitter, Facebook Marketplace) and Google searches with search queries of the form '(animal name) for sale'. Upon visual inspection, the images from web searches were found to be identical in nature to the images of animals found on the peer to peer website, and the source of the images appeared to be indistinguishable just by looking at the images. However, it is challenging to obtain data from these sources at volume needed to train models, due to accessibility issues (e.g., automated script blocks in data download), website regulations (e.g., terms prohibiting the use of data) and overall scarcity. See supplementary Fig. S1 for example of images from various sources. Thus, models trained on the collected data can be widely applicable to multiple digital platforms. We built a customised image downloading script in Python 3.6 (using the library BeautifulSoup) that was executed periodically to scan for new listings on the website and download images along with the associated text for that post. All images for each of the listings were transferred directly onto a cloud storage platform[5] which hosted all the data for the models. A total of 3051 images covering 58 species (including species listed in Appendix I, II and III of the Convention on International Trade in Endangered Species of Wild Fauna and Flora - CITES, see Supplementary Table S1 for the full list) were downloaded over a 4 month period.

The negative class of images comprised of animals in a non-sale context, essentially wild animals in their natural settings. Image classification models are known to be sensitive to small patterns in the image

---

(Olah et al., 2017). Therefore, to avoid positive detection purely based on particular species characteristics, we obtained images of the same species as in the positive class. We randomly sampled 1000 images from the positive class and made a list of all species mentioned for sale. While the number of new species found diminished strongly as we manually progressed through the images, no new species were found past image number 902 from the sample, therefore indicating that the sample covered most of the species from the positive class). Wild images were obtained from the iNaturalist database[6] which is an application allowing wildlife observers to record their sightings and including images. Data were obtained using the iNaturalist API (Application Programming Interface) implemented in Python 3.6 using the library pyinaturalist.[7] Only images with the quality designated as 'research grade' were used and transferred to the same cloud storage as mentioned above. 'Research grade' is a designation in the iNaturalist data when an observation passes a quality control process and multiple observers agree on the identification of the species. A total of 2505 images were obtained, thus creating a near balanced class division in the data. The combined 5556 images formed the first dataset designated as *data_no_bg*.

Another set of images was sourced to supplement the negative class of images. The model is intended to assign any image other than an image containing an animal for sale, as a negative detection. However, to test if the model would erroneously base the decisions on random objects in the positive class without taking into account the full context, we made a list of objects present in the positive class images such as table, chair, carpet, cage, human hands etc. (see full list in Supplementary Table S2). Negative sampling of the feature space has been known to affect learning outcomes of models (Saeidi et al., 2017). The list was compiled by manually viewing 500 images of animals in captivity and noting down objects in the background. The images were sourced using Google images as the search engine[8] allows for specific and flexible search queries such as 'empty cage for sale'. A custom built script written in Python 3.6 was used to source images based on a search query. Search queries were based on the list of background objects. Overall, 705 images were collected using this method and the combined total of all images (6261) allowed creating a second dataset designated as *data_bg*. Note that for both datasets the binary class division is near balanced with a slight tilt towards one class (positive in *data_no_bg* and negative in *data_bg*).

Following common practice, the data were split into 80 % train and 20 % test set. The test set was designated as *test_in*. However, to gain insights into model performance in real life settings we created two other test sets which contained images from a different source and mimic images from general social media. As the model is intended to monitor the sale of wild animals on digital market places and social media, the assessment of the model in that scenario (images representing general social media) is important. Thus, we had a test dataset that was partly out of distribution (features in the data can be different from the training set). As part of the "out of distribution" data, an image is made up of several different features that the model learns to detect. These features can be lines, curves, abstract shapes, colours and various other tiny patterns. The model learns to identify these set of features along with their combinations, which affect the classification decision made by the model. This set of features makes up the "distribution of features" and is akin to a statistical distribution. The model only learns the distribution present in the training data and since the test data is only a small part removed out of the training data, it tends to have the same feature distribution as the training data. Therefore, models tested in this manner do not indicate the performance of the model when applied in real world conditions as the feature set (or distribution) in the real world is potentially unlimited (or at least much wider than the

training set). Hence, a stronger test of the model is to create an "out of distribution" data where test data originate from a source which is not the same as the training data. These images were obtained from the Flickr dataset hosted on the website Kaggle.[9] The first out of distribution test data had 998 images from Flickr and 250 from the peer to peer website selling exotic animals and was designated as *test_out_pet*. The images from the website were downloaded at a different time as not to overlap with the data source used for *test_in* (total of 1248 images) and was designated as *test_out_pet*. Flickr included images of domestic pets such as cats and dogs. Many wild cat species resemble domestic cats, especially as kittens, while some dogs resemble wolves. Since the models were not trained to detect exact species, it was likely that such images tricked the models into making a positive decision for 'animal in captivity'. Thus, we created a second out of distribution test set that did not contain images of domestic pets as to measure the impact this could have on model performance. This dataset contained 1246 images (same 250 images sourced from the exotic animal website and 996 from Flickr). Flickr images are associated with text that describes the image (e.g. "a man walking a dog"). This facilitates filtering of images based on the description. We used Wordnet (Fellbaum, 2010) classes to exclude all images for which the description had a mention of the Wordnet class 'animal'. The Wordnet 'animal' class ensures all dogs and cats are covered along with some other domestic pets that may be in the data (although a visual inspection of a random sample of 100 images from Flickr did not result in any domestic pet other than cats and dogs). We called these data *test_out_nopet*. See Section 3 for the impact of the two test sets and the importance to test beyond the 80–20 % split.

2.3. Model training

Models were trained with three different training protocols. First, all eight models versions (Alexnet, Densenet121, Densenet201, ResNet18, ResNet152, Squeezenet, VGG11 and VGG19) were trained with transfer learning technique on two datasets (*data_bg and data_no_bg*) resulting in 16 models. Model training was carried out using the Pytorch library (Paszke et al., 2019) on an NVIDIA GPU (RTX 2070) with a fixed random seed. For transfer learning, models were pre-trained on ImageNet data and only the final layer of the model was modified to have two linear outputs for the two classes. Training was carried out with an early stopping protocol. Thus, training was terminated if there was no improvement in model performance for 15 consecutive epochs. Models with the best performance (lowest Cross-entropy Loss) on the *test_in* data were saved as final models. Supplementary Fig. S2 shows all models tend to approach a plateau after roughly 15 epochs (see supplementary Table S3 for additional information on training parameters).

