Identifying Illegal Wildlife Trade On E-commence Using Deep Learning and NLP Techniques

1. Introduction

Illegal wildlife trade (IWT) is a known long outstanding concern that needs to be addressed. It initially started from manual trafficking and later spread across to e-commerce websites, social media, as well as the dark web. The internet made access to illegal trades easier by providing connectivity to a wide variety of audiences and now these users access various e-commerce websites and get illegal items at their fingertips. Internet platforms aiding such activities have increased the already existing concerns.

The motivation behind researching illegal wildlife trade is that research in this area is relatively rare. The trade-in of wildlife can be both legal and illegal which can be influenced by a variety of factors, for example, the need for medicine to cure a disease is different compared to poaching. It is also a known fact that the legal trade is often overshadowed by crime on wildlife.

2. Background and related work

Any crime that involves the illegal trade, smuggling, poaching, or capture of endangered species, or protected wildlife is known as illegal wildlife trafficking. Some examples of illegal wildlife trade are the poaching of elephants for ivory and tigers for their skins and bones. However, countless other species are similarly overexploited, from marine turtles to timber trees.

Due to the desire for a wide range of goods, millions of plants and animals are captured from their natural habitats each year as whole animals or animal body parts, which are marketed as food, pets, plants, leather, and medicines including bushmeat, unusual pets, jewelry, and accessories like chess sets, furs for applications ranging from jackets to traditional costumes, and trophies. A concerningly big amount of this trade is illegal, endangering the lives of many species.

(1) Research in illegal wildlife trading with machine learning is rare. There are some papers on NLP techniques that are using the CN2 classifier with a specific if-then rule to identify illegal elephant ivory sales on eBay. This research does not address the wildlife trade in general.

(2) Another one on surveys, formatting different groups of people and asking them to complete the survey about images. They have applied statistics and people's decisions to compare experts, but it is not generalizing the IWT identification.

(3) Few papers are on designing the pipeline and framework to detect IWT particularly rhino species on Twitter but not on implementation.

(4) A paper on identifying animals via computer vision but not particularly on IWT.

3. Problem definition

Sometimes legal images can be mistaken as illegal which makes it harder to determine so, we need a concept advancement in this research area that can properly determine the illegal wildlife trade. Our goal is to identify illegal wildlife trade activities in e-commence platforms using Deep learning and NLP techniques.

4. Data and Methodology

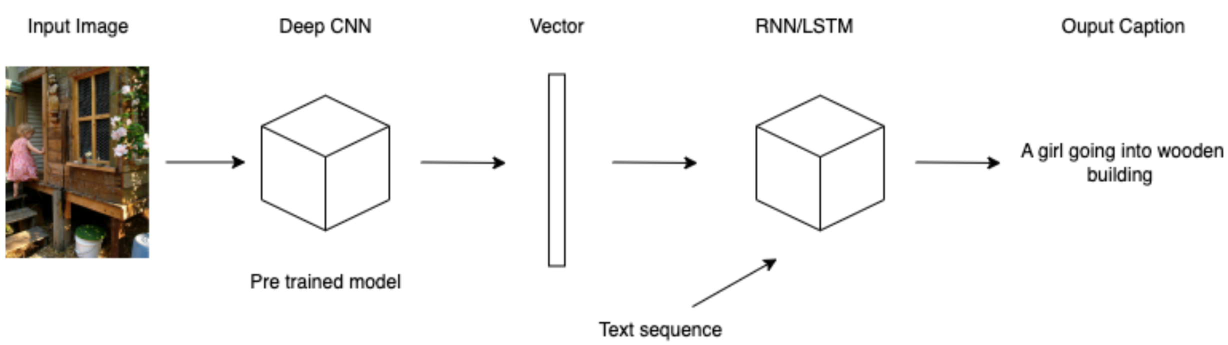
4.1 Data Collection

As data specific to Illegal wildlife trading is not readily available on the internet, collected data through web scraping. Many in python can scrape from websites like urllib, selenium, etc. From the selection, selenium is more flexible in interacting with buttons, URLs, and text on websites and can be used for fully automating the scraping process. For URLs and keywords related to wildlife trafficking taken from exports, written scraping scripts for each URL separately.

The script will search all keywords and automatically download the ad (advertisement) results related to the keywords. From every ad result script will download images, scrape image URLs, and information related to the ad like its title, price, location, details, and description of the item which are written into a text file. Manually filtered the ad results from web scrapping to avoid irrelevant search results.

