**Project Report On**

Applying Data Science Pipelines, DevOps Practices, ​

& Computing Techniques to Image Classification Tasks ​

Nikitha Patel, Harman Kaur, Shubham Bansal, Krishna Agarwal, Manas Shukla

University Of Calgary

Data 608 - Developing Big Data Applications

8/04/2022

**Introduction:**

In high-level vision, indoor scene recognition is a difficult open problem. In the indoor domain, most scene recognition models that work well for outdoor situations perform badly. The major problem is that, while some indoor scenes (for example, hallways) can be well described by global spatial features, others (for example, bookstores) are better described by the things they include. In general, we require a model that can use both local and global discriminative information to solve the interior scene recognition problem.

**Dataset:**

The database contains 67 Indoor categories, and a total of 15620 images. The number of images varies across categories, but there are at least 100 images per category. All images are in .jpg format. All images have a minimum resolution of 200 pixels in the smallest axis.



**Figure 1**. The following is a list of the 67 indoor scene categories that we employed in our Project. To make it easier to see the range of scenario genres considered here, we've divided them into five large scene groups. There are 15620 photos in the database. In the smallest axis, all photos have a minimum resolution of 200 pixels.

The below are the 67 indoor scene categories that we employed in our project:

['.DS\_Store', 'children\_room', 'pantry', 'meeting\_room', 'bedroom', 'restaurant', 'florist', 'church\_inside', 'greenhouse', 'concert\_hall', 'toystore','toystore', 'bakery', 'gameroom', 'bathroom', 'corridor', 'laundromat', 'dentaloffice', 'office', 'stairscase', 'bowling', 'casino', 'lobby', 'fastfood\_restaurant', 'library', 'laboratorywet', 'grocerystore', 'dining\_room', 'tv\_studio', 'classroom', 'restaurant\_kitchen', 'bookstore', 'studiomusic', 'cloister', 'shoeshop', 'hospitalroom', 'movietheater', 'prisoncell', 'clothingstore', 'poolinside', 'operating\_room', 'jewelleryshop', 'kindergarden', 'livingroom', 'mall', 'hairsalon', 'videostore', 'locker\_room', 'museum', 'computerroom', 'artstudio', 'gym', 'airport\_inside', 'garage', 'nursery', 'waitingroom', 'elevator', 'warehouse', 'kitchen', 'inside\_bus', 'bar', 'closet', 'deli', 'auditorium', 'inside\_subway', 'trainstation', 'buffet', 'subway', 'winecellar']

**Image classification Vs Object detection:**

*Image classification:* Image classification, in simple terms, is a technique for classifying or predicting the class of a single object in an image. This technique's major purpose is to precisely identify the characteristics in an image.

How Image classification works?

Image classification approaches are broadly classified as parametric and non-parametric, supervised and unsupervised, as well as hard and soft classifiers. This technique gives results for supervised classification based on the decision boundary produced, which is primarily based on the input and output provided while training the model. However, in the case of unsupervised classification, the technique produces results based on its own examination of the input dataset; characteristics are not directly given to the models.

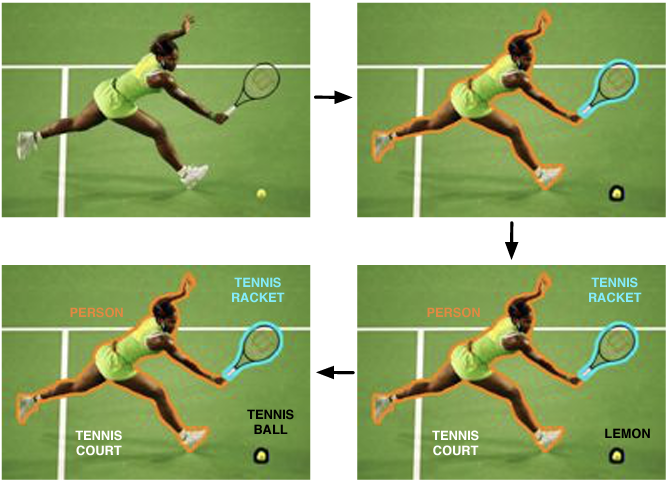
The main steps in image classification techniques are determining a relevant classification system, feature extraction, selecting good training samples, image pre-processing and selecting an appropriate classification method, post-classification processing, and finally assessing overall accuracy. The inputs in this technique are typically images of specific things, and the outputs are predicted classes that define and match the input objects. Convolutional Neural Networks (CNNs) are the most widely used neural network model for image classification.



**Figure 2:** Identifies the bookshelves in the background and classify images from Library class.

*Object Detection:* Object detection is defined as determining where objects are situated in a given image, also known as object localization, and which category each object belongs to, also known as object classification. Object detection is a form of image classification technique that, in addition to categorizing, determines the position of object instances in natural images from a large number of predefined categories.

This technique has the capability to search for a specific class of objects, such as cars, people, animals, birds, etc.