We selected four models that were best performing (accuracy above 80 %) among the 16 (Densenet121, Densenet201, VGG11 and VGG19) and carried out two more training experiments. First, we trained these models on the same data as mentioned above but trained the models from scratch without pre-training to see if features for this task could be better captured. Second, we used the combination of transfer learning and training the full model with weights of all the layers modified. For the purpose of combination training, we first trained models for 10 epochs by fine-tuning of final layer but keeping the rest of the weights of the deeper layers frozen, as described above, and then unfroze all layers for the remaining epochs and applied the early stopping protocol. Early stopping was increased to 35 epochs because we had to reduce the learning rate for a stable learning progression (see supplementary Table S3). Similarly, batch size was reduced from 75 to 15 when all layers were trained to avoid memory issues.

Image augmentation (cropping, flipping, normalising and rotating) was used during training to reduce over-fitting to spurious features and expanding information in the training image distribution (Shorten and

---

Khoshgoftaar, 2019). Slightly different augmentation pipelines were used for the two classes to remove watermarks in specific places in the two classes, so as to reduce the possibility of the models using watermarks as features (see supplementary Table S4 for details on the augmentation pipeline).

2.4. Feature visualisation

We used feature visualisation to gain insights on how the model was classifying the images, thus attempting to explore which aspects of an image led to a positive prediction (Olah et al., 2017). A popular method for feature visualisation is to generate saliency maps (Simonyan et al., 2014) which highlight the areas of an image that lead to maximum impact on the weight gradients for the target class, thus indicating which human interpretable aspects of the image are involved in decision making by the model. Feature visualisation was implemented using the Flashtorch[10] library. We selected the top ten images from the test set which led to the highest activation of the output unit class, as they indicate the strongest decision bias for identification. We then generated saliency maps for those images.

3. Results

3.1. Effect of image background on model performance

Models were evaluated using the f-score and accuracy metrics on three different test sets (see Section 2.2 for details and supplementary Table S5, Table S6 and Table S7 for all the evaluation figures). Fig. 2 shows the performance of the eight models trained using two datasets, *data_bg*, in which a proportion of training images (11 %) consisting of random background objects, denoted as 'background' in the figure, and *data_no_bg* for which no specific background images exist, denoted as 'no background'. Models were then evaluated on three different test sets *test_in* (20 % set aside from training data), *test_out_nopet* (general images outside training distribution but without any pets) and *test_out_pet* (as previous but with some images of pets). Highest performance of all models (f-score mean: 0.94, s.d.: 0.01; accuracy mean 94.18, s.d.: 1.31) was seen on *test_in* indicating that the models were fairly successful in discriminating the context of the two classes. However, the performance dropped when the models were tested outside their range of training distribution and the top performing models were Densenet121, Densenet201, VGG11 and VGG19 with f-scores mean 0.76, s.d. 0.01, and accuracy mean 84.19, s.d. 1.53. Squeezenet (f-score 0.63, accuracy 58.39) and Alexnet (f-score 0.67, accuracy 66.74) were among the lowest performing models. Overall all models performed remarkably better on the out of distribution datasets when trained with the *data_bg* dataset compared to *data_no_bg* even tough the performance on within distribution test set(*test_in*) was similar for all models. When tested in the fine tuning paradigm, *test_bg* had f-score mean: 0.70, s.d. 0.04 and accuracy mean: 74.53, s.d 8.65, while *test_no_bg* had f-score mean: 0.61, s.d. 0.009 and accuracy mean: 47.67, s.d. 4.19. Thus, indicating the added generalisation ability of the model when trained on *data_bg*, evident only when tested out of distribution. Slightly better performance of all models was found on the *test_out_nopet* (f-score mean: 0.72, s.d.: 0.04; accuracy mean 76.92, s.d.: 9.1) compared to *test_out_pet* (f-score mean: 0.69, s.d.: 0.03; accuracy mean 72.92, s.d.: 7.4). Images of common pets caused a drop in performance of the models due to a strong similarity of pet images to target positive class animal images.

3.2. Effect of training methods

Accuracy and f-score of the four top performing models (Densenet121, Densenet201, VGG11 and VGG19) is shown in Fig. 3. Models

---

trained from scratch consistently under performed (f-score mean: 0.58, s.d.: 0.01; accuracy mean 62.85, s.d.: 4.2) compared to the 'combi_train' method (f-score mean: 0.80, s.d.: 0.04; accuracy mean 86.89, s.d.: 4.3). Training from scratch was worse in performance than the fine-tuning method, while the 'combi_train' method performed better than fine-tuning (indicated by horizontal lines in Fig. 3). The two top performing models were VGG11 (f-score: 0.87, accuracy 92.69) and Densenet121 (f-score: 0.85, accuracy 91.88), which were the smaller versions of the respective architectures in size. As for the fine-tune train method, evaluation on *test_out_nopet* was slightly better than *test_out_pet* for 'combi_train' but comparable for 'scratch_train'.

With lower performance on the out of distribution data compared to the within distribution data, we tested the precision of models to identify positive class images only, since this is of practical importance for the application of the models in real life scenarios. Table 1 lists the precision values for all the models trained with *data_bg* for positive class only as compared to the mean precision of the positive and negative classes. The models showed high precision to detect images of animals in captive contexts, except for the 'scratch_train' case. Highest precision was seen in the 'combi_train' case, with VGG19 as the top performer. Thus when the model flags the image as positive it has a lower tendency for false positives. Identical precision values across the two datasets are due to the fact that class distribution is near identical in the two test datasets.

3.3. Model efficiency

Fig. 4 shows the gain in performance for every additional parameter added to the architecture as calculated for models trained with the fine-tune paradigm on *test_in*. Squeezenet was the lowest performing model and also had the lowest number of parameters, thus we used it as the baseline to measure the impact of additional parameters on performance. 'gain' is the ratio of the increase in the f-score of the model for every additional parameter added over the Squeezenet baseline. See supplementary Table S8 for the number of parameters for each model. Densenet121 was found to have the highest gain, indicating that each additional parameter gave the maximum contribution in boosting the f-score above the baseline compared to all other architectures.

