**Injured animal Detection and Alert generation system**

By: Akash Jangid

Project Title:

**Recognition, Monitoring, and Alert via Camera Trap Images**

**:Surveillance with Deep Learning**

## Base Paper/Reference Paper:

##### **Automated Recognition of Wild Animal Species in Camera Trap Images Using Deep Learning Models**

**Citation:**

R. Thangaraj, S. Rajendar, S. M, R. S. K, S. Sasikumar and C. L, "Automated Recognition of Wild Animal Species in Camera Trap Images Using Deep Learning Models," 2023 Third International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT), Bhilai, India, 2023, pp. 1-5, doi: 10.1109/ICAECT57570.2023.10117922. keywords: {Deep learning;Image recognition;Biological system modeling;Wildlife;Ecosystems;Transfer learning;Cameras;Animal species recognition;Convolution neural network;Inception V3;camera trap;transfer learning;fine-tuning;Data Augmentation},

**Venue (Conference/Journal/Transaction) & other details:**

Published in: [2023 Third International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT)](https://ieeexplore.ieee.org/xpl/conhome/10117587/proceeding)

Date of Conference: 05-06 January 2023

Publisher: IEEE

Conference Location: Bhilai, India

##### **Enhanced Multimodal Understanding: Integrating CNN and LSTM for Advanced Image Semantic Description**

**Citation:**

H. Lei, "Enhanced Multimodal Understanding: Integrating CNN and LSTM for Advanced Image Semantic Description," 2023 International Conference on Applied Physics and Computing (ICAPC), Ottawa, ON, Canada, 2023, pp. 1-5, doi: 10.1109/ICAPC61546.2023.00006. keywords: {Deep learning;Visualization;Adaptation models;Semantics;Feature extraction;Convolutional neural networks;Artificial intelligence;Convolutional Neural Networks;Long Short-Term Memory networks;deep understanding},

**Venue (Conference/Journal/Transaction) & other details:**

Published in: [2023 International Conference on Applied Physics and Computing (ICAPC)](https://ieeexplore.ieee.org/xpl/conhome/10497884/proceeding)

Date of Conference: 27-29 December 2023

Publisher: IEEE

Conference Location: Ottawa, ON, Canada

1. [UniRel: Unified Representation and Interaction for Joint Relational Triple Extraction](https://github.com/wtangdev/unirel)

## Abstract:

With the evolution of Animal monitoring technologies such as radio tracking, Wireless Sensor Network (WSN), Satellite and Global Positioning System (GPS); motion sensor cameras called camera traps (CTs), have become a pivotal tool due to their ability to capture images without human intervention. While CTs are effective to some extent, the automatic identification of wild animals in CT images remains challenging due to variations in the image properties.

1. Large volume of images capturing animals in various postures
2. Instances where no animals were present
3. Animals appearing very far away, beyond human eyesight
4. Images with extreme close-up shots
5. Images containing lens flare or captured under diverse lighting and weather conditions

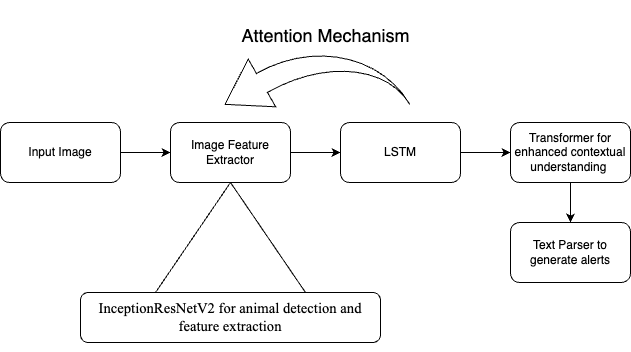
Manual identification of animal species from CT images is challenging due to these above factors and therefore, automating this process is crucial to efficiently monitor wildlife.

Learning models such as DeneseNet169, InceptionResNetV2, ResNetV2-50, and Xception for the automatic identification and recognition of individual animals in different patterns. The effectiveness of the models is measured using precision, recall, and F1 score indicators. Finally, a description of useful images is generated using LSTM, text is parsed and timely alerts are generated in emergency situations.

## Contribution of the Papers :

1. [[1]](#_8dhbg2q9xtog) introduces and evaluates various deep learning models such as DenseNet169, InceptionResNetV2, ResNetV2-50, and Xception for the automatic identification and recognition of individual animals in different patterns. It compares the performance of different deep learning models based on various metrics and provides insightful analysis, identifying InceptionResNetV2 as the top-performing model with an accuracy of 94.82%.
2. The paper also, through experimental validation, conducts experiments using real-world data obtained from the Open Dataset V6+, demonstrating the practical applicability of the proposed models in wild animal identification.
3. [[1]](#_8dhbg2q9xtog) outlines a transfer learning approach for improving animal species recognition using pre-trained CNN models, incorporating techniques such as deep feature extraction, convolution layers, ReLU activation functions, max-pooling layers, fine-tuning with SoftMax classifiers, and provides comprehensive details on dataset distribution, network parameters, and training protocols to support reproducibility and future exploration in the domain.
4. [[2]](#_wh4sqyeqi4ty) on the other hand provides an innovative framework merging Convolutional Neural Networks (CNN) with Long Short-Term Memory networks (LSTM) to enhance image semantic description, emphasizing deep understanding over mere recognition. It is leveraging CNNs for detailed visual feature extraction and LSTM for sequential data processing to generate accurate and contextually rich textual descriptions.
5. Evaluation using BLEU and ROUGE metrics for rigorous quantitative assessment of linguistic quality and contextual alignment with reference descriptions.

