Problem Definition:

# Topic: Image Tagging

Domain: AI, ML/DL, Image Processing

There are already popular models that will allow to find and determine what objects are present in an image. This use case takes the same one level further, it also contains association of tags to the images to improve their searchability. The same can be done on 2 levels, the basic level being to tag using similar meaning words which is quite simple. And the more advanced version of the same being that the interns will have to find associated words too (so, for example, if crown is an object detected, king and queen should also be tagged)

Objectives and Features Completed:

The objectives of our project completed are:

1. Loaded a pretrained model - We have successfully loaded a pretrained model - RESNET 50 (Residual Network with 50 layers.) which is trained on imagenet dataset.
2. Loaded a dataset - ImageNet dataset has over **14 million images** maintained by **Stanford University**. It is extensively used for a large variety of Image related deep learning projects. The images belong to various **classes** or **labels.**
3. Specify image transformations- Various transformations like resize , center crop, normalization, etc are specified.
4. Load the image and pre-process it - Read the input image and perform the image transformations specified.
5. Model Interference - Use the pre-trained weights to find out the output vector. Each element in this output vector describes the **confidence** with which the model predicts the input image to belong to a particular class.
6. Forward Pass: Based on the scores obtained, display the predictions.
7. Give the synonyms for the prediction of the image using Natural Language Toolkit (NLTK) Library.

Alternative approaches that you found/looked into:

1. By using different pretrained models like YOLO , Resnet101 and Alexnet.

# 2. Another approach was to train the model .The training set contains the labeled source images and the unlabeled target images. It is first trained on both ImageNet (the source domain) and personal photos (the target domain) by pre-training and fine-tuning for discovering shared middle-level feature abstractions across domains. Once the shared feature abstractions are learned, the top layer with ontology priors is further trained. In the testing stage, the resultant parameters W and B can be transferred to the target domain to obtain the middle-level feature representations (a bottom-up transfer) and high-level confidence scores (a top-down transfer)

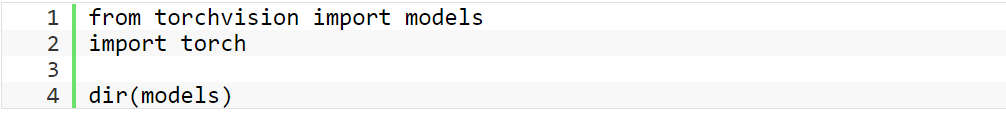
1. By using keras from tensorflow library. Keras provides a higher-level abstraction to the underlying ML toolkit. Hence, prototyping quick ML frameworks for feasibility is agnostic to the changes brewing underneath.

How To Use:

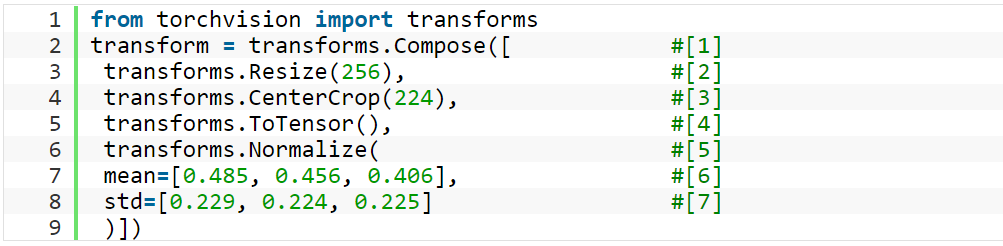
Initially we installed TorchVisionmodule using the command given below



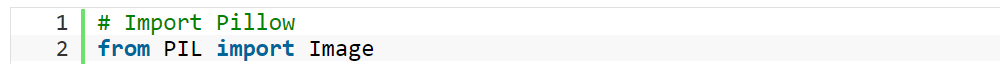
Next, we imported models from torchvision module



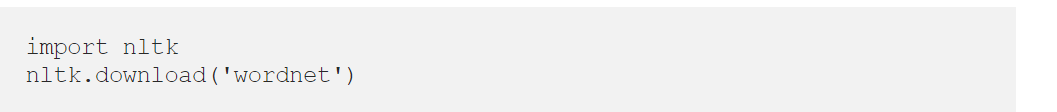
Once we have the model with us, the next step is to transform the input image so that they have the right shape and other characteristics like mean and standard deviation. We can pre-process the input image with the help of **transforms** present in **TochVision** module.



Now to input image and carry out the image transformations we have specified above. We have used Pillow (PIL) module extensively with TorchVision as it’s the default image backend supported by TorchVision.



Next, we imported NTLK library and installed Wordnet to get a collection of words and vocabulary from English language that are related to each other and grouped in some other ways.



Strengths And Weaknesses:

We can here summaries some of the strength and weaknesses of our project.

Strength:

* Human error and bias are minimized (fewer decisions are required by the user).
* More uniform classes are produced as output.
* Spectrally distinct classes present in the data may be revealed which were not initially apparent to the user.
* Our project goes one step further and has an added advantage over other pre-trained models; In the output our model gives out words related to object in the image instead of just stating the obvious.
* We have to keep the backbone part obtained from the pretrained model fixed and only allow the parameters of the classifier to change. This approach is ideal when you want to train a model quickly or without much computational resources.

Weakness:

* Spectral grouping produced by the classifier may not correspond to the information classes of interest to the user.
* There is limited control over the 'menu' of classes.
* The performance of our model in this case might not be the best because the pretrained backbone may suffer from domain adaptation.

Conclusion:

In a nutshell, this internship has been an excellent and rewarding experience. We can conclude that there has been a lot that we’ve learnt from our work here. Two skills that we developed during the course of this internship are managing time efficiently and secondly working together in a group environment. We were able to build a good connection with each other and we have a much better understanding in regards to deep learning and more specifically object detection. Moreover, we would like to thank our mentors Hetul Mehta, Burhan Plumber, Irfan Siddavatam sir and Ashwini Dalvi Ma’am; they left no stone unturned in guiding us to the best of their knowledge. Overall, our internship has been a success and being a part of such a great organization speaks for itself, we couldn't be more thankful.