3.4. Feature visualisation

The two top performing models (Densenet121 and VGG11 with 'combi_train') were selected for feature visualisation to observe which features contribute the most for the positive class identification. Images which led to high output activation for the class labels were selected for visualisation, Fig. 5 (a) shows a prediction of Densenet121 for a fennec fox (*Vulpes zerda*). Along with the prominent features of the fox, like the eyes and nose, strong gradients are also found for the cage in the background, thus adding context to the animal being in the cage. Similarly, in the prediction for a marbled cat kitten (*Pardofelis marmorata*) (b), the gradients while being strong on the cat shape and fur pattern are also strong on the human hand, again signifying the captive context. Along the same lines images for the negative class, example shown in Fig. 5 (c), show a strong gradient response for the features of the animal, ring-tailed lemur (*Lemur catta*) and also the surrounding features of the forest, while (d) shows that just the presence of an animal, whiptail wallaby (*Macropus parryi*), is not enough to illicit a positive class response even though it is a close-up of the wallaby similar to captive images. See supplementary Fig. S3 for more examples of images of the negative class where human in natural surroundings with no animal presence or humans in urban settings including a close-up of a face, are not enough to have positive class identification.

Similarly, we performed feature visualisation for incorrect predictions of the model for images which produce high activation of output units for the opposite class with respect to the desired class. Fig. 6 shows False positives on top (a) European hedgehog (*Erinaceus europaeus*), an image where a wild hedgehog is photographed in an urban
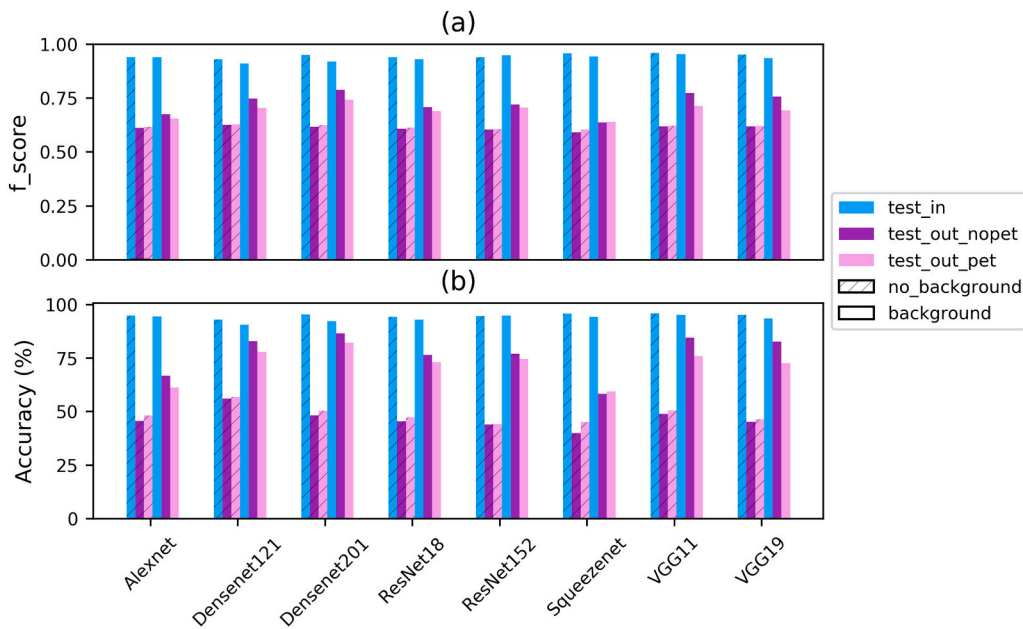
Fig. 2. f-score and Accuracy of eight models spanning a combination of two training methods and three test sets. Test methods indicated by 'background', in which 11 % of training images consisting of random background objects (f-score mean: 0.70, s.d. 0.04; accuracy mean: 74.53, s.d 8.65), and 'no background' (f-score mean: 0.61, s. d. 0.009; accuracy mean: 47.67, s.d. 4.19) for which no specific background images exist. Three test sets indicated by *test_in*(20 % set aside from training data) (f-score mean: 0.94, s.d.: 0.01; accuracy mean 94.18, s.d.: 1.31), *test_out_nopet* (fscore mean: 0.72, s.d.: 0.04; accuracy mean 76.92, s.d.: 9.1) (general images outside training distribution but without any pets) and *test_out_pet* (as previous but with images of pets)(fscore mean: 0.69, s.d.: 0.03; accuracy mean 72.92, s.d.: 7.4). See main text.
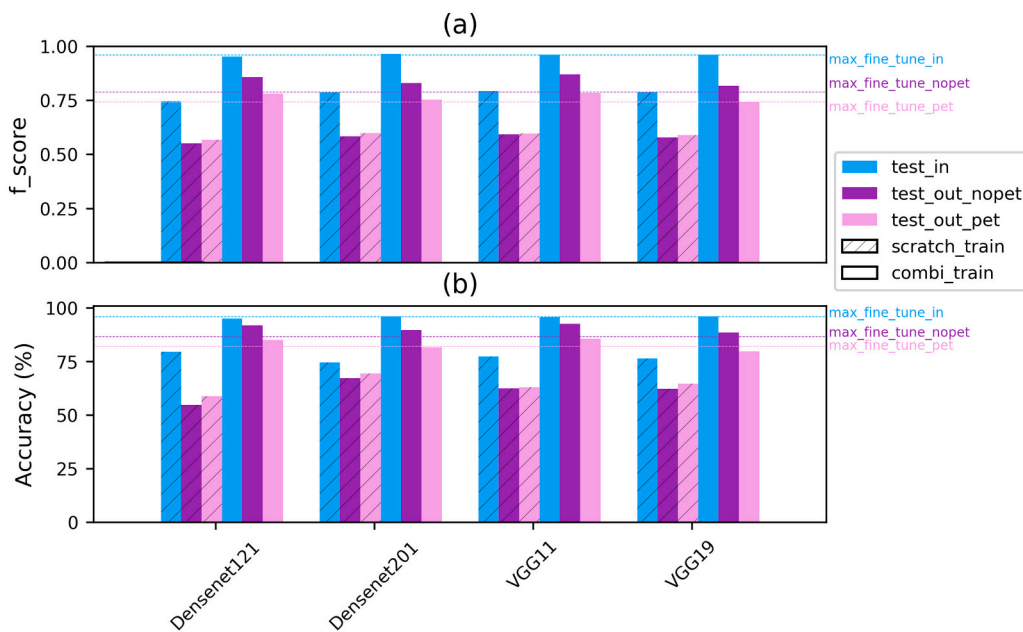
Fig. 3. Four top performing models having highest performance in fine-tune paradigm (Densenet121, Densenet201, VGG11 and VGG19) were tested for two additional training paradigms, training the full model from scratch (i.e. no pre-training), denoted as 'scratch_train' (f-score mean: 0.58, s.d.: 0.01; accuracy mean 62.85, s.d.: 4.2), and combination of fine-tuning and then training the full model once the final layer is semi-trained, denoted as 'combi_train' (f-score mean: 0.80, s.d.: 0.04; accuracy mean 86.89, s.d.: 4.3). 'combi_train' training method had higher performance than fine-tune method. Horizontal lines indicate the highest performance in finetune paradigm (Fig. 2) for the three respective test sets. Models were trained only with the *data_bg* dataset due to its advantage on generalisation and evaluated similarly on the three different test sets as for the fine-tuning paradigm.

