Wildlife: Exploring Wildlife Trafficking through Animal-Related Ads

Jin Zhou
New York University
New York, USA
jz3928@nyu.edu

Junzhe Zhou
New York University
New York, USA
jz3709@nyu.edu

Chhatrapathi Sivaji Lakkimsetty
New York University
New York, USA
cl7203@nyu.edu

I. Introduction

The illegal wildlife trade (IWT) poses significant threats to biodiversity, ecosystems, and human health. To combat this illegal activity, the recent researches propose diverse set of methods ranging from X-ray imaging to genetic analysis.

In "Detecting illegal wildlife trafficking via real time tomography 3D X-ray imaging and automated algorithms," 3D X-ray CT technology is used together with machine learning to develop an algorithm that is capable of identifying concealed wildlife within luggage and cargo with high accuracy. The researchers utilized 294 scans from 13 rare species to train the model, and they are able to achieve a detection rate of 82% with a 1.6%false positive rate. [1], [2]

The focus shifts to the impact of IWT on spreading infectious diseases in both nature and human habitats in research paper "Illegal Wildlife Trade and Emerging Infectious Diseases:Pervasive Impacts to Species, Ecosystems and Human Health". The authors conduct a comprehensive literature review on 82 papers from 1990 to 2020 in order to fully understand the link between IWT and certain types of pathogens. [3]

The study "A Survey on Identification of Illegal Wildlife Trade" highlights the unawareness of the general public about illegal animal trafficking by explaining several ways of trade that is happening on the web. The author Nalluri first explains how trades are conducted on the "clear web". On legitimate platforms like Ebay, advertisements regarding animal trade are being sent to potential interested buyers via recommendation systems on a weekly basis. Since these ads exists for only a short period of time, it is difficult to track them back to their source. [4] The other lesser known space is dark web. A P2P network and a Tor browser are tools for cybercriminals to mask their identities and conduct illegal businesses. This is why investigating crime on the dark web is even more difficult. Later in the paper, author recommends the use of deep learning to identify both sellers and buyers. More specifically, researchers can use APIs to collect large dataset from social media platform like twitter. Then, a deep learning model can help identify visual, verbal, and audiovisual content regarding IWT. [4], [5]

II. Problem Formulation

The main issue we are addressing in this project is identifying online advertisements for wild animal trading to potentially reduce the problems mentioned above. To achieve this, Machine Learning will be utilized to classify information from images and texts. This problem can be divided into three smaller tasks: data gathering, image analysis, and text analysis.

- For data gathering, advertisements need to be collected from multiple trading sites such as eBay. Then, the collected data will be cleaned with weak supervision, for example, remove the ads that did not provide an image or text description, to avoid bad data and reduce the accuracy of the model.
- Then, the images provided need to be processed into relative text or numerical information that can be analyzed. This step can be completed either separately to provide more inputs for a model with a simpler structure, or passed in along with other data but have to be evaluated differently.
- Finally, the text data such as description and location needs to be analyzed alongside either the modified image data or the original image to form a final classification of the advertisement, with the final output being either whether it is wild animal trading or the type of product the advertisement is trying to promote.

Using the final result from this three-step processing, we will be able to distinguish whether an advertisement is attempting to trade wild life.

III. Related Work

There have been multiple efforts in image-to-text classification using machine learning. The paper "Deep Learning for Image-to-Text Generation: A Technical Overview" by Xiaodong He highlights the difficulty of detecting and understanding the relationships between various elements within an image. He et al. proposes two solutions that enhance video captioning. [6]First one is an end-to-end framework, which is also called vector sequence learning. Basically, the encoder processes the image to generate a global visual feature vector using Convolutional Neural Networks(CNNs) and the decoder will use that output to generate a caption using Recurrent Neural Networks(RNNs). The second method is the attention mechanism. This allows the model to focus on specific parts of an image rather than the entire image. By dynamically focusing on different subregions of an image, the caption generated will evolve over time. To evaluate their

results, they used both automatic metrics and human studies. A quantifiable metric would be the fraction of n-grams between the hypothesis and the inference, which we can also consider using in our research. [6]

Another novel approach uses Maximum Mean Discrepancy (MMD) to help in the image-to-text synthesis process, rather than relying solely on the traditional Generative Adversarial Networks (GANs). Das et al. created two separate autoencoders, where one is for texts and the other is for images. [7]For the image autoencoder, ResNet50 is used to first transform images into high-dimentional vectors. For the text autoencoder, the authors employ a pre-trained LSTM on the One Billion Word Benchmark dataset. For the translation between embedding spaces of the two modalities, the authors compare the performance of a MMD-based mapping network and a GAN-base mapping network. This study offers an advanced multi-modal learning technique that outperforms the traditional image-to-text frameworks. [7]

More recently, GPT-4Vision has been known to benchmarking image-to-text taks using Large Language Models(LLMs). It is capable of accurately extracting both text and visual features from user input images. In further details, similar with aforementioned work, the features are extracted and mapped to a semantic space. [8] Then, the data is fed into an LLM like GPT so that text generation can begin. It is potentially useful to include the use of LLM-based models to compare the performance of different classifiers that we build.

