Deep Learning for Advanced Robot Perception

Assignment 3: Parsing data and Regularization

This homework will provide the foundations for working with large datasets. As a part of this homework you will:

- Download the VOC PASCAL data set
- You will implement an image parser.
- You will develop your Deep Learning model (reduce overfitting using Dropout Regularization)

Getting the dataset

Make sure you are connected to the internet. Get the VOC PASCAL 2012 dataset.

To download the training/validation data, see the development kit (http://host.robots.ox.ac.uk/pascal/VOC/voc2012/#devkit).

The training data provided consists of a set of images; each image has an annotation file giving a bounding box and object class label for each object in one of the twenty classes present in the image. Note that multiple objects from multiple classes may be present in the same image. Annotation was performed according to a set of guidelines (http://host.robots.ox.ac.uk/pascal/VOC/voc2012/guidelines.html) distributed to all annotators.

A subset of images is also annotated with pixel-wise segmentation of each object present, to support the segmentation competition.

Images for the action classification task are disjoint from those of the classification/detection/segmentation tasks. They have been partially annotated with people, bounding boxes, reference points and their actions. Annotation was performed according to a set of guidelines (http://host.robots.ox.ac.uk/pascal/VOC/voc2012/action_guidelines/index.html) distri buted to all annotators.

The data has been split into 50% for training/validation and 50% for testing. The distributions of images and objects by class are approximately equal across the training/validation and test sets. Statistics of the database are online.

Parsing data (50%)

The dataset contains an XML file for each picture. The XML file describe the location of several object and their labels of the object/action in the picture.

1.- First approach: generating an intermediate data set

Your python code should parse the XML file and then extract the elements from the corresponding image that were described.

The elements extracted from the image should be organize as a new dataset where the labels become folder names and the sub-images extracted became image files.

All the images should be resized to 224x224 without losing image resolution.

This process should be automatic for the whole VOC PASCAL dataset.

Once the new dataset is generated you are require to generate a single binary file to compress the whole data

2.- Second approach: using the RAM

Your python code should parse the XML file and then extract the elements from the corresponding image that were described.

The elements extracted from the image should be organize as a new dataset. This process should be done in memory without writing any temporary file in the HD.

All the images should be resized to 224x224 without losing image resolution.

This process should be automatic for the whole VOC PASCAL dataset.

Once the new dataset is reorganized you are require to generate a single binary file to compress the whole data.

You are required to submit your code and you are not required to submit the final binary file. DO NOT modify the file structure of the VOC PASCAL dataset.

Building a Deep Learning model with dropout (50%)

A simple and powerful regularization technique for neural networks and deep learning models is dropout. (Lecture 6)

Dropout is a regularization technique for neural network models proposed by Srivastava, et al. in their 2014 paper Dropout: A Simple Way to Prevent Neural Networks from Overfitting. Dropout is a technique where randomly selected neurons are ignored during training. They are dropped-out randomly. This means that their contribution to the activation of downstream neurons is temporally removed on the forward pass and any weight updates are not applied to the neuron on the backward pass.

The dataset we will use in this section of the homework is the Sonar dataset. This is a dataset that describes sonar chirp returns bouncing off different surfaces. The 60 input variables are the strength of the returns at different angles. It is a binary classification problem that requires a model to differentiate rocks from metal cylinders.

It is a well understood dataset. All of the variables are continuous and generally in the range of 0 to 1. The output variable is a string M for mine and R for rock, which will need to be converted to integers 1 and 0. The dataset contains 208 observations.

This dataset is a standard benchmark problem. This means that we have some idea of the expected skill of a good model. Using cross validation, a neural network should be able to achieve performance around 84% with an upper bound on accuracy for custom models at around 88%. You can learn more about this dataset on the UCI Machine Learning repository.

You will find three different implementation of a Deep Learning Model solving this problem in Files/week 5/ Reduce overfitting.

Your assignment consists in outperform these three implementation by changing several parameters of the network (probability, number of neurons, number of layers, activation function, number of epochs and batch size). You should not change the optimizer.

Your submission should consist in a python code (Keras+TensorFlow) of the best configuration for the three approaches and a report showing the final architecture and the evolution of accuracy and loss for each experiment.