environment, and (b) sika deer (*Cervus nippon*) again a seemingly wild individual in an urban environment. False negative examples are (c) bactrian camel (*Camelus bactrianus*) for sale, although in this image the camel appears to be in a natural looking background, the error originates from the model weights focusing on an obscure part of the image (sky). While (d) is an image of a serval (*Leptailurus serval*) where even though the cat is in a cage the surroundings are dominated by natural looking elements (leaves and branches).

4. Discussion

In this study, we introduced a method for automatic identification of images pertaining to wildlife trade and evaluated the models by creating a bespoke dataset pertaining to the sale of wild animals. We explained and applied our methods based on images obtained from a website known to post images of mammal species for sale. The novelty of our work is in creating a specialised dataset to distinguish between the context in which an animal is present and demonstrating that the model learns this context as shown by feature visualisation for correct prediction of the context.

Using different models allowed us to learn important aspects useful in the implementation of machine vision applications in detecting the sale of wild animals on the internet. First, it is crucial to evaluate the model on 'truly' unseen data as exhibited in Fig. 2 and Fig. 3. It is common practice to test the models on a small fraction of data set aside from training but derived from the same source as the training data. Unless the training data are of a very large size and covering most of the possible cases that the model will encounter in real applications, there is a high likelihood that models will overfit to the feature set of source (Radosavljevic and Anderson, 2013; Subramanian and Simon, 2013).

Table 1
Precision to detect positive class only (animals in captivity) compared with mean precision of both the classes combined, for all models trained with *data_bg*. VGG19 in the 'combi_train' paradigm has the top performance on detecting positive hits.

Model	Precision for positive class		Mean Precision for both classes	
	test_out_nopet	test_out_pet	test_out_nopet	test_out_pet
Alexnet	0.95	0.94	0.77	0.74
Densenet121	0.80	0.80	0.82	0.79
Densenet201	0.82	0.82	0.85	0.82
ResNet18	0.88	0.88	0.81	0.79
ResNet152	0.92	0.92	0.83	0.81
Squeezenet	0.91	0.91	0.71	0.71
VGG11	0.92	0.92	0.88	0.82
VGG19	0.91	0.91	0.86	0.80
Densenet121 - scratch_train	0.63	0.63	0.58	0.61
Densenet201 - scratch_train	0.52	0.52	0.62	0.63
VGG11 - scratch_train	0.68	0.68	0.65	0.65
VGG19 - scratch_train	0.61	0.61	0.62	0.63
Densenet121 - combi_train	0.94	0.94	**0.93**	**0.88**
Densenet201 - combi_train	0.95	0.95	0.92	0.87
VGG11 - combi_train	0.92	0.92	0.93	**0.88**
VGG19 - combi_train	**0.96**	**0.96**	0.91	0.86

Bold values indicates model with highest precision value.
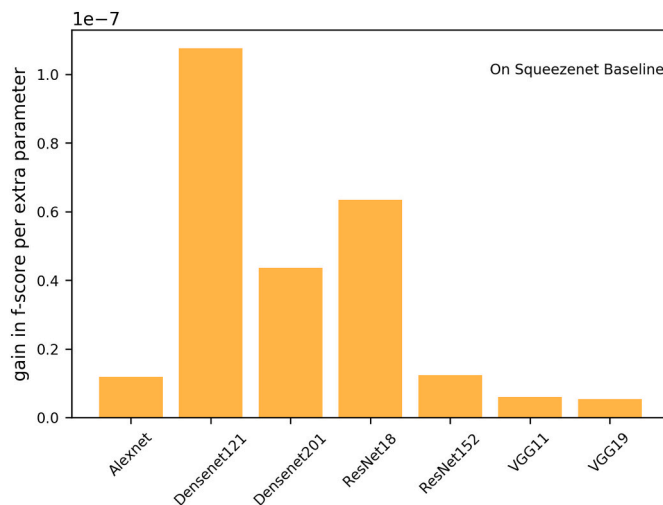
Fig. 4. Efficiency of the models measured as the increase in f-score from the Squeezenet baseline, for every additional parameter of the model compared to the number of parameters for Squeezenet. Squeezenet has the lowest number of parameters and also the lowest f-score. Densenet121 even though slightly lower f-score than VGG19 (0.95 vs 0.96) has significantly lower number of parameters (6,955,906 vs 139,589,442).

Thus, we recommend that a separate test data is created from a source different to the training data, and that it mimics real world data. Second, it is possible to improve the generalisability of the model and reduce the phenomenon of the model basing decisions on spurious objects by curating the data to include unrelated images containing the random objects that may appear in the target class. As most images of captive exotic animals are in a man made surrounding, including random images of man made objects (negative features) improved the performance of the model in the out of distribution data. This again emphasises the need to evaluate the model on 'truly' unseen data as the difference in performance and improved generalisability was seen on the out of distribution data and not on the 20 % test split. Third, the three different

training methods resulted in varying performances of the model. Transfer learning and fine-tuning the final layer for the specified task is an increasingly popular method in many fields including conservation science applications (Miao et al., 2019; Tabak et al., 2018). Training the model from scratch led to poor performance, most likely due to the smaller size of the data than required to train large models. Training deep models requires a large set of data, typically of the order of 10^6 to capture enough basic features of images. In situations for specific tasks such as wildlife trade, the datasets are quite small and transfer learning can help fill the gap by using pre-trained models. However, even though appreciable results were seen by fine tuning the final layer only, the best results were obtained using a combination of fine-tuning the last layer till the gradient decent becomes slower, and then letting all the layers learn the weights for subsequent epochs of training. This prevents the deeper layers from being destructively changed during the initial phase of learning and seemed to give added benefits on feature learning for the task when all layers were trained in a slow manner (reducing the learning rate). Fourth, the model is intended to act as an automated service to flag sales of wild animals to inform wildlife trade analyses. At the same time it can be used by online platforms and law enforcement agencies to help prevent the sale of prohibited live animals. In this regard, it is important for the model to not have many false positives which might affect its utility in real applications. The model showed high precision in detecting the images when an animal is present in a captive context. Hence, ensuring that when an image is flagged, it is of a high likelihood to be of interest. Fifth, as Densenet121 had the highest efficiency for the number of parameters used, it might be the best model to be used in such tasks of binary classification of context detection. Further experiments with diverse datasets are required to confirm the efficiency results.

