Wildlife Image Classification Using CNN

Nitish Rathore,

Niveditha Kundapuram,Vishal J Lodha  
Student   
Department of CSE  
People Education Society UniversityBengaluru, India  
[pes1ug20cs270@pesu.pes.edu](mailto:pes1ug20cs270@pesu.pes.edu)  
[pes1ug20cs271@pesu.pes.edu](mailto:pes1ug20cs271@pesu.pes.edu)  
[pes1ug20cs507@pesu.pes.edu](mailto:pes1ug20cs507@pesu.pes.edu)

Gayathri.T  
Department of CSEPeople Education Society University  
Bengaluru, India

Dr.MamathaiH R  
Department of CSE(Computer Science Engineering).PES University  
Bengaluru, India

***Abstract*—Image classification is one of the core problems in the field of vision with a large variety of practical applications. Deep Learning has evolved over the years it’s use cases have diversified. In this paper, we present two detection models - custom CNN and custom YOLOv3. CNN model recognizes and classifies wildlife images into two classes - ‘Coyotes’ and ‘Dogs’, whereas the YOLO model classifies the images into- ‘Horses’ and ‘Zebras’. The models are trained on the Florida-Wildlife-Camera-Trap dataset. We have built our own CNN model which would predict between two similar looking classes in both day and night vision.**

***Keywords—CNN, YOLO, neural network, convolutional, deep learning, wildlife image classification,tensorflow***

1. Introduction

The incidents of wildlife infiltrating residential areas specially villages during night time always results in accidents or loss of lives, it is an issue which needs to be addressed and resolved soon, often lack of knowledge among the people specially villagers and less awareness results in tragic incidents.The ability of Deep learning methods to build precise models for extracting features from the dataset and classify images which do not require human intervention aids to the issue raised.

In this project, we deploy two Convolutional Neural Networks with TensorFlow framework, one for image classification between dogs and coyotes and the other between Horses and Zebras. This consists of a set of steps- we first import the dataset and rescale the images which best fits the model for prediction, Features from the images are extracted and fed into the CNN-model for training on the dataset, which is then being tested on.

## Trail Camera images from the forests of Florida.

The original dataset consists of 104,495 images of 22 animal classes. We have chosen two classes- Dogs and Coyotes and compiled our own dataset that consists of 340 images in the ratio 57.31 **:** 42.69.

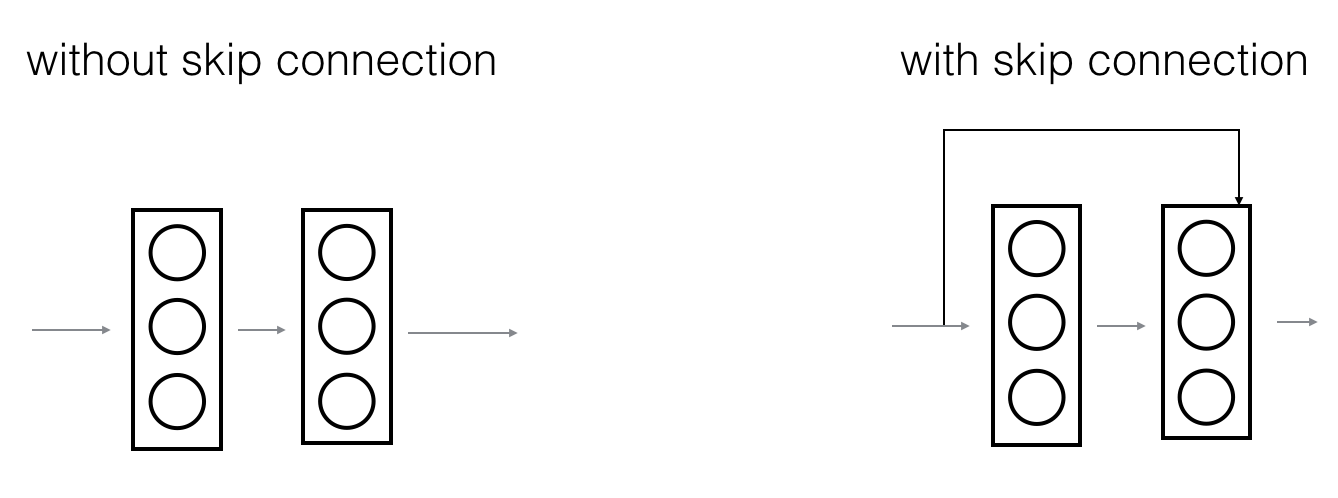
## Data drawbacks.

