tiger3

by Tejalal Choudhary

Submission date: 18-Jul-2019 12:38PM (UTC+0530)

Submission ID: 1152862858

File name: AMUR_TIGER_DETECTION_without_references.pdf (1.03M)

Word count: 1738

Character count: 9126

AMUR TIGER DETECTION USING DEEP LEARNING

Shankho Ghosh, Ketan Muddalkar, Ajay Patel, Vishnu B Nair

Balmukund Mishra, Divya Acharya, Tejalal Choudhary

I. ABSTRACT

In our ecosystem, wildlife plays a key role in sustaining different natural processes in nature. They are as important as humans. So, protection and conservation of wildlife becomes vital, especially those which are at risk of being extinct. One such species is the Amur tiger, which is categorised as endangered. Traditional methods for detecting them have time and again proved insufficient. Therefore, in our project we attempt to provide a more efficient and reliable method for detection of this endangered species using deep learning techniques. To facilitate our project, we are provided with ATRW(Amur Tiger Re-identification in the wild) datasets by CVWC(Computer Vision for Wildlife Conservation which contains 2485 and 277 photos(along with their annotations) for training and validation respectively. We deploy three different object detection model for training and select the best one for its implementation.

II. INTRODUCTION

Wildlife is a national resource which helps in maintaining ecological balance. Wildlife is a key part of our ecosystem, and disturbing it would cause imbalance in nature. So, it becomes imperative that we conserve all them.

When speaking of wildlife conservation, the issue of the decrease in the number of Amur tiger calls for some attention. Amur tiger is categorised as extremely endangered species due to an alarming rate of decrease in their numbers. Poaching and habitat loss has led to reduce their number to mere 500. Several rules and regulations have been introduced to curb this menace and several organizations have taken initiatives to save them. Traditionally, methods of attaching transmitters to wildlife has been used. But there are many problems associated with this like image classification, image localization sensor failures, difficulty in scaling large populations and so on.

Deep learning techniques for object detection may be used to provide more effective ways of identifying and monitoring the number of Amur tigers, especially coupling it with unmanned aerial vehicles like drones. But using deep learning comes with its own challenges-detecting an object from a cluster of objects that may be overlapped becomes difficult, while detection of objects presented individually is lot easier. Also, wildlife data contains a large number pose variations, complex backgrounds etc which makes object detection more complex process. Current datasets lack systematic benchmark protocols. This makes for requiring new datasets and establishing appropriate benchmark protocols for the same.

III. RELATED WORKS

Object detection is extensively used for identification of person. There have been various types of projects made based on person detection and reidentification. Projects such as human detection in forests[1], Also, there has been work on identifying various diseases in various person which also based on object detection based on persons. Apart from person re identification there has also been extensive works on vehicle detection such as identifying specific vehicle, collision avoidance system and so on.

Based on wildlife, there has been work on identifying animals such as tiger detection, elephant detection etc. Such projects includes automatically identifying, counting, and describing wild animals in captured in cameras[2], a practical animal detection and collision avoidance system[3], animal head detection[4]. But, because of smaller data size and limited annotations associated with it very limited researches are done on these datasets. For overcoming these limitations, we use ATRW datasets which are of large data size and proper annotations. There has also been a project based on detection and re-identification of Amur tiger based on their pose[5].

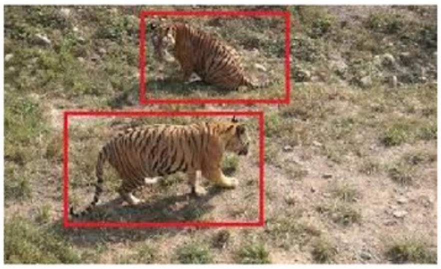


Fig.1 Tigers Detected

In our project we use SSD lite mobilenet v2, SSD inception v2 and faster R-CNN inception for tiger detection. SSD (single shot multibox detector) and SSD Lite is mostly used for real time object detection while mobilenet v2 is used for improving performance of mobile models. Using this model, we detect tiger with less time and high accuracy. Faster R-CNN is used less for real time object Detection. With Regional Proposal Network it take very less time for detection than RCNN and Fast RCNN. We use inception v2 with our model that increase accuracy and reduce the computation complexity.