We foresee a number of applications of the proposed methods in the context of helping monitor online wildlife trade. A first important application is to cost-effectively 'flag' content that requires further verification by social media platforms and authorities. In particular, focusing on species for which trade is illegal would help identify in real-time image content that is potentially connected to illegal wildlife trade, (see also (Cardoso, 2023) for another example of computer vision models used in the identification of images of pangolins). As machine learning methods apply to various forms of data, the proposed methods can also be used to potentially identify illegal wildlife trade content from X-ray scans (Pirotta et al., 2022; Brandis et al., 2018). Especially in the case of trade in bird species that are kept in cages this should be highly feasible. While the proposed methods were developed in accordance with data privacy and protection principles (Di Minin et al., 2021), they could potentially be used by relevant law enforcement agencies. Importantly, the proposed methods are openly available and can be used and adapted by other scientists and practitioners for the purpose of investigating wildlife trade on digital platforms. As these methods develop further and are known, illegal wildlife traders may change the way they post images as to avoid automated identification of suspicious content. Methods to trick a machine learning model and bypass detection are called"adversarial attacks" (Kurakin et al., 2016), however, they do require technical expertise to execute. Techniques to make machine learning models robust to adversarial attacks is an evolving research topic and involves training models with specific adversarial inputs (Kurakin et al., 2016). Finally, it remains important to further develop multimodal models that can process and relate information from multiple modalities (i.e., text, images, audio, etc.) and that can reduce the probability of misclassification.

5. Limitations

Using feature visualisation to observe areas of images which lead to strong classification decisions, the two top performing models seemed to have learned the general concept of combining animal features with background contexts of captivity (i.e., cages, human body parts,
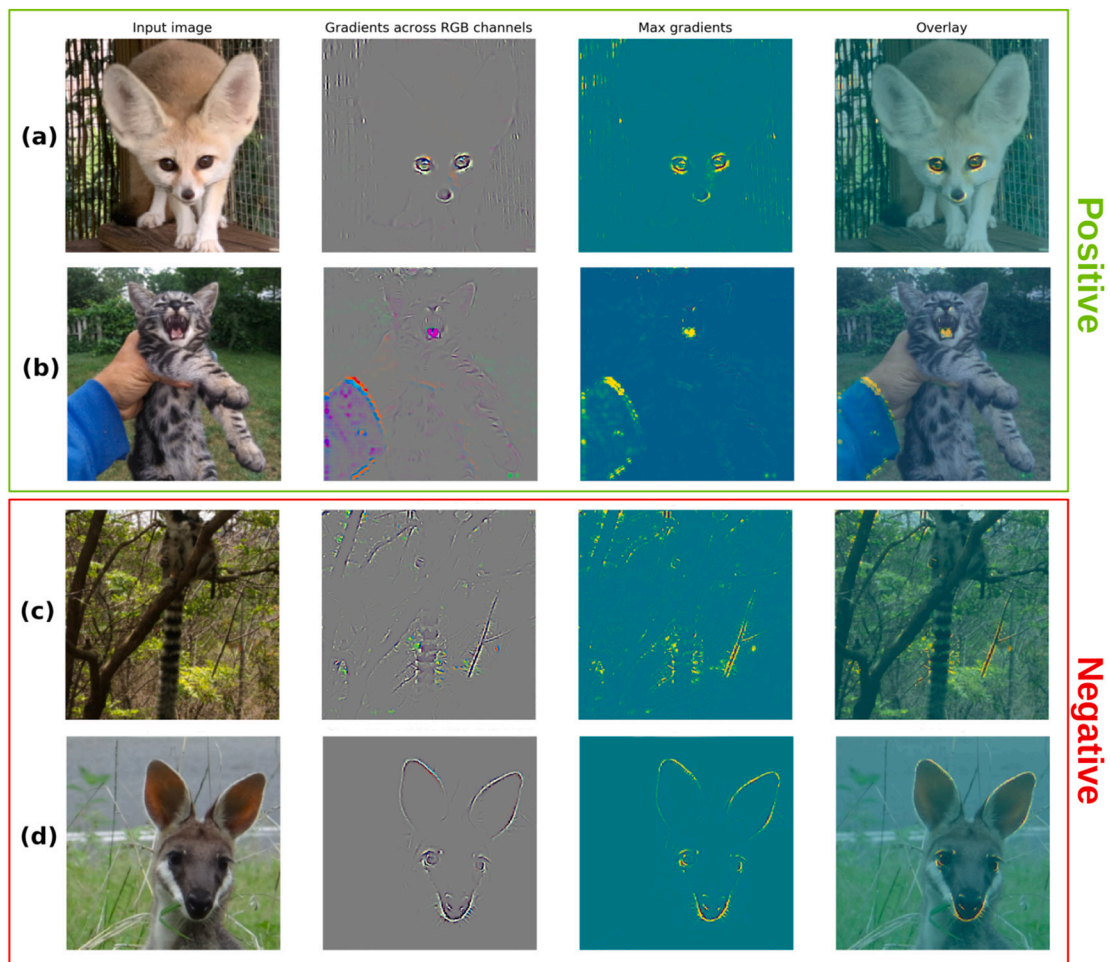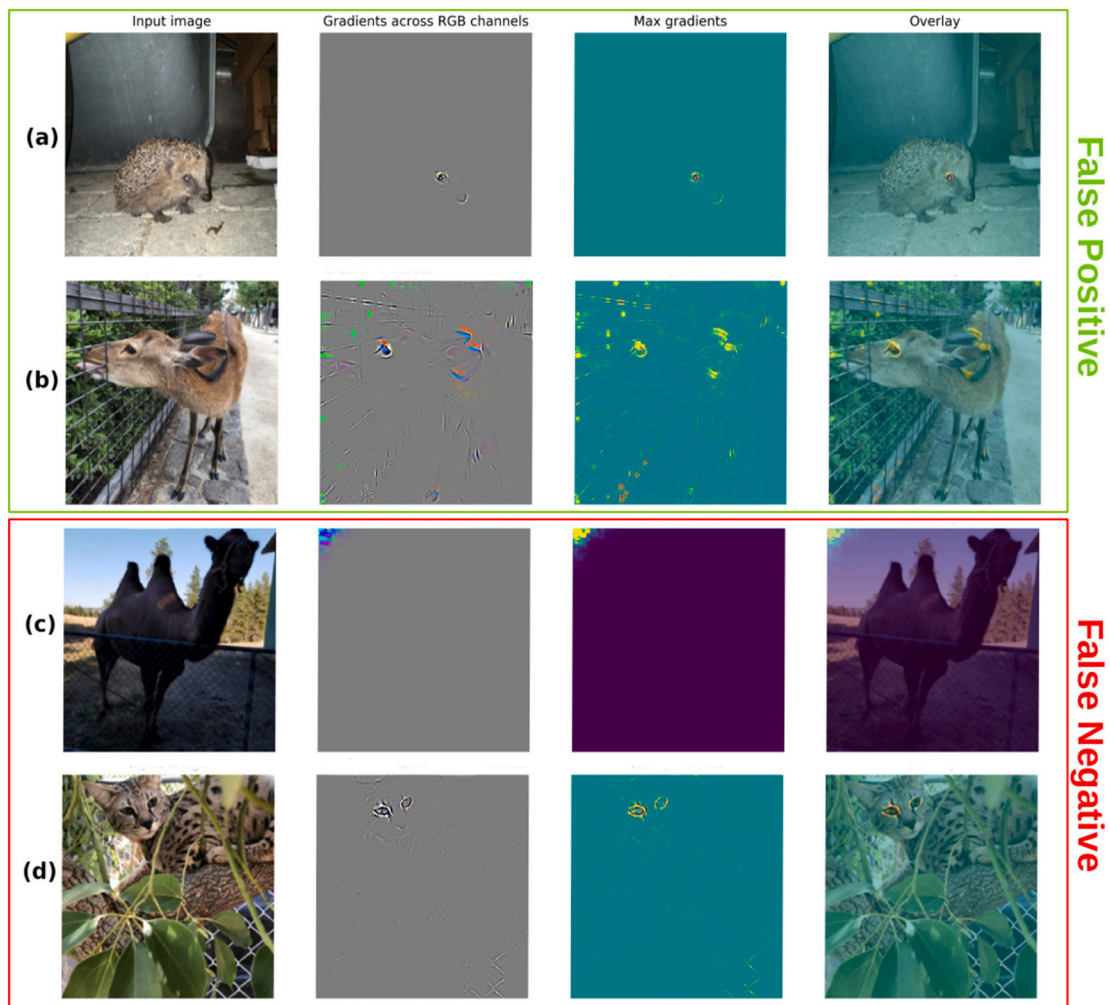
Fig. 5. Feature visualisation for images with highest output activation. Top two rows for positive predictions (animals in captive settings) while bottom two rows for negative prediction (animals in natural settings). (a) A captive fennec fox (*Vulpes zerda*) shows strong gradient response for facial features such as eyes and nose along with the background cage; (b) A marbled cat kitten (*Pardofelis marmorata*) evokes strong response for facial features along with a human hand holding it; (c) The tail pattern of a ring-tailed lemur (*Lemur catta*) and the surrounding tree pattern contribute the most to a negative decision; (d) Similar to (a), facial features of a whiptail wallaby(*Macropus parryi*) evoke strong gradient response but the absence of man-made surroundings lead to a negative decision. Overall, feature visualisation exhibits that model decision is based on 'animal + surroundings'.