Trail cameras capture images throughout the day and well into the night. However, the images of classes required by our model were captured during the night and inevitably suffered illumination and quality issues. We also found that in some images the animals were obscured by the vegetation around, which gave rise to ambiguity in their class prediction. The size of our dataset was also small according to conventional standards for datasets used in Deep Learning models.

II.BACKGROUND AND RELATED WORK

ResNet50-

We began with ResNet50, as it is a reputed model particularly in the domain of Computer Vision. It was a fundamental breakthrough in the field of DL as it allowed us to train more than 150+ neural networks in an efficient way. It solved the problem of vanishing gradients by introducing ‘skip connections’.



Skip connections aid in flow of output from earlier layers to later layers, allowing the later layers to perform at least as good as the earlier layers and not worse. Thus, it mitigates the vanishing gradient problem.

The ResNet-50 model comprises of 5 phases each with a convolution and Identity block. Each convolution and identity block have 3 convolution layers. The model has over 23 million trainable parameters.

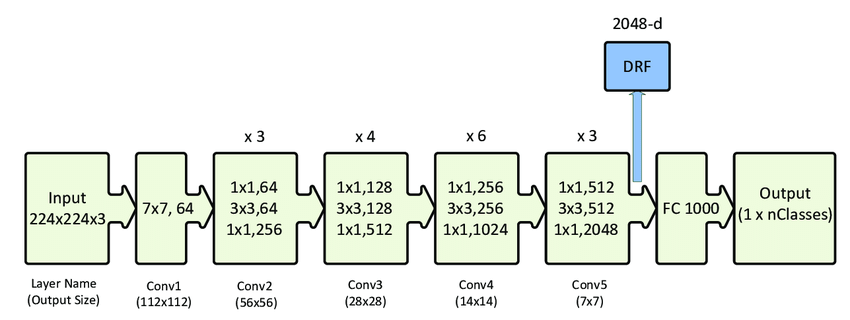


Image: Resnet50

CNN-

In neural networks, CNN is the primary technique used for image classification. The incredible power to learn from data and environment makes it the best choice to efficiently solve complex tasks for researchers.

CNN model takes the input image and breaks the image into a matrix of pixels and store the pixel values in respective locations.

Convolution Layer

This layer is used to collect information from the image by traversing throughout the complete picture in boxes or grids without altering any properties of the images fed.

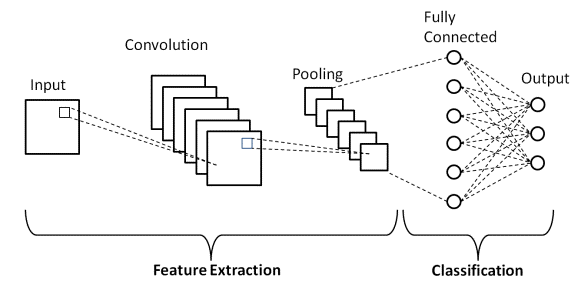
As illustrated down, these are the different images we get when we apply unique features for gathering information 

Image:Convolution Neural Network

Strides

It's the pixel value that shifts over the input matrix.If stride is 3 then it means that we shift over three pixels, the smaller the stride more feature maps are generated

Global Padding

Sometimes the filter does not fit the input image. We have two options:

•Pad the picture with zeros (zero-padding) so that it fits

•Drop the part of the image where the filter did not fit. This is called valid padding which keeps only valid part of the image.

ReLU

ReLU is a matrix product of the pixels to add non linearity to the data. There are other non direct methods, for example, tanh or sigmoid. The majority of prefer this because working of ReLu is the best

ReLU’s purpose is to keep the gradient always high if the neuron activates and so it does not saturate and the neural network isn’t linear. As, the actual data would want our neural network to learn would be positive.