IV. Data Insights

Numerous sources offer advertisements related to wildlife, spanning a broad spectrum of wildlife products, both directly and indirectly connected. However, sifting through the vast internet landscape to gather this information is not computationally efficient for this project. Therefore, our research narrows its focus to a single prominent online platform, eBay.com, as the primary source for data collection. The ads found on eBay.com provide a rich dataset, notably abundant in listings directly linked to animal products. Our dataset comprises various elements such as image URLs, descriptions, titles, prices, domains, locations, and more from these advertisements. To effectively comprehend and utilize this data, it's essential to analyze and process both the image and textual information. This approach transforms raw data into structured, meaningful insights that accurately reflect the characteristics of the advertisements.

Table 1: Data overview: attributes and their descriptions, and the number of records that contain the attributes

Attributes	Description	# of records
url	The ad URL	954,684
title	Product Advisement title	946,732
text	The page text	954,684
product	Name of the product	954,684
description	Description of the product	805,449
domain	Website where the product is posted	954,684
image	URL of the image	787,185
retrieved	time when the page was downloaded	954,684
category	The category listed for that product	25,038
production date	Production date of the product	5,786
price	Price of the product	682,652
currency	Currency of the price	679,717
seller	Seller name	8,910
seller_type	the category the seller is listed	27,483
location	Location of product	25,150
zero_shot_label	zero shot classifier results	954,684
zero_shot_prob	zero shot label probability	954,684
id	UUID used as filename for images	954,684

V. Methods, Architecture, and Design

This project is architected around two critical pipelines, pivotal to its comprehensive development and eventual success. Below we elucidate the organization and operational specifics of these pipelines.

A. Data Collection Pipeline

In the extensive digital ecosystem, advertisements for wildlife products are prevalent, with both direct and indirect relevance. However, the sheer scale of online data poses computational challenges for our project. Consequently, we concentrate our data gathering efforts on eBay.com, a principal source. This pipeline aims to methodically harvest product listings from eBay and analogous platforms, thereby generating an 'animal_products' dataset. This dataset includes essential attributes such as Image URL, Product Sales, Seller Information, Location, Text Descriptions, and Titles, all of which are instrumental in furnishing deep insights. Utilizing a Jupyter notebook service, the 'Minio_data' pipeline provides a shared data storage solution, facilitating the dissemination of sizable datasets to users as a read-only resource within their notebook environments.

B. Data Inference Pipeline

Crucial to the project, this pipeline focuses on the processing, cleansing, transforming, and analytical examination of the raw data to extract valuable insights. It commences with the purification of raw data, a prerequisite for detailed analysis. The Image URL field, rich in visual data, necessitates transformation to unlock its information potential. This pipeline incorporates several key procedures:

C. Feature Transformations

- **Image to Text (ResNET, Attention Layers):** A significant correlation exists between raw image data and text descriptions, necessitating preprocessing via ResNET and attention layers. These techniques are indispensable for converting images and text into actionable insights, significantly boosting the classifier's performance. An application of this method is Visual Question Answering (VQA), where models generate responses based on both textual and visual inputs. [9]
- **Text Transformation (Transformers, LLMs):** Transformations employing Transformers revolutionize language comprehension and the integration of LLMs, enhancing the semantic processing of text descriptions and categorical features.

D. Data Cleaning

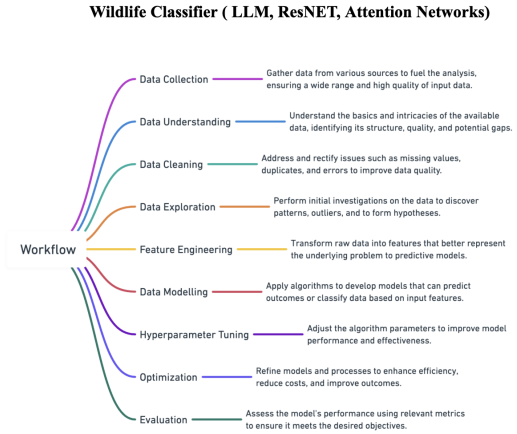
Identified as both laborious and crucial, this process tackles duplicates, missing values, and outliers to ensure data integrity and applicability.