IV. METHODOLOGY

1. PRE PROCESSING

As stated earlier, the ATRW dataset has images in jpeg format and annotations in XML format. This project uses tensorflow to train three model, that is, Faster R-CNN inception v2, SSD inception v2 and SSDLite mobilenet. In order to use tensorflow we need to have a tf (tensorflow) record file format, which is tensorflow's custom binary storage format. For this, the annotations which were in XML format was converted into CSV(Comma Separated Values) file format using Python. Then, this CSV file was subsequently converted to tf record, also using Python.

2. TRAINING

Now that all the preprocessing is done, we are all set for the next step ,that is , training. For this project, we decided to train three models which were Faster RCNN inception v2, SSD inception v2 and SSDLite mobilenet.

a. Faster RCNN inception v2

Faster R-CNN contains two networks: first, is Region Proposal Network (RPN) which is responsible for generating regions and second, a classifier which classifies the regions generated by RPN. The RPN is also responsible for predicting the possibility of the anchor being in background or foreground. The work of CNN model is to drag a sliding-window across multiple possible regions which are generated based on *k* fixed-ratio anchor boxes. Each region proposal contains an objectness score(a.k.a Confidence Score) and 4 coordinates representing the bounding box of that particular region. Now these region proposals are fed straight into Fast R-CNN. Pooling layers, a bunch of fully-connected layers, and later a softmax classification layer and bounding box regressor is additionally added. The biggest edge Faster R-CNN(refer Fig. 2 and Fig. 3) has over other models is that it is much faster than its predecessors i.e. R-CNN and Fast R-CNN[8]. Also detection accuracy is increased. On the downside, there is a small disadvantage as far as accuracy is concerned real time speed is required. Also, the architecture of Faster RCNN is much more complex.

Fig.2 Faster RCNN architecture[9]

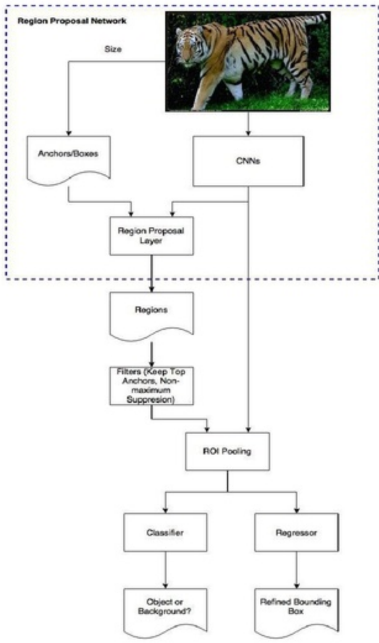


Fig.3 Faster RCNN Flowchart[10]

b. SSD inception v2

SSD stands for Single Shot Detector which does detection and classification in a single shot. The major idea behind SSD is the use of information provided by deep layers and shallow layers. The information extracted by deep layers can be used to classify bigger objects and information extracted by shallow layers can be used to classify small objects. To do this, we pass the image through a sequence of convolutional layers, with several sets of pre-defined feature maps at different scales. A 3x3 convolutional filter is used to assess a small set of default bounding boxes for each location in each of these feature maps. These default bounding boxes are equivalent to Faster R-CNN's anchor boxes. For each anchor box, the bounding box offset and their class probabilities are predicted simultaneously. During training, the ground truth box is matched with these predicted boxes based on IoU value. The model returns a positive value if the IoU value is > 0.5. This value is can be tweaked to achieve better localisation results. The main advantage of SSD is that it is a much faster alternative than Faster R-CNN as it does detection and classification in one single step. On the downside, SSD classifies and draws bounding boxes from every single position in the image, using multiple unique shapes, at several different scales. As a result, a significant amount of bounding boxes are generated.