household items, etc). However, there are also some examples (see supplementary Fig. S4) when the strongest predictors are some obscure parts of the image. Sometimes, sellers add text to the image (e.g contact information) and the model uses parts of the text as features. When the text is at a consistent location (e.g., name of the website) it has been cropped out during training so the model does not learn the classification based on watermarks, complete removal of spurious text however is not guaranteed. Similarly, text can sometimes be found on the images of wild animals and the impact is mitigated using random crop augmentation (see Section 2.3). Upon visual inspection of a sample, such cases did not appear to be too frequent in the data. It is also difficult to obtain images particular to the sale of wild animals and the positive class of images are sourced centrally from a single website. The website selected is a peer to peer digital sale platform, ensuring that the sales posts come from a diverse range of sellers and contains a diverse range of animals. Although it is expected that the data represent the nature of digital sales posts in general (examined by manual searched over the web), it is important for future experiments to source images of wild animals for sale from multiple platforms to enhance the model capabilities. Furthermore, images of domestic pets, common on social media, have the potential to trick the model as seen through the results for *test_out_nopet* vs *test_out_pet* (Fig. 2 and Fig. 3). Certain dog and cat breeds, especially as cubs, resemble wild canids and felids. Although

challenging, training models for species detection as in (Miao et al., 2019; Tabak et al., 2018) will help mitigate the issue to a certain extent or an exclusive pet detection model may be required. From an end user perspective, many times sales posts can be associated with some text information which can be used to further filter and refine relevant posts. In addition, wildlife trade exists for a wide variety of species spanning different taxonomic groups. The data we gathered of animals for sale was restricted by the website for mammals only. Further work is needed to expand the data for different taxonomic groups that would allow a diverse range of stakeholders to apply automated detection of wildlife trade. Finally, while having a strong potential to automatically analyse and flag online sale of threatened species, machine learning methods do possess implementation hurdles as they require specific infrastructure and expert personnel to manage the algorithms. This creates a need for the algorithms to be packaged in easy to use software tools that can be readily implemented with minimal training of on-field stake holders.

