Web Scraping:

Scraping folder has site\_urls.txt which contains URLs of various websites and wildlife keywords.txt which contains search keywords related to wildlife. Folder also has the web scraping scripts for each URL. Some sites are not responding to web scraping due to following errors, indicated those URLs in filename with NW (not working) for example urls\_0\_NW.ipynb

Errors:

* Blocking due to detection of automation script
* Input box or links are hidden and selenium not picking it
* 500 error
* Captcha or login
* Html changing for every ad

Data folder structure: (indexing naming convection)

Diagram, schematic

Description automatically generated

Table

Description automatically generated

* 0.txt file contains description of first ad, for example

{

"name" : "Alligator Purse",

"price" : "$35",

"location" : "Lakeland, FL, US",

"details" : "Designer Homes Saturday Auction by Designer Homes of Central

Florida",

"description" : “Lieber black alligator clutch handbag with convertible shoulder

strap”

"url" : Url link

}

* 0\_imgs.txt contains images URLs from first ad, for example

["https://image.invaluable.com/housePhotos/ShowplaceAntiques/12/732412/H20259-L303867646\_original.jpg","https://image.invaluable.com/housePhotos/ShowplaceAntiques/12/732412/H20259-L303867648\_original.jpg"]

* Manually filtered for images related to wildlife (data\_filter folder)
* Data\_process.ipynb will process the data as required for the model into data\_process folder

Image captioning: (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)

Diagram

Description automatically generated with medium confidence

* When given an input image to a pre trained deep learning model like VGG, Xception, Inception and Resnet.
* The deep learn model will extracts features from images which is one of the inputs to RNN

Diagram

Description automatically generated

* The above image is VGG16 pretrain deep learn model, we will consider only up to last but 2 layers (upto 1\*1\* 4096) in order get features instead of classification output
* The other input to the model RNN is text data from the captions as below

Text

Description automatically generated

* Model must predict the next words and from the sentence by taking images features as input\_1 and text as input\_2 as shown below

Diagram

Description automatically generated

* model\_vgg.ipynb, model\_xception.ipynb, model\_inception.ipynb, model\_resnet.ipynb are python script for respective VGG, Xception, Inception and Resnet models

Multi-label-classification:

* Image tagging is a multilabel classification model. Considered images as input
* For labels,

Table

Description automatically generated with medium confidence

A picture containing diagram

Description automatically generatedTitle: tiger eye large carved skull

If words exist in list of filtered unique words from wildlife keywords.txt

Label: tiger skull

Title: eye large carved

* If any word from title does not exist in filtered list, then keyword for image is taken. For example, for image 0\_545\_0\_0.jpg keyword by which we have searched is tiger+skull (label)
* Taken weights from pretrained mobilenet\_v2\_100\_224 model and did transfer learning and updates weight according to our images.
* Multi\_label\_classifiction.ipynb is the python script from image tagging.