E. Advanced Feature Transformations

Leveraging SparkMLib, this strategy emphasizes creating new features from existing data based on product location, category, and seller information to improve predictive accuracy concerning specific variables.

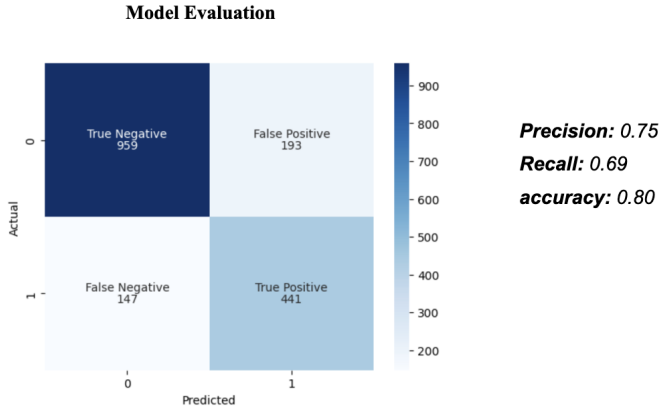
F. Workflow

Each pipeline is intricately constructed to maximize the utility of the collected data, ensuring its precise processing and analysis to fulfill the project's goals.



VI. Results

The existing codebase has Data collection and Inference pipelines, with some default parameters and architecture, which we leveraged to get insights from the data. As of now, we only have the results from the previous team that worked on this project. Using these pipelines, we achieved these following metrics on the testing set :



Acknowledgment

We thank Professor Juliana Freire for giving us access to the existing dataset and codebase on NYU JupyterHub.

References

[1] V. Pirotta, K. Shen, S. Liu, H. T. H. Phan, J. K. O'Brien, P. Meagher, J. Mitchell, J. Willis, and E. Morton, "Detecting illegal wildlife trafficking via real time tomography 3d x-ray imaging and automated algorithms," *Frontiers in Conservation Science*, vol. 3, 2022. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fcosc.2022.757950

[2] G. K. Moloney and A.-L. Chaber, "Where are you hiding the pangolins? screening tools to detect illicit contraband at international borders and their adaptability for illegal wildlife trafficking," *PLOS ONE*, vol. 19, pp. 1–25, 04 2024. [Online]. Available: https://doi.org/10.1371/journal.pone.0299152

[3] E. R. Rush, E. Dale, and A. A. Aguirre, "Illegal wildlife trade and emerging infectious diseases: Pervasive impacts to species, ecosystems and human health," *Animals*, vol. 11, no. 6, 2021. [Online]. Available: https://www.mdpi.com/2076-2615/11/6/1821

[4] S. Nalluri, S. J. R. Kumar, M. Soni, S. Moin, and K. Nikhil, "A survey on identification of illegal wildlife trade," in *Proceedings of International Conference on Advances in Computer Engineering and Communication Systems*, C. Kiran Mai, B. V. Kiranmayee, M. N. Favorskaya, S. Chandra Satapathy, and K. S. Raju, Eds. Singapore: Springer Singapore, 2021, pp. 127–135.

[5] E. Di Minin, C. Fink, H. Tenkanen, and T. Hiippala, "Machine learning for tracking illegal wildlife trade on social media," *Nature Ecology & Evolution*, vol. 2, no. 3, pp. 406–407, 2018. [Online]. Available: https://doi.org/10.1038/s41559-018-0466-x

[6] X. He and L. Deng, "Deep learning for image-to-text generation: A technical overview," *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 109–116, 2017.

[7] A. S. Das and S. Saha, "Self-supervised image-to-text and text-to-image synthesis," *CoRR*, vol. abs/2112.04928, 2021. [Online]. Available: https://arxiv.org/abs/2112.04928

[8] X. Zhang, Y. Lu, W. Wang, A. Yan, J. Yan, L. Qin, H. Wang, X. Yan, W. Y. Wang, and L. R. Petzold, "Gpt-4v(ision) as a generalist evaluator for vision-language tasks," 2023.

[9] F. Phe, "Paying attention to text and images for visual question answering," https://blog.dataiku.com/paying-attention-to-text-and-images-for-visual-question-answering, 2023.