In conclusion, together with advanced methods for text classification, our methods offer the potential to investigate online wildlife trade in a cost-efficient manner. We foresee their use as part of scientific research investigating illegal wildlife trade and online monitoring activities by law enforcement agencies and online platforms. The same methods should now be applied to automatically identify wildlife trade images for other taxonomic groups beyond mammals and for multiple

Fig. 6. Feature visualisation for incorrect predictions (similar to Fig. 5). Top two rows for false positive predictions (animals predicted in captive settings but are wild) while bottom two rows for false negative prediction (animals in predicted natural settings but are captive). (a) European hedgehog (*Erinaceus europaeus*), a wild individual in an urban space, has strong gradient response for facial features but lacks discernible background response; (b) sika deer (*Cervus nippon*), a wild deer along a fence evokes strong gradient response for facial features and the fence which appears like a cage, thus leading to an incorrect response; (c) bactrian camel (*Camelus bactrianus*) a captive individual in a natural looking surrounding, however, the incorrect response is likely due to the model focussing on spurious image features in the top left corner; (d) serval (*Leptailurus serval*), a captive cat in a cage but has dominant natural looking elements like a tree branch and leaves.

online sources at the same time. Future work should now focus on creating new training datasets for further improvement of the models.

Credit authorship contribution statement

R.K. and E.D.M. designed research; R.K. developed the scripts to collect data, train and analyse machine learning models; R.K. and E.D.M. wrote the paper.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgment

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.biocon.2023.109924.

References

't Sas-Rolfes, Michael, Challender, Daniel W.S., Hinsley, Amy, Veríssimo, Diogo, Milner-Gulland, E.J., 2019. Illegal wildlife trade: scale, processes, and governance. Annu. Rev. Environ. Resour. 44 (1), 201–228.

Bellard, C., Genovesi, P., Jeschke, J.M., 2016. Global patterns in threats to vertebrates by biological invasions. Proc. R. Soc. B 283, 20152454.

Bellis, Laura M., Pidgeon, Anna M., Radeloff, Volker C., St-Louis, Véronique, Navarro, Joaquín L., Martella, Mónica B., 2008. Modeling habitat suitability for greater rheas based on satellite image texture. Ecol. Appl. 18 (8), 1956–1966.

Bengio, Yoshua, 2012. Deep learning of representations for unsupervised and transfer learning, 02 Jul. In: Guyon, Isabelle, Dror, Gideon, Lemaire, Vincent, Taylor, Graham, Silver, Daniel (Eds.), Proceedings of ICML Workshop on Unsupervised and Transfer Learning, volume 27 of Proceedings of Machine Learning Research. Bellevue, Washington, USA, pp. 17–36. PMLR.

Bezerra-Santos, Marcos A., Mendoza-Roldan, Jairo A., Andrew Thompson, R.C., Dantas-Torres, Filipe, Otranto, Domenico, 2021. Illegal wildlife trade: a gateway to zoonotic infectious diseases. Trends Parasitol. 37 (3), 181–184.

Brandis, Kate J., Meagher, Phoebe J.B., Tong, Lydia J., Shaw, Michelle, Mazumder, Debashish, Gadd, Patricia, Ramp, Daniel, 2018. Novel detection of provenance in the illegal wildlife trade using elemental data. Sci. Rep. 8 (1). October.

Cardoso, Ana Sofia, Bryukhova, Sofiya, Renna, Francesco, Reino, Luís, Chi, Xu, Xiao, Zixiang, Correia, Ricardo, Di Minin, Enrico, Ribeiro, Joana, Vaz, Ana Sofia, 2023. Detecting wildlife trafficking in images from online platforms: a test case using deep learning with pangolin images. Biol. Conserv. 279, 109905. January.

Chen, Ruilong, Little, Ruth, Mihaylova, Lyudmila, Delahay, Richard, Cox, Ruth, 2019. Wildlife surveillance using deep learning methods. Ecol. Evol. 9 (17), 9453–9466.

Deng, Jia, Dong, Wei, Socher, Richard, Li, Li-Jia, Li, Kai, Fei-Fei, Li, 2009. ImageNet: a large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp. 248–255. June.

Di Minin, Enrico, Fink, Christoph, Tenkanen, Henrikki, Hiippala, Tuomo, 2018. Machine learning for tracking illegal wildlife trade on social media. Nat. Ecol. Evol. 2 (3), 406–407.

Di Minin, Enrico, Fink, Christoph, Hiippala, Tuomo, Tenkanen, Henrikki, 2019. A framework for investigating illegal wildlife trade on social media with machine learning. Conserv. Biol. 33 (1), 210–213.

Di Minin, Enrico, Fink, Christoph, Hausmann, Anna, Kremer, Jens, Kulkarni, Ritwik, 2021. How to address data privacy concerns when using social media data in conservation science. Conserv. Biol. 35 (2), 437–446. March.

Dube, T., Gumindoga, W., Chawira, M., 2014. Detection of land cover changes around lake mutirikwi, Zimbabwe, based on traditional remote sensing image classification techniques. Afr. J. Aquat. Sci. 39 (1), 89–95.

Feddema, Kim, Paul Harrigan, K., Nekaris, Anne Isola, Maghrifani, Dila, 2020. Consumer engagement behaviors in the online wildlife trade: implications for conservationists. Psychol. Mark. 37 (12), 1755–1770. December.

Feddema, Kim, Harrigan, Paul, Wang, Shasha, 2021. The dark and light sides of engagement: an analysis of user-generated content in wildlife trade online communities. Australas. J. Inf. Syst. 25. Jun.

Fellbaum, Christiane, 2010. Wordnet. In: Theory and Applications of Ontology: Computer Applications. Springer, pp. 231–243.

He, Kaiming, Zhang, Xiangyu, Ren, Shaoqing, Sun, Jian, 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). June.

Huang, Gao, Liu, Zhuang, Maaten, Laurens Van Der, Weinberger, Kilian Q., 2017. Densely connected convolutional networks. In: n Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700–4708.

Huang, Xiaowei, Kroening, Daniel, Ruan, Wenjie, Sharp, James, Sun, Youcheng, Thamo, Emese, Min, Wu., Yi, Xinping, 2020. A survey of safety and trustworthiness of deep neural networks: verification, testing, adversarial attack and defence, and interpretability. Comput. Sci. Rev. 37, 100270. August.

Iandola, Forrest N., Han, Song, Moskewicz, Matthew W., Ashraf, Khalid, Dally, William J., Keutzer, Kurt, 2016. Squeezenet: Alexnet-level Accuracy With 50x Fewer Parameters and < 0.5 mb Model Size. arXiv preprint arXiv:1602.07360.

Khan, Asifullah, Sohail, Anabia, Zahoora, Umme, Qureshi, Aqsa Saeed, 2020. A survey of the recent architectures of deep convolutional neural networks. Artif. Intell. Rev. 53 (8), 5455–5516.