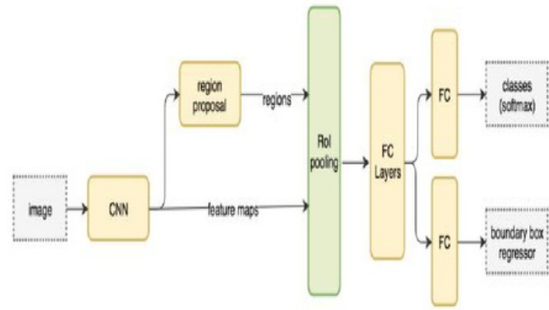


Fig.4 SSD Workflow[11]

c. SSDLite Mobilenet

SSDLite is basically a variant of SSD. Here, employs Mobilente v2 as the backbone and has depth wise separable convolutions for the SSD layers.The main advantage of using SSDlite is that it is much faster than the SSD itself and need less computing resources. But, it comes at a cost, accuracy. Fig.5 and Fig. 6 give us a clear idea about the architecture of this model.

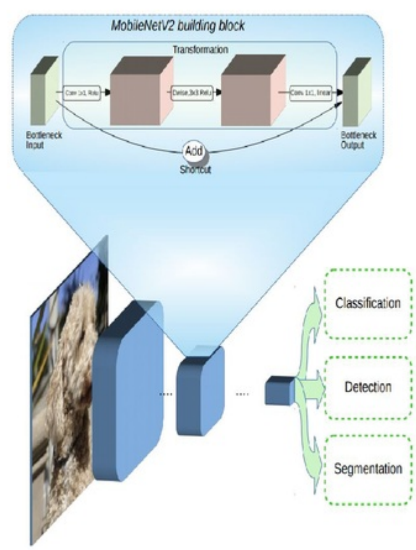


Fig.5 Mobilenet v2 Building Block

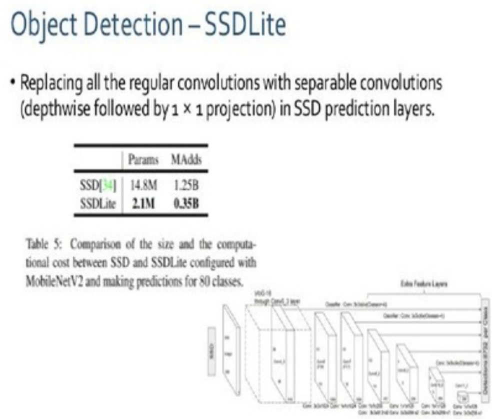
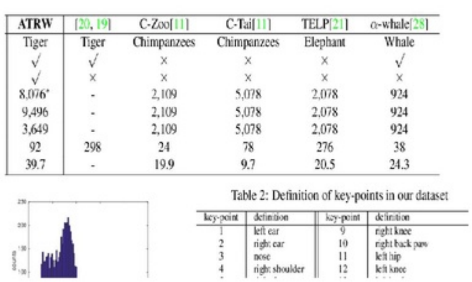


Fig.6 SSD Lite using Mobilenet v2 [13]

V. EXPERIMENTAL RESULTS

1. THE DATASET

This project uses ATRW datasets which is a large scale dataset where the images are shot in unconstrained environment. The dataset contains high resolution images (1920x1080) which have 2485 images along with their annotations for training and 277 images along with their annotations for validation. All the images in the dataset are in jpeg format and the annotations are in XML format. Fig.2 shows the comparison between ATRW dataset and other dataset formats.



