**W251 HW5**

**Hui Michelle Sun**

**Section 1. Monday 6:30pm**

1. **What is TensorFlow? Which company is the leading contributor to TensorFlow?**

From Wikipedia. ‘TensorFlow is a free and open-source software library for dataflow and differentiable programming across a range of tasks. It is a symbolic math library, and is also used for machine learning applications such as neural networks’.. Google

1. **What is TensorRT? How is it different from TensorFlow?**

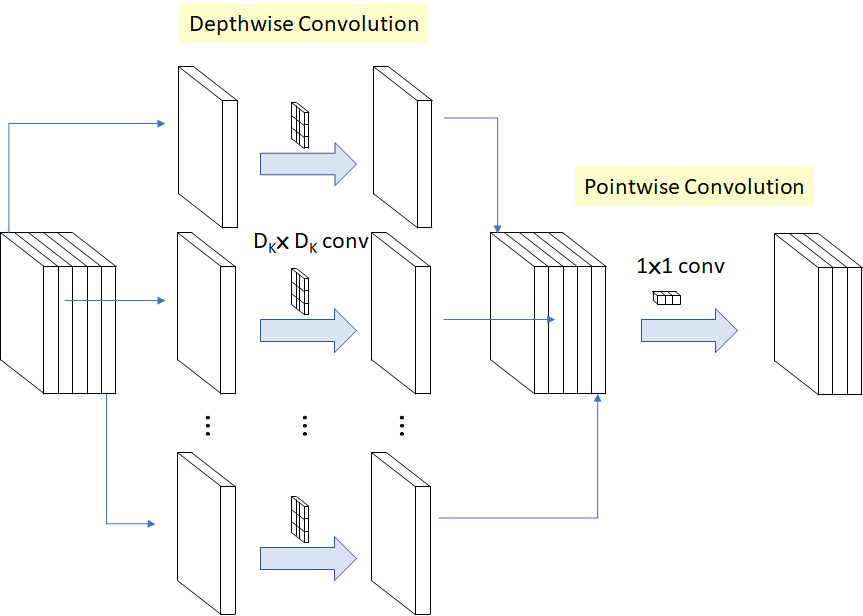
From nvida website ‘TensorRT-based applications perform up to 40x faster than CPU-only platforms during inference. With TensorRT, you can optimize neural network models trained in all major frameworks, calibrate for lower precision with high accuracy, and finally deploy to hyperscale data centers, embedded, or automotive product platforms. ensorRT is built on CUDA, NVIDIA’s parallel programming model, and enables you to optimize inference for all deep learning frameworks leveraging libraries, development tools and technologies in CUDA-X for artificial intelligence, autonomous machines, high-performance computing, and graphics.’

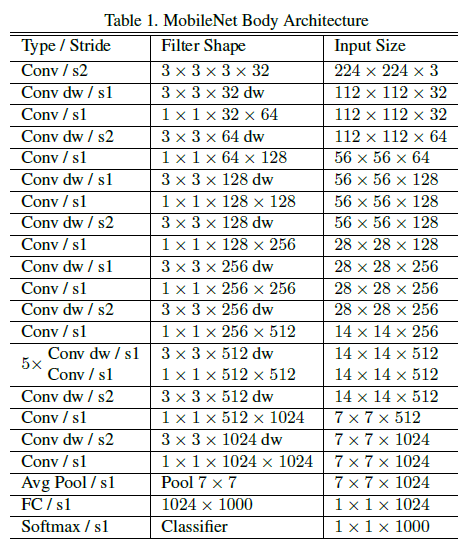
1. **What is ImageNet? How many images does it contain? How many classes?**

From Wikipedia ‘The ImageNet project is a large visual [database](https://en.wikipedia.org/wiki/Database) designed for use in [visual object recognition software](https://en.wikipedia.org/wiki/Outline_of_object_recognition) research. More than 14 million images have been hand-annotated by the project to indicate what objects are pictured and in at least one million of the images, bounding boxes are also provided. ImageNet contains more than 20,000 categories with a typical category, such as "balloon" or "strawberry", consisting of several hundred images. The database of annotations of third-party image [URLs](https://en.wikipedia.org/wiki/URL) is freely available directly from ImageNet, though the actual images are not owned by ImageNet.’