## Proposed Novelty (Yours):

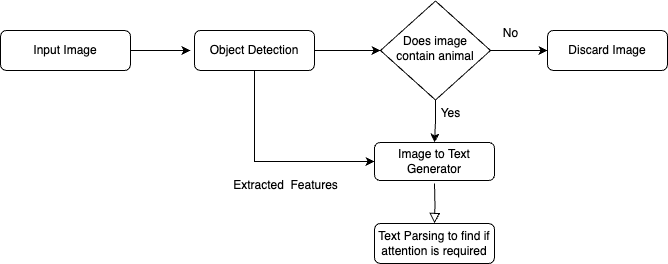


Proposed Network

We are proposing a network that combines aspects from two research papers. From [[1],](#_8dhbg2q9xtog) we found that InceptionResNetV2 works well with camera trap images to identify wildlife, helping us segregate useful images from trash. [[2]](#_wh4sqyeqi4ty) proposed a network using LSTM and attention mechanisms to generate image descriptions. We will parse this text for alarming words like “human”, "blood", "injury", and "dead" to trigger an alarm, creating an end-to-end wildlife monitoring application.

**Benefits:**

* Leverages strengths of both image recognition and NLP.
* Provides a more robust solution compared to relying solely on one source of information.



Proposed Flow of Network

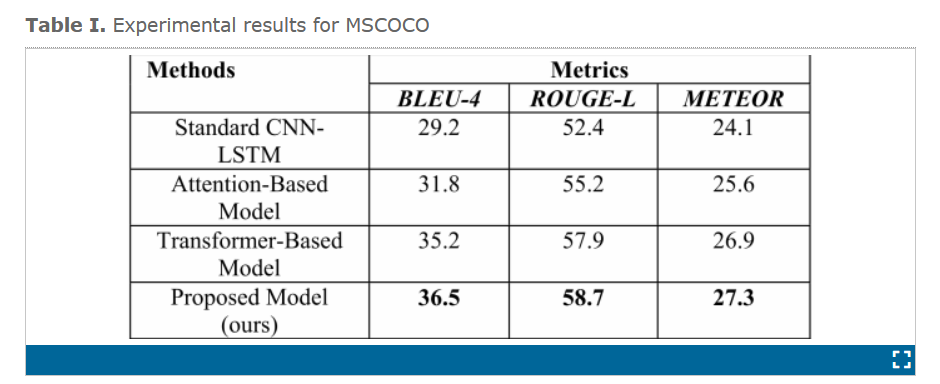
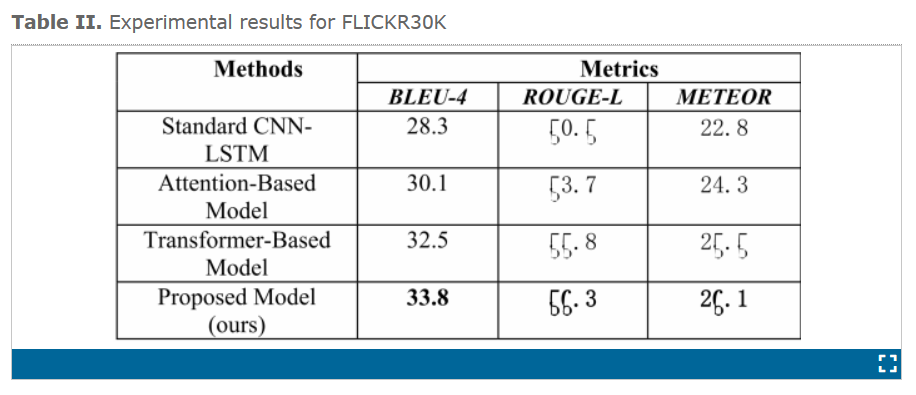
## Results:

The below table compares the metric values of the different models for wild animal classification. Finally, based on the table the results illustrate that InceptionResNetV2 outperforms compared to other models with an accuracy of 94.82%.The precision, recall, and F1 score metrics further underscore the superiority of InceptionResNetV2, with values consistently at 0.94 across these measures. This remarkable performance highlights the model's ability to effectively classify wild animals with high precision and recall, striking a balance between minimizing false positives and false negatives.

While other models such as DenseNet 169, ResNet V2-50, and Xception also demonstrate respectable performance, their metrics fall short of matching InceptionResNetV2's exceptional accuracy.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Models | Accuracy % | Precision | Recall | F1 Score |
| DenseNet 169 | 88.38 | 0.89 | 0.88 | 0.88 |
| Inception ResNet V2 | 94.82 | 0.94 | 0.94 | 0.94 |
| ResNet V2-50 | 87.99 | 0.88 | 0.87 | 0.88 |
| Xception | 89.26 | 0.89 | 0.89 | 0.89 |

Comparison of Performance Metrics

From the second paper, in continuation of the image to text using LSTM with different datasets- 

the obtained results align with expectations.

**Challenges:**

* Requires a larger dataset for training, including images with corresponding descriptions that mention the animal and its state.
* Implementing the attention mechanism can add complexity to the model.

Despite this, they hold promise for deeper exploration, particularly in developing an optimized alert system based on captioned text derived from camera trap images. The results for our dataset are to be anticipated but we are hopeful that it will be insightful for further investigation towards an optimal way to make an alert system and informed conservation efforts through captioned text based on the camera trap images by the CT cameras.