ATRW	[20, 19]	C-Zoo[11]	C-Tai[11]	TELP[21]	α-whale[28]
Tiger	Tiger	Chimpanzees	Chimpanzees	Elephant	Whale
√	√	×	×	×	√
√	×	×	×	×	×
8,076*	-	2,109	5,078	2,078	924
9,496	-	2,109	5,078	2,078	924
3,649	-	2,109	5,078	2,078	924
92	298	24	78	276	38
39.7	-	19.9	9.7	20.5	24.3

Table 2: Definition of key-points in our dataset

key-point	definition	key-point	definition
1	left ear	9	right knee
2	right ear	10	right back paw
3	nose	11	left hip
4	right shoulder	12	left knee

Fig.7 ATRW dataset released by CVWC[3]

2. RESULTS

Faster RCNN, SSD, SSDlite mobilenet were the models used for training. When the dataset was trained on Faster RCNN model over 94% accuracy was observed (Fig.8). While using SSD, an accuracy of over 90% was observed (Fig.9) and when SSDlite was trained over 95% accuracy was observed (Fig.10). The complete results of all the three models that were trained are given in the following table.

Table 1. Comparison of results of all three models

	Faster R-CNN inception v2	SSD inception v2	SSDlite mobilenet
mAP @IoU 0.5	0.9437	0.9055	0.9554
Steps	2,00,000	1,00,000	1,30,000
Loss (classification)	0.123126	5.8326	3.052575
Loss (localization)	0.112125	0.662244	0.442702

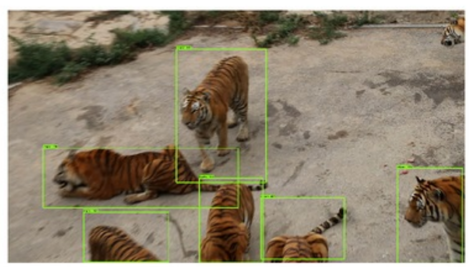


Fig.8 Faster R-CNN results

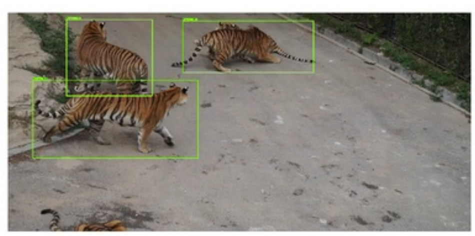


Fig.9 SSD results

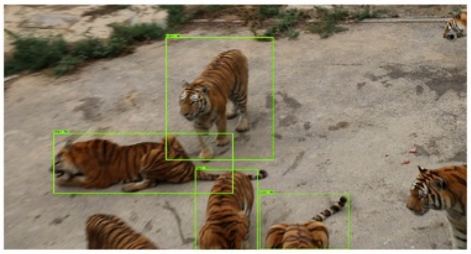


Fig.10 SSDLite results

From the above observations, we can divide the applications of these models based on 2 scenarios.

First, where computation power is available and low latency is required, Faster RCNN inception v2 is best suited. Second, for scenarios with low computation power and high latency SSDlite mobilenet is recommended.

VI. CONCLUSION

The aim of the project was to detect Amur tiger using deep learning which is an endangered species. We were provided with ATRW datasets which contained high resolution images for training as well as validation. Using this dataset, we trained three different models-Faster RCNN, SSD and SSDLite. We observed that for high latency scenario we can use SSDLite mobilenet and for low latency scenario Faster RCNN is to be used. Thus, we provided a novel method to detect Amur tigers using deep learning techniques for its conservation. This project can be further progressed by training on tiger pose detection and Re-ID datasets by CVWC and deploying it on mobile computers like NVIDIA Jetson Nano extending upto drones.

tiger3

ORIGINALITY REPORT

10%	0%	2%	10%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to University of Wales Institute, Cardiff Student Paper	6%
2	Submitted to University of Greenwich Student Paper	1%
3	Qing Tang, Kang-Hyun Jo. "Analysis of Various Traffic Sign Detectors Based on Deep Convolution Network", 2019 IEEE/SICE International Symposium on System Integration (SII), 2019 Publication	1%
4	Submitted to Higher Education Commission Pakistan Student Paper	1%
5	Submitted to CSU, San Jose State University Student Paper	1%
6	Submitted to University of Sheffield Student Paper	1%
7	Álvaro Arcos-García, Juan A. Álvarez-García,	<1%

Luis M. Soria-Morillo. "Evaluation of deep neural networks for traffic sign detection systems", Neurocomputing, 2018

Publication

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		