4.2 Methodology

(5) Image captioning is a technique that will use both Deep Learning and NLP to determine the caption for a given image input.



The image will process through a Deep learning neural network which can be any pre-train model like VGG, or ResNet to get the feature vector and it is one of the inputs to RNN as shown in the above figure. The other input is the text sequence which might be words or a sequence of words. By taking these two inputs LSTM with L2 regularization will predict the next sequence of words (next word prediction) from the input sequence as shown below. Later from these sentences randomly chosen a word instead of the entire sentence and calculated accuracy instead of the BLEU score.

X2(text sequence)	Y(word to predict)
start	A
start A	girl
start A girl	into
start A girl into	wooden
start A girl into wooden	building
start A girl into wooden building	end

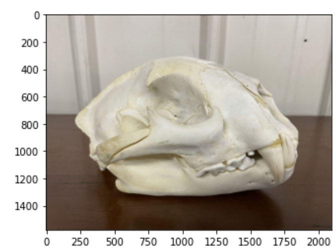
Another model was implemented which is Image Tagging a multi-label classification model that gives a word as a label by taking an image as input. Weights from the pre-trained model mobilenet_v2_100_224 are taken and updated those weights according to our data which is called transfer learning. Labels are the words taken from the captions (sentences) checking if the words belong to keywords of wildlife and randomly selecting from words that existed else the keyword is taken.

5. Results/Evaluation

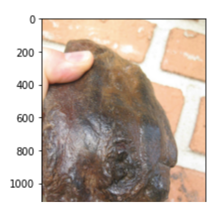
VGG model with entire sentences as captions trained for 20,30,50 epochs and below are the BLEU scores and loss respectively.

	Epochs - 20	Epochs - 30	Epochs - 50
BLEU-1	0.290381	0.326050	0.338046
BLEU-2	0.165062	0.203879	0.219267
Loss	0.8219	0.1945	0.0596

```
--------------------Actual----------------
startseq mountain lion skull endseq
--------------------Predicted-------------
startseq mountain mountain skull endseq
```

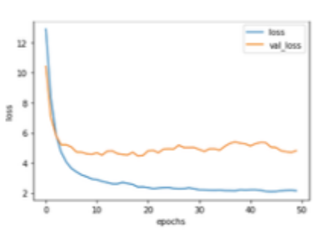


```
--------------------Actual--------------------
startseq fossil petrified mammoth molar tooth 14.4 cm endseq
--------------------Predicted-------------------
startseq petrified petrified molar molar 14 4 cm endseq
```

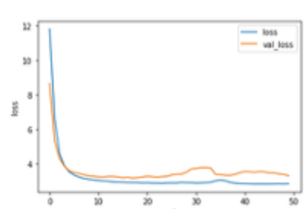


BLEU scores are low, and the predicted results as shown in images are not accurate as there is repetition of words.

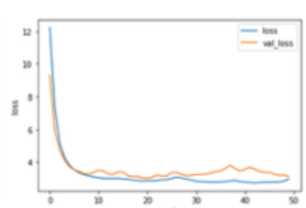
VGG model



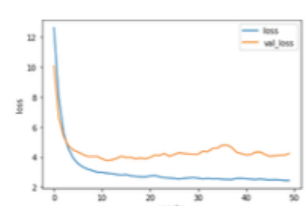
Xception



InceptionV3



ResNet50

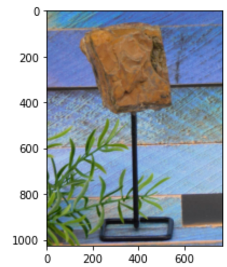


	VGG model	Xception	InceptionV3	ResNet50
test	0.22727272727272727	0.005050505050505051	0.0	0.15151515151515152
train	0.806070826306914	0.3243395165823496	0.26194491287240024	0.5930297920179877

Trained different model as above with L2 regularization = 0.01, epoch = 50, batch size= 64 and early stopping. These models are trained on randomly selected a word instead of entire sentence. Calculated accuracy instead of bleu score. Accuracy = number of predicted words in actual/ length of predictions.

```
---------------------Actual---------------------
alligator purse
--------------------Predicted--------------------
startseq purse endseq
```