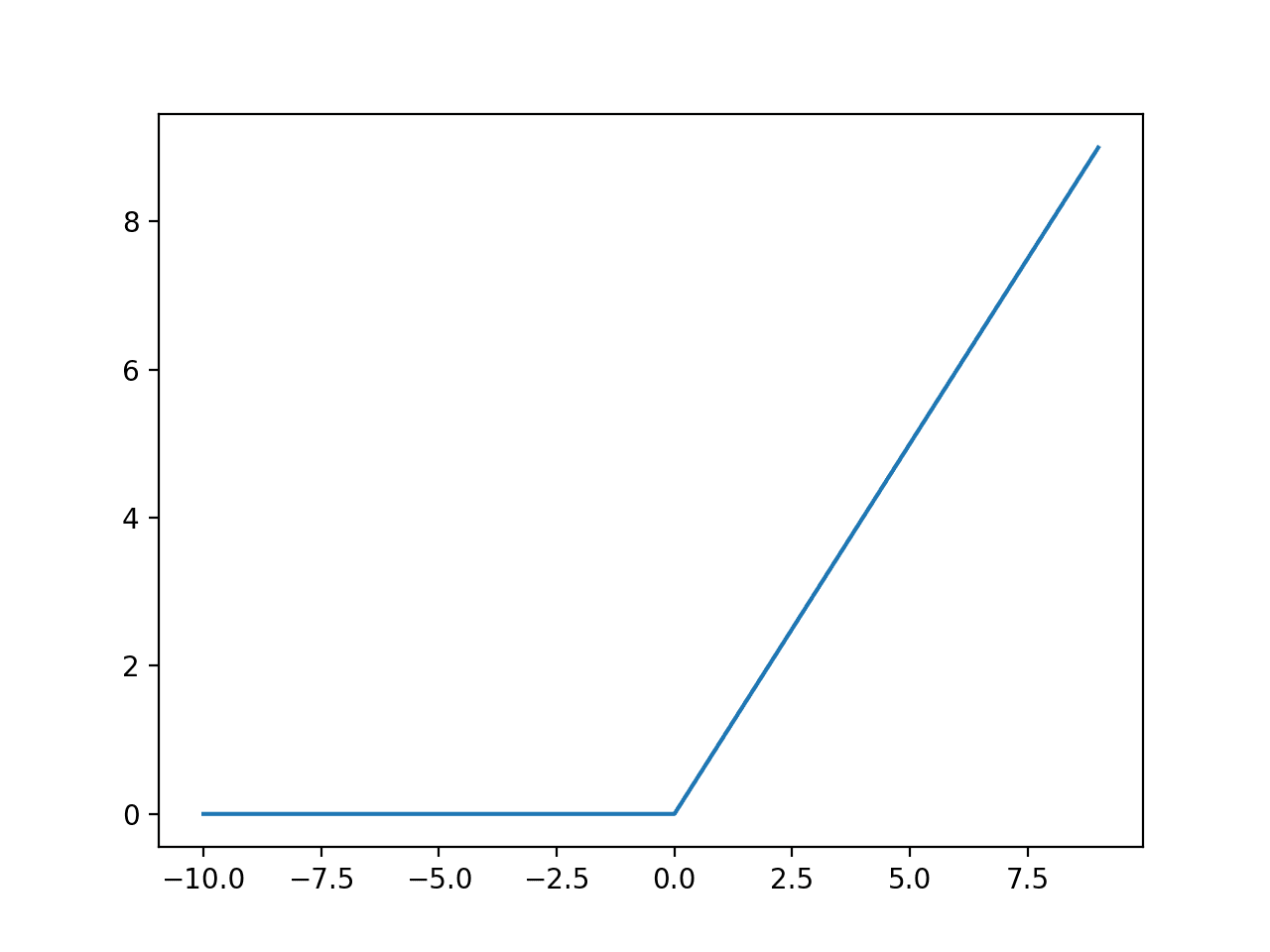


Image: ReLU Function

Soft Max

Its is a mathematical function which converts some random vectors to probabilities such that sum of all is equal to 1.It gives a relative scaling of vectors