Krizhevsky, Alex, Hinton, Geoffrey, 2009. Learning Multiple Layers of Features From Tiny Images.

Krizhevsky, Alex, Sutskever, Ilya, Hinton, Geoffrey E., 2012. Imagenet classification with deep convolutional neural networks. Adv. Neural Inf. Proces. Syst. 25, 1097–1105.

Kulkarni, Ritwik, DiMinin, Enrico, 2021. Automated retrieval of information on threatened species from online sources using machine learning. Methods Ecol. Evol. 12 (7), 1226–1239. May.

Kurakin, Alexey, Goodfellow, Ian, Bengio, Samy, 2016. Adversarial Machine Learning at Scale.

Lavorgna, Anita, 2014. Wildlife trafficking in the internet age. Crime Sci. 3 (1). May.

Maxwell, S., Fuller, R., Brooks, T., et al., 2016. Biodiversity: The ravages of guns, nets and bulldozers. Nature 536, 143–145. https://doi.org/10.1038/536143a.

Miao, Zhongqi, Gaynor, Kaitlyn M., Wang, Jiayun, Liu, Ziwei, Muellerklein, Oliver, Norouzzadeh, Mohammad Sadegh, McInturff, Alex, Bowie, Rauri C.K., Nathan, Ran, Yu, Stella X., Getz, Wayne M., 2019. Insights and approaches using deep learning to classify wildlife. Nat. Sci. Rep. 9 (1). May.

Microsoft, 2021. First-of-its-kind Multispecies Ai Model to Detect Illegal Wildlife Trafficking is Ready to Roll Out to Airports. November.

Morcatty, Thais Q., Feddema, Kim, Nekaris, K.A.I., Nijman, Vincent, 2021. Online trade in wildlife and the lack of response to COVID-19. Environ. Res. 193, 110439. February.

Olah, Chris, Mordvintsev, Alexander, Schubert, Ludwig, 2017. Feature visualization. Distill 2 (11). November.

Paszke, Adam, Gross, Sam, Massa, Francisco, Lerer, Adam, Bradbury, James, Chanan, Gregory, Killeen, Trevor, Lin, Zeming, Gimelshein, Natalia, Antiga, Luca, Desmaison, Alban, Kopf, Andreas, Yang, Edward, DeVito, Zachary, Raison, Martin, Tejani, Alykhan, Chilamkurthy, Sasank, Steiner, Benoit, Fang, Lu, Bai, Junjie, Chintala, Soumith, 2019. Pytorch: an imperative style, high-performance deep learning library. In: Advances in Neural Information Processing, 32. Curran Associates, Inc, pp. 8024–8035.

Patrignani, Andres, Ochsner, Tyson E., 2015. Canopeo: a powerful new tool for measuring fractional green canopy cover. Agron. J. 107 (6), 2312–2320.

Pirotta, Vanessa, Shen, Kaikai, Liu, Sheldon, Phan, Ha Tran Hong, O'Brien, Justine K., Meagher, Phoebe, Mitchell, Jessica, Willis, Joel, Morton, Ed, 2022. Detecting illegal wildlife trafficking via real time tomography 3d x-ray imaging and automated algorithms. Front. Conserv. Sci. 3. September.

Qing, Xu., Li, Jiawei, Cai, Mingxiang, Mackey, Tim K., 2019. Use of machine learning to detect wildlife product promotion and sales on twitter. Front. Big Data 2, 28.

Radosavljevic, Aleksandar, Anderson, Robert P., 2013. Making better maxent models of species distributions: complexity, overfitting and evaluation. J. Biogeogr. 41 (4), 629–643.

Saeidi, Marzieh, Kulkarni, Ritwik, Togia, Theodosia, Sama, Michele, 2017. The effect of negative sampling strategy on capturing semantic similarity in document embeddings. In: Proceedings of the 2nd workshop on semantic deep learning (SemDeep-2), pp. 1–8.

Shorten, Connor, Khoshgoftaar, Taghi M., 2019. A survey on image data augmentation for deep learning. J. Big Data 6 (1). Jul.

Simonyan, Karen, Zisserman, Andrew, 2014. Very Deep Convolutional Networks for Large-scale Image Recognition. arXiv preprint arXiv:1409.1556.

Simonyan, Karen, Vedaldi, Andrea, Zisserman, Andrew, 2014. Deep inside convolutional networks: visualising image classification models and saliency maps. In: Workshop at International Conference on Learning Representations.

Singh, Abhineet, Pietrasik, Marcin, Natha, Gabriell, Ghouaiel, Nehla, Brizel, Ken, Ray, Nilanjan, 2020. Animal detection in man-made environments. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). March.

Siriwat, Penthai, Nijman, Vincent, 2020. Wildlife trade shifts from brick-and-mortar markets to virtual marketplaces: a case study of birds of prey trade in Thailand. J. Asia-Pac. Biodivers. 13 (3), 454–461. September.

Stringham, Oliver C., Moncayo, Stephanie, Hill, Katherine G.W., Toomes, Adam, Mitchell, Lewis, Ross, Joshua V., Cassey, Phillip, 2021. Text classification to streamline online wildlife trade analyses. PLOS ONE 16 (7), e0254007. July.

Subramanian, Jyothi, Simon, Richard, 2013. Overfitting in prediction models – is it a problem only in high dimensions? Contemp. Clin. Trials 36 (2), 636–641. Nov.

Sung, Yik-Hei, Fong, Jonathan J., 2018. Assessing consumer trends and illegal activity by monitoring the online wildlife trade. Biol. Conserv. 227, 219–225. November.

Tabak, Michael A., Norouzzadeh, Mohammad S., Wolfson, David W., Sweeney, Steven J., Vercauteren, Kurt C., Snow, Nathan P., Halseth, Joseph M., Di Salvo, Paul A., Lewis, Jesse S., White, Michael D., Teton, Ben, Beasley, James C., Schlichting, Peter E., Boughton, Raoul K., Wight, Bethany, Newkirk, Eric S., Ivan, Jacob S., Odell, Eric A., Brook, Ryan K., Lukacs, Paul M., Moeller, Anna K., Mandeville, Elizabeth G., Clune, Jeff, Miller, Ryan S., 2018. Machine learning to classify animal species in camera trap images: applications in ecology. Methods Ecol. Evol. 10 (4), 585–590. Nov.