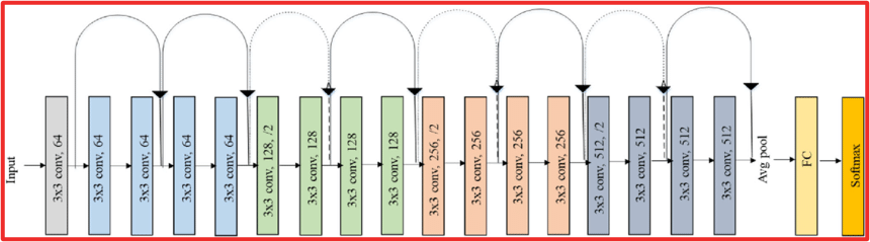
**Figure 3:** An original image is segmented into objects; each object is categorized.

**Data Transformation & Loading:**

**ResNet18:**

* Resnet is a pretrained network that was trained on images in 1000 object categories, such as keyboard, mouse, pencil, and many animals.
* This model was the winner of the 2015 ImageNet competition.
* It is trained on ImageNet dataset which consists of more than 14 million images.
* ResNet-18 is a convolutional neural network that is 18 layers deep.
* It contains 11.174M training parameters.

**Architecture:**



ResNet18 consists of 18 layers in total ,17 convolutional layers and one Fully connected layer. If we look at the above architecture it can be seen that it uses a 3X3 filter for every convolutional layer and also make use of the skip connections which prevent the problem of vanishing gradient descent.

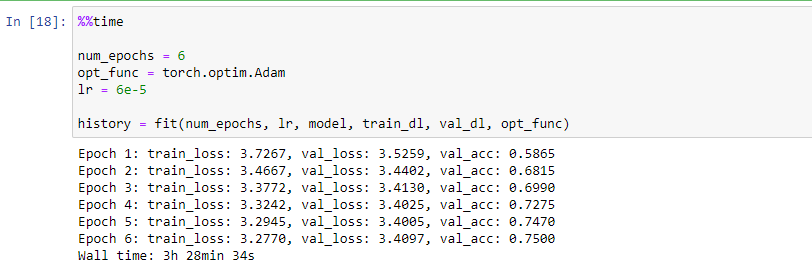
**Model Fitting:**

The initial step before using a pretrained ResNet model was to modify the Fully connected layer. The reason behind it was that the ResNet original training was done on the dataset which consisted of 1000 unique categories, and the Indoor Scene Recognition dataset consisted of 67 indoor categories.

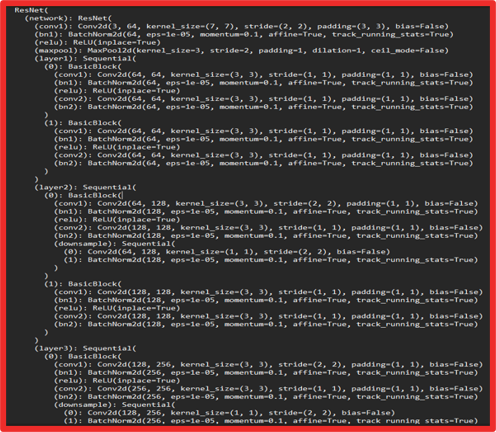
Therefore, to make sure that the pretrained model can be used on our dataset we changed the output features in the Fully Connected Layer to 67.If we look at the ResNet architecture it shows that the activation function used for the original training was SoftMax function, but for fitting our indoor scene training dataset we used a sigmoid function.

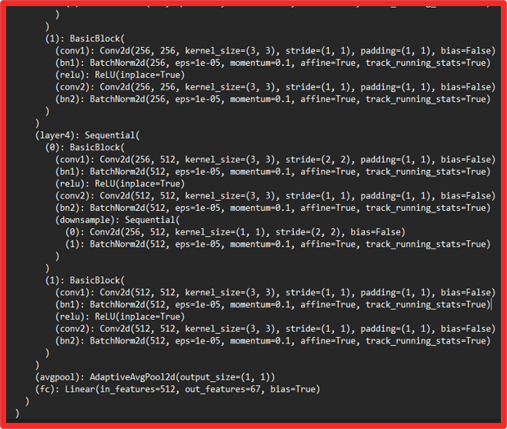
The code for fitting the training set on the pretrained ResNet18 Model is shown below:





The final ResNet18 model with all the layers is shown below:





The training parameters for our Model are as follows:

**Batch size:** It is the number of samples processed before the model parameters are updated. For training our model we used a batch size of 25.

**Learning rate:** For our model we used a learning rate of 6e-5. It is a hyperparameter that controls how much to change the model in response to the estimated error each time the model weights are updated.

**Number of epochs:** The number of complete passes through our training set was 6.

**Loss function:** Cross Entropy loss, measures the performance of the classification model whose output is a probability value between 0 and 1. For our model we are using a sigmoid function in the last layer, which gives the output in the range of 0 and 1, therefore, cross entropy loss was a reliable choice.

**Optimizer:** Adam Optimization function which is stochastic gradient descent method.

**Model Evaluation:**

For evaluating the performance, of our model, we plotted training and validation loss with increasing number of epochs. It can clearly be seen in the below given graph that the training loss kept decreasing with increasing the number of epochs, but the validation loss was almost constant after 3 epoch. This indicates that there was not a significant amount of change in model parameters after 3 epochs. Therefore, we decided not to train our model for more than 6 epochs, as it can lead to overfitting.