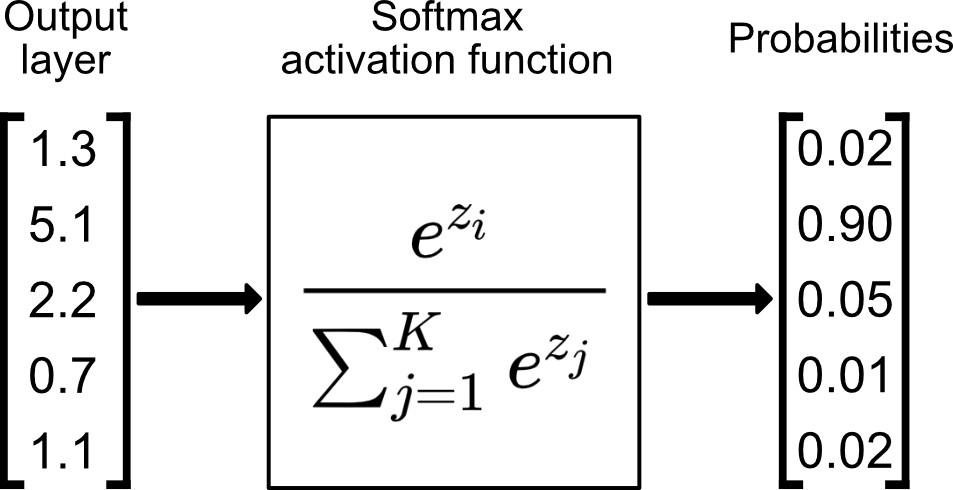


Image:SoftMax Function

Sigmoid Function

It is a logistic function which classifies points between values 0 to 1 such that 50% values are below 0.5 and 50% values above 0.5.It is given by σ(x) = 1/(1+exp(-x)).

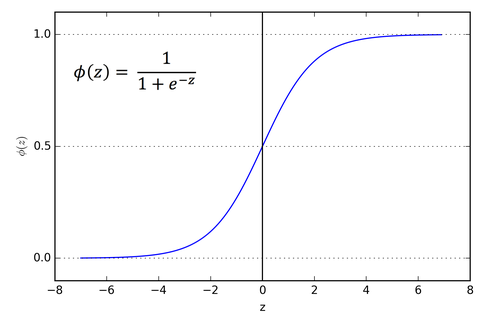


Image: Sigmoid Function

Global Pooling Layer

Global Pooling layers decrease the limit when the photos are unreasonably colossal. Global pooling in a similar manner is called downsampling which lessens the dimensionality of each guide as well as holds critical information. The characteristic types of pooling are:

• GlobaliMax Pooling

• Global Min Pooling

• Global Average Pooling

Pooling is a process where we decrease the shape of our network, it is done in many ways but Global Max Pooling is one of the well known used methods. In this method the maximum element from the kernel is chosen and placed to form a new shape of network.

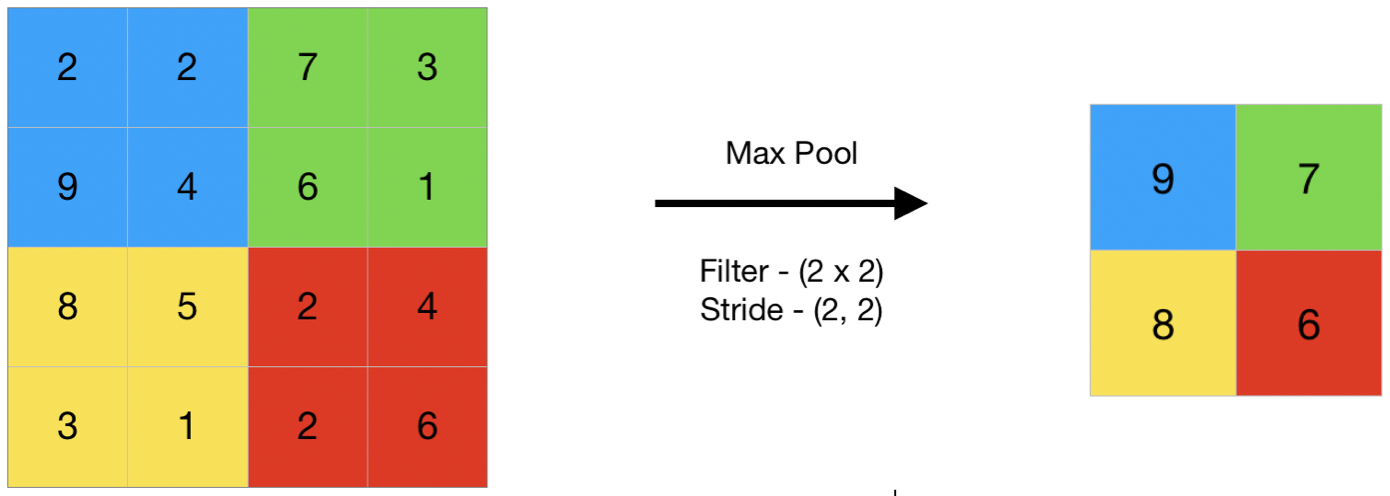


Image:Pooling

Fully Connected Layer

Fully connected layer or also called dense layer means that each neuron in the following layer of neuron is powered by all neurons of the present neuron,hence forming a full connection. In many well known AI models almost all the layers are fully connected hence helping to give a precise output.