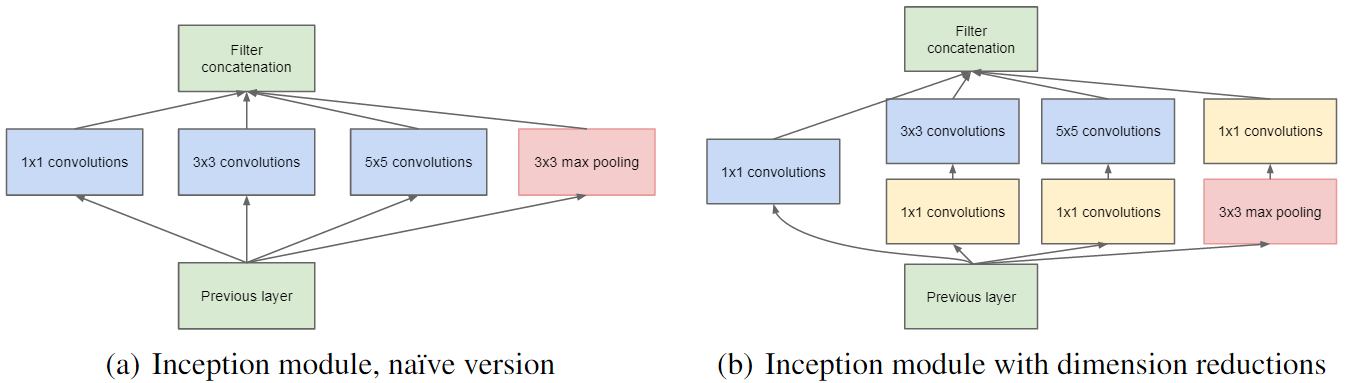
1. **Please research and explain the differences between MobileNet and GoogleNet (Inception) architectures.**

MobileNet: reference <https://towardsdatascience.com/review-mobilenetv1-depthwise-separable-convolution-light-weight-model-a382df364b69>





Inception



1. **In your own words, what is a bottleneck?**

Bottleneck the previous layers of the neural network before the classification layer. The fact that it is called bottleneck does not mean that it will create any delay for the whole process, it is just referring to the previous layer of the neural network (since the last classification layer is more compact then the rest of the layers).

1. **How is a bottleneck different from the concept of layer freezing?**

Bottleneck refer to sequence of layer before the classification layer. And the layer freezing is used during the deep learning, we freeze one or more layers to conduct training.

1. **In the TF1 lab, you trained the last layer (all the previous layers retain their already-trained state). Explain how the lab used the previous layers (where did they come from? how were they used in the process?)**

The previous layers are the model weights from the pre-trained model. Then we use a classification layer added on top of the pre-trained model (previous layers). This is the concept of transfer learning.

1. **How does a low --learning\_rate (step 7 of TF1) value (like 0.005) affect the precision? How much longer does training take?**

The Default Learning Rate (LR)=0.01

When LR=0.005, Validation accuracy is 0.91, test accuracy=0.93

Time spent is 1m31sec.

1. **How about a --learning\_rate (step 7 of TF1) of 1.0? Is the precision still good enough to produce a usable graph?**

When LR=1.0, the validation accuracy is 0.88, and the test accuracy = 0.92, time spent is 1m20sec. Is accuracy is good enough to produce a useable graph.