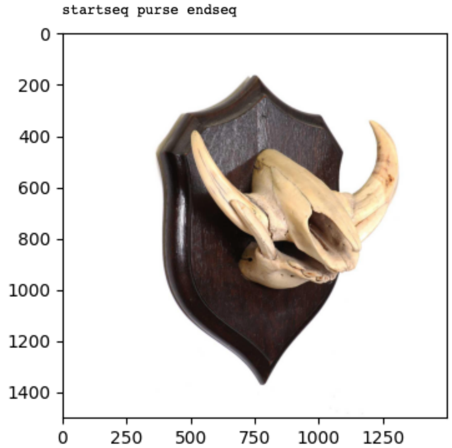
```
---------------------Actual---------------------
tiger eye display rock stone lapidary specimen
--------------------Predicted--------------------
startseq lapidary endseq
```

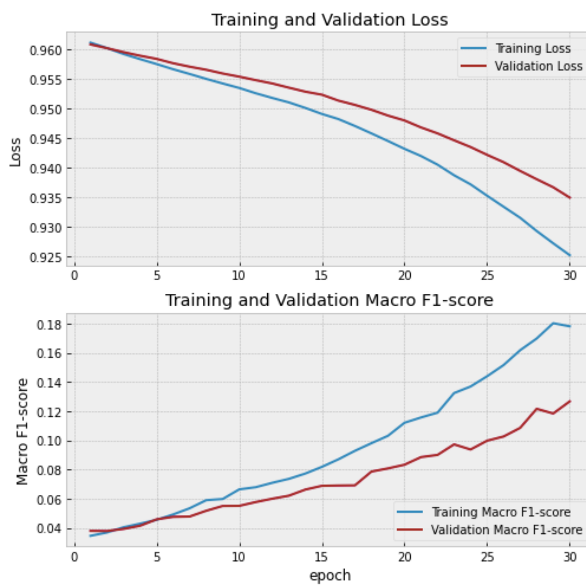


The test accuracy is low and even train accuracy. VGG model performed better compared to other models. The predicted results from VGG as shown in images is picking the words form captions

Multilabel classification was trained for 30 epochs, got macro f1 score is 0.13 which is the average F1 score of every class label as below.

```
---------------------Actual---------------------
alligator purse
--------------------Predicted--------------------
startseq purse endseq
```






```
actual
[list(['lizard'])]

Prediction
['birds', 'bone', 'egg', 'elephant', 'feather', 'feet', 'hair', 'leopard', 'lizard', 'paw', 'skeleton', 'skin', 'skul
l', 'tooth']
```

The Macro F1 score is very low and predicted results in images are picking the words but these results are with threshold above 0.5 of probability.

```
actual
[list(['taxidermy'])]

Prediction
['bird', 'birds', 'bone', 'egg', 'feather', 'feet', 'hair', 'leopard', 'lion', 'lizard', 'lynx', 'skeleton', 'skull',
'taxidermy', 'tooth']
```

6. Discussion

Train VGG models on the data that was processed for multilabel classification as before the words are randomly taken instead of matching with keywords of wildlife. As accuracy is low, we must try different hyperparameter tuning and adding more data for CNNs would help and we must check if there is any class imbalance for the multilabel classification model as the average F1 is low.

References

1. Hernandez-Castro and Roberts (2015), Automatic detection of potentially illegal online sales of elephant ivory via data mining. PeerJ Comput. Sci. 1:e10; DOI 10.7717/peerj-cs.10
2. Austen GE, Bindemann M, Griffiths RA, Roberts DL. 2018. Species identification by conservation practitioners using online images: accuracy and agreement between experts. PeerJ 6:e4157 https://doi.org/10.7717/peerj.4157
3. Conservation Biology, Volume 33, No. 1, 210–213 C 2019 The Authors. Conservation Biology published by Wiley Periodicals, Inc. on behalf of Society for Conservation Biology. DOI: 10.1111/cobi.131045
4. Parham, J., Stewart, C., Crall, J., Rubenstein, D., Holmberg, J., and Berger-Wolf, T. (2018). “An animal detection pipeline for identification,” in 2018 IEEE Winter Conference on Applications of Computer Vision (WACV) (Lake Tahoe, NV: IEEE), 1075–1083.
5. S. Amirian, K. Rasheed, T. R. Taha and H. R. Arabnia, "Automatic Image and Video Caption Generation With Deep Learning: A Concise Review and Algorithmic Overlap," in *IEEE Access*, vol. 8, pp. 218386-218400, 2020, doi: 10.1109/ACCESS.2020.3042484