MIRnet:

Determined to recuperate excellent picture content from its debased rendition, picture reclamation partakes in various applications, for example, in photography, security, clinical imaging, and remote detecting. In this model, we execute the MIRNet model for low-light picture enhancement, a completely convolutional design that learns an advanced arrangement of elements that joins logical data from various scales, while at the same time protecting the high-goal spatial subtleties.

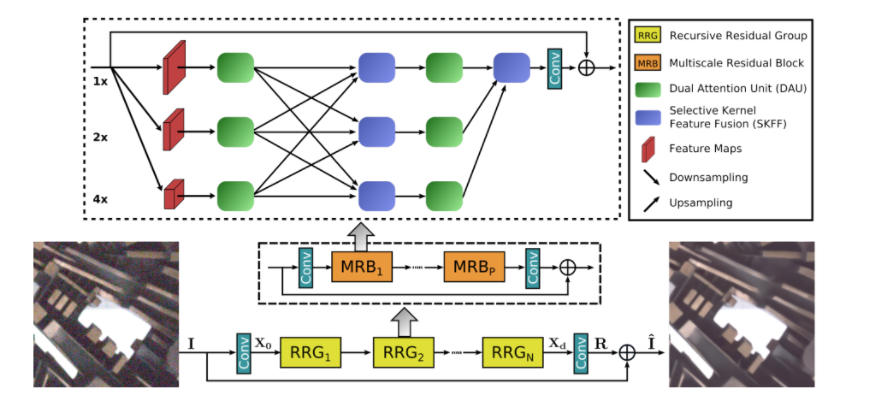


Image:MIRnet model

Features:

A feature extraction model processes a basic plan of features across various scales, while staying aware of the main focused features to shield near perfect spatial nuances.

A reliably repeated part for data interchange, where the features across multi-resolution branches are ceaselessly consolidated for additional created depiction learning.

This model has been trained on the LOL Dataset which has 485 pairs of images, that is a low light image and a similar good light photo. Each photo is reshaped to128x128 and then processed through to get weights for the neural networks.

Selected Kernel Feature Fusion(SKFF in short):

It is one of the key features of the MIRnet model.This unit performs dynamic Adjustment of receptive fields with the help of Fuse structured CNN and Select structured CNN.

Fuse operation : A global feature descriptor is made by this Fuse structured CNN, it is done by manipulating the information from a multi-resolution stream.It uses Global Pooling across contiguousidimensions.

Select operation : This operation uses these features produced by fuse operations to recreate the feature map by changing this values according to the new results. It uses a Softmax function to the above features

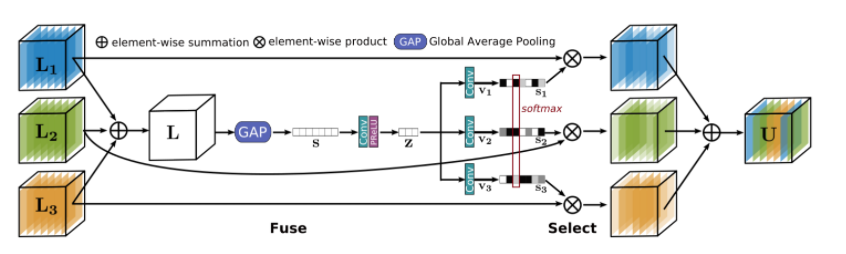


Image: Selected Kernel Feature Fusion

Dual attention Unit:

Now after features have been extracted by the SKFF feature extractor, we pass these features through the Dual Attention Unit(DAU) unit which allows only useful features to pass through and suppresses less useful features. A unit is required which can share information about the spatial and channel features with the tensor, this is done by Dual Attention Unit which comprises Channel Attention unit and Spatial Attention unit.

Channel Attention Mechanism : It applies squeeze and excitation operations by using a pooling method which takes average across spatial dimensions. It consists of two CNN layers with sigmoid function.

Spatial Attention Mechanism : It exploits inter-spatial dependencies of features on convolution It generates a spatial attention map and uses two types of pooling Global Max and Global Average Pooling along with with a convolution network with a sigmoid activation

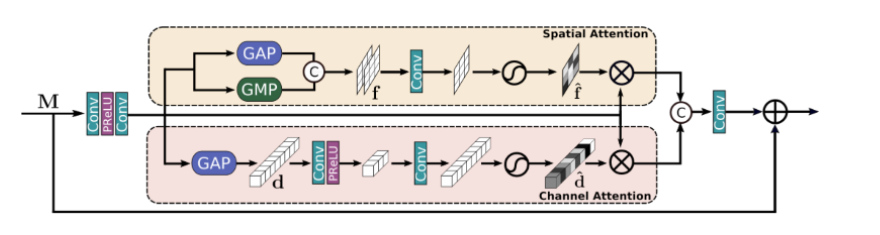


Image:Dual attention unit

Multi-Scale Residual Block:

It generates a spatial precise output of a low light image in such a way that a high resolution of the image. It consists of CNNs running parallely . This parallel feature helps in keeping the high resolution of the image intact with the help of low resolution features. The MIRnet model applies a recursive approach to solve the problem of low light enhancement.

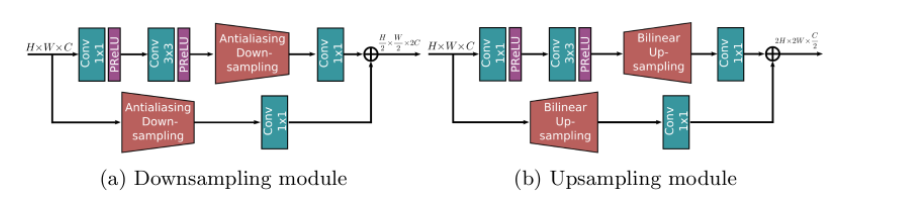


Image: Multi-Scale Residual Block

The MIRnet model helps us to regain many small features which have been lost due to the presence of low light. This model helps in pre-processing the image such that many key features can be extracted from the image.



Image: MIRnet enhancement

YOLOv3:

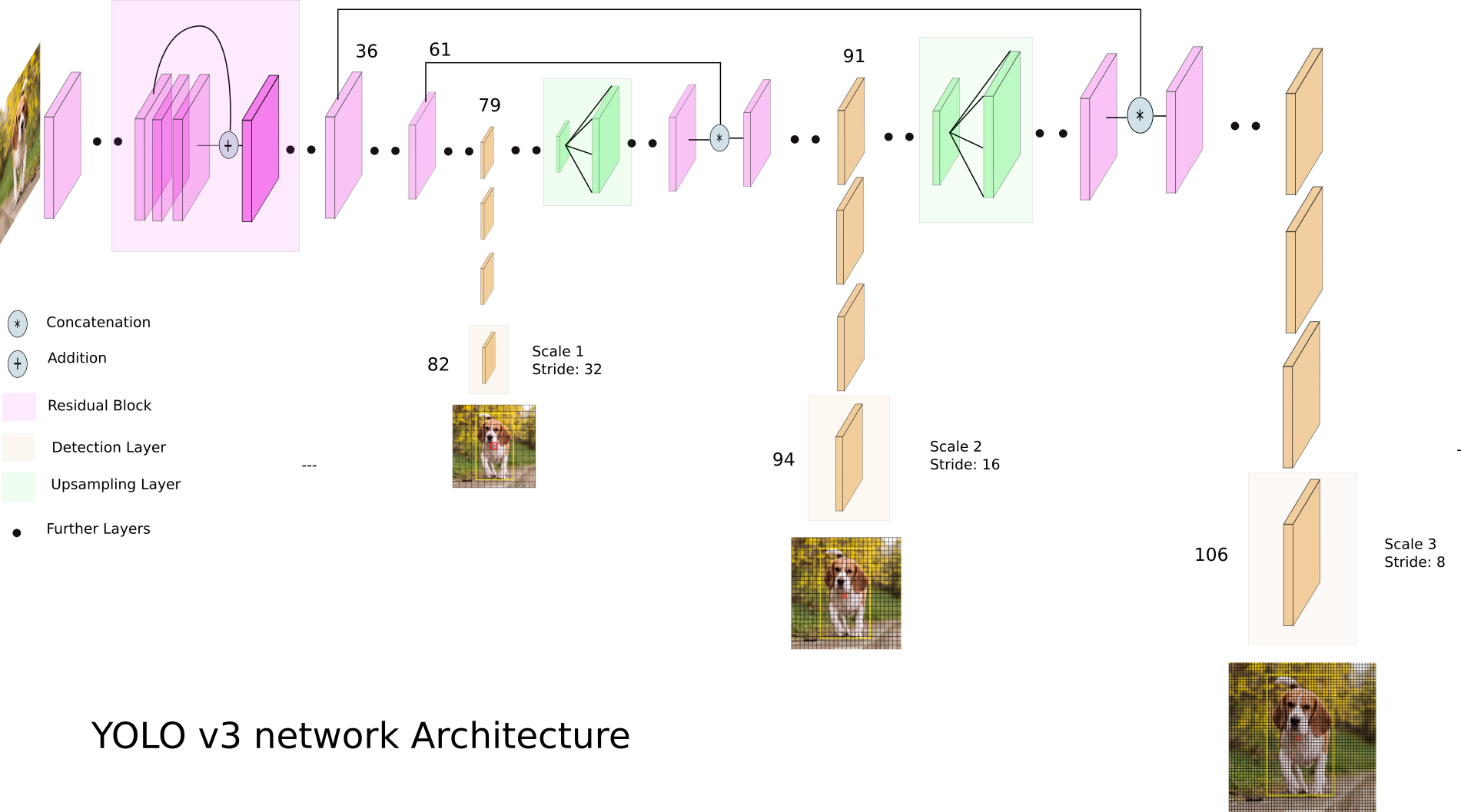
YOLO detection model is one of the most apt models for real-time object detection with very high accuracy. This is due to the entanglement of the network called Darknet, which originally has 53 convolutional layers. Supplementary fifty-three layers are included for task of object detection. This gives the 106 full convolutional layer for the YOLOv3 model.

The model is equipped with skip connections and unsampling. It makes detections at three different regions at three different scales, by applying 1x1 kernels of three different sizes.

The shape of the detection kernel is calculated by the formula-

**1 x 1 x (B x (5 + C)),**

B represents number of bounding boxes around the object and C represents number of classes we are studying overall



The images are downsampled by 32, 16 and 8 at three different regions of the network.

The prediction of object confidence and class are done by logistic regression. YOLOv3 doesn’t use SoftMax in the later layers. It calculates each class score predicted by logistic regression and compares the value with a threshold. All the scores above the threshold are displayed on the box. Thus, YOlOv3 executes multi-label classification of objects.

# III. EXPERIMENTS AND RESULTS

On trying different methods, we got the following results.

1. *Pre-Processing:Resizing and reshaping and running through MIRnet model*
2. *Classification Model  
   a. ResNet50*

**

Resnet does not have a class for black panther so it is not able to classify the animal and since some of the features match to curly coated retriever it gave accuracy of 62.14%

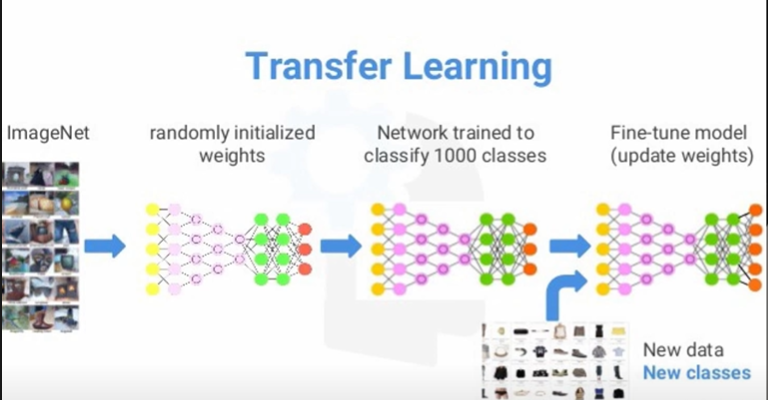
*b. Custom CNN*



|  | Validation | Testing |
| --- | --- | --- |
| Accuracy | 85.18% | 89.13% |
| Loss | 17.91% | 14.72% |

*b. YOLOv3*

We have implemented Transfer Learning with YOLOv3, where we took a pre-trained model and formatted certain aspects of the model’s last three layers and applied it to our classification problem.



The formula used to calculate max\_batches and filters are

max\_batches= 2000\*n filters = (n+5)\*3,

n represents number of classes

As n=2 for our model, we have changed to max\_batches and filters to 4000 and 21 accordingly.

The batches and subdivisions are also changed according to the size of our dataset.

The shape of our detection model calculated by the formula given in \_\_\_\_, comes to 1x1x21.

The model predicts the class of the animal with an accuracy of around 90-92%.

IV. CONCLUSIONS

*Drawbacks:*

We have only worked on improving the low light presence in the images which is helping us to give accurate answers for some images but we have excluded images where only parts of the body of the animal are visible.

We have only classified coyotes-dogs and Horse-Zebra, but there are many other animals that may cause harm if they enter a settlement area, and many more that need to be studied for further advancements in the field of biology and other interdisciplinary fields.

As we have used around 300 images our model weights aren't perfect hence our accuracy is less.

V. FUTURE WORK

We have achieved decent accuracy with our models for wildlife image detection. We plan to improve on our models to achieve higher accuracy and wish to extend our project to Real- Time detection implementing more efficient models. We would also like to improve our implementation from binary classification to multi-class classification by working on different animal classes.

##### References

1. TensorFlow: Image Recognition (2018). <https://www.tensorflow.org/tutorials/image_recognition>.
2. Rohrer, B: How do Convolutional Neural Networks work? (2016). <http://brohrer.github.io/how_convolutional_neural_networks_work.html>.
3. Rangdal, Mukesh | Hanchate, Dinesh. (2014).  [(PDF) Animal Detection Using Histogram Oriented Gradient | International Journal IJRITCC - Academia.edu](https://www.academia.edu/8476934/Animal_Detection_Using_Histogram_Oriented_Gradient)
4. Mehdi Laziri Supervised by Dr. Naeem Nisar Sheikh<http://www.aui.ma/sse-capstone-repository/pdf/spring-2019/DETECTING_ANIMALS_USING_COMPUTER_VISION.pdf>
5. [Ruilong Chen](https://onlinelibrary.wiley.com/action/doSearch?ContribAuthorRaw=Chen,+Ruilong) | [Ruth Little](https://onlinelibrary.wiley.com/action/doSearch?ContribAuthorRaw=Little,+Ruth), [Lyudmila Mihaylova](https://onlinelibrary.wiley.com/action/doSearch?ContribAuthorRaw=Mihaylova,+Lyudmila) | [Richard Delahay](https://onlinelibrary.wiley.com/action/doSearch?ContribAuthorRaw=Delahay,+Richard) | [Ruth Cox](https://onlinelibrary.wiley.com/action/doSearch?ContribAuthorRaw=Cox,+Ruth) <https://doi.org/10.1002/ece3.5410>.
6. Akshay Patil| Tejas Chaudhari |Ketan Deo | Kalpesh Sonawane | Rupali Bora. [(PDF) Low Light Image Enhancement for Dark Images (researchgate.net)](https://www.researchgate.net/publication/346809857_Low_Light_Image_Enhancement_for_Dark_Images)
7. M.Praveena | V.Pavan Kumar | R. Asha Deepika | Ch.Sai Raghavendhar | J.Rahul Sai Reddy [F10610486S319.pdf (ijitee.org)](https://www.ijitee.org/wp-content/uploads/papers/v8i6s3/F10610486S319.pdf)

[8] Introduction to neural networks: Deep learning. Analytics Vidhya. (2020, May 13). Retrieved December 8, 2021, from https://www.analyticsvidhya.com/blog/2018/10/introduction-neural-networks-deep-learning/.

[9] Prabhu. (2019, November 21). Understanding of convolutional neural network (CNN) - deep learning. Medium. Retrieved December 7, 2021, from https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148.

[10] B. H. Curtin and S. J. Matthews, "Deep Learning for Inexpensive Image Classification of Wildlife on the Raspberry Pi," 2019 IEEE 10th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), 2019, pp. 0082-0087, doi: 10.1109/UEMCON47517.2019.8993061.