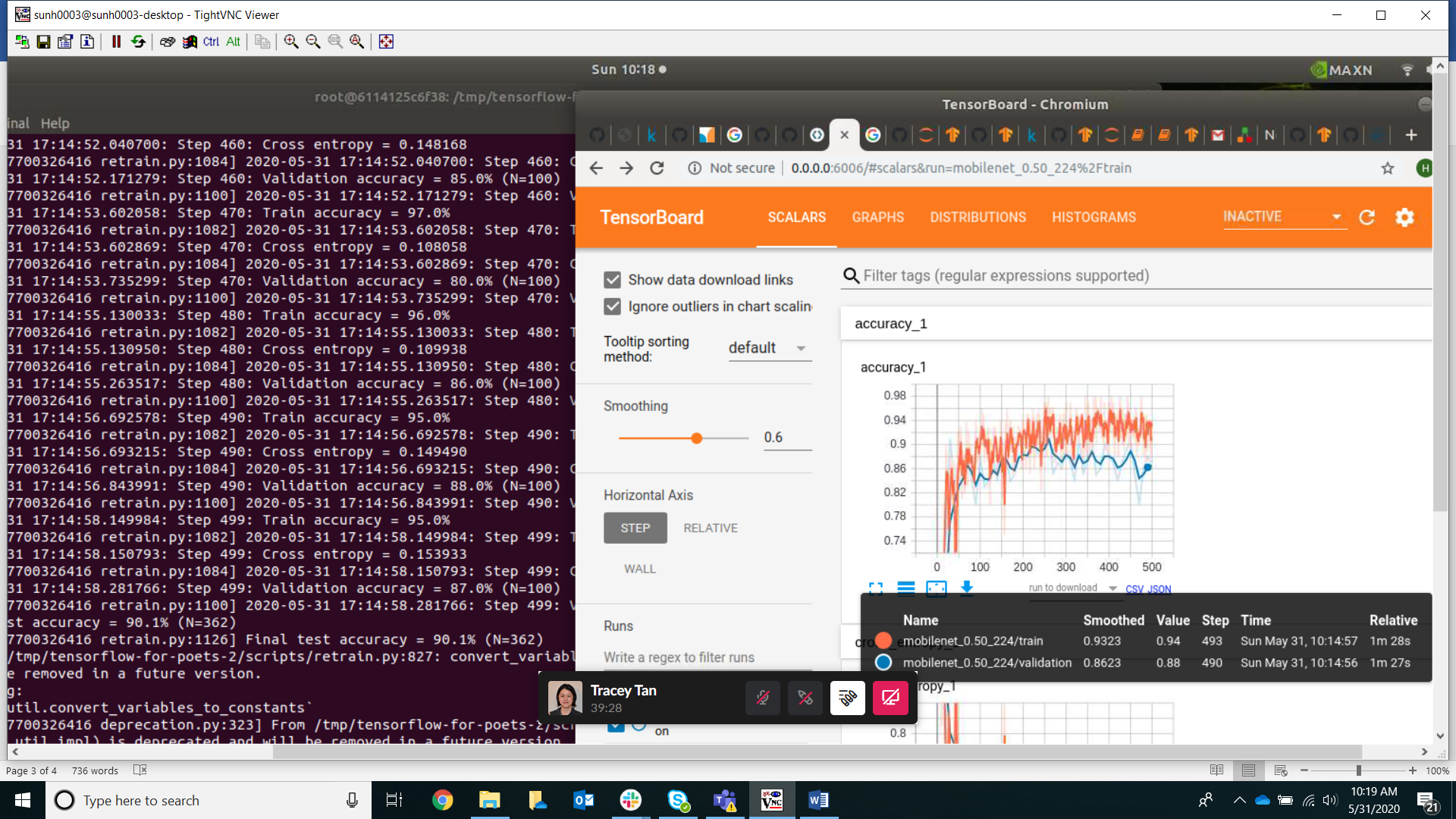
1. For step 8, you can use any images you like. Pictures of food, people, or animals work well. You can even use [ImageNet](http://www.image-net.org/) images. How accurate was your model? Were you able to train it using a few images, or did you need a lot?

I used Caltech bird images libaray. 2 classes. Each class there are around 50 photos.

The train accuracy is 99% but the test accuracy is low only 77%. GPU time=1m22sec

1. Run the TF1 script on the CPU (see instructions above) How does the training time compare to the default network training (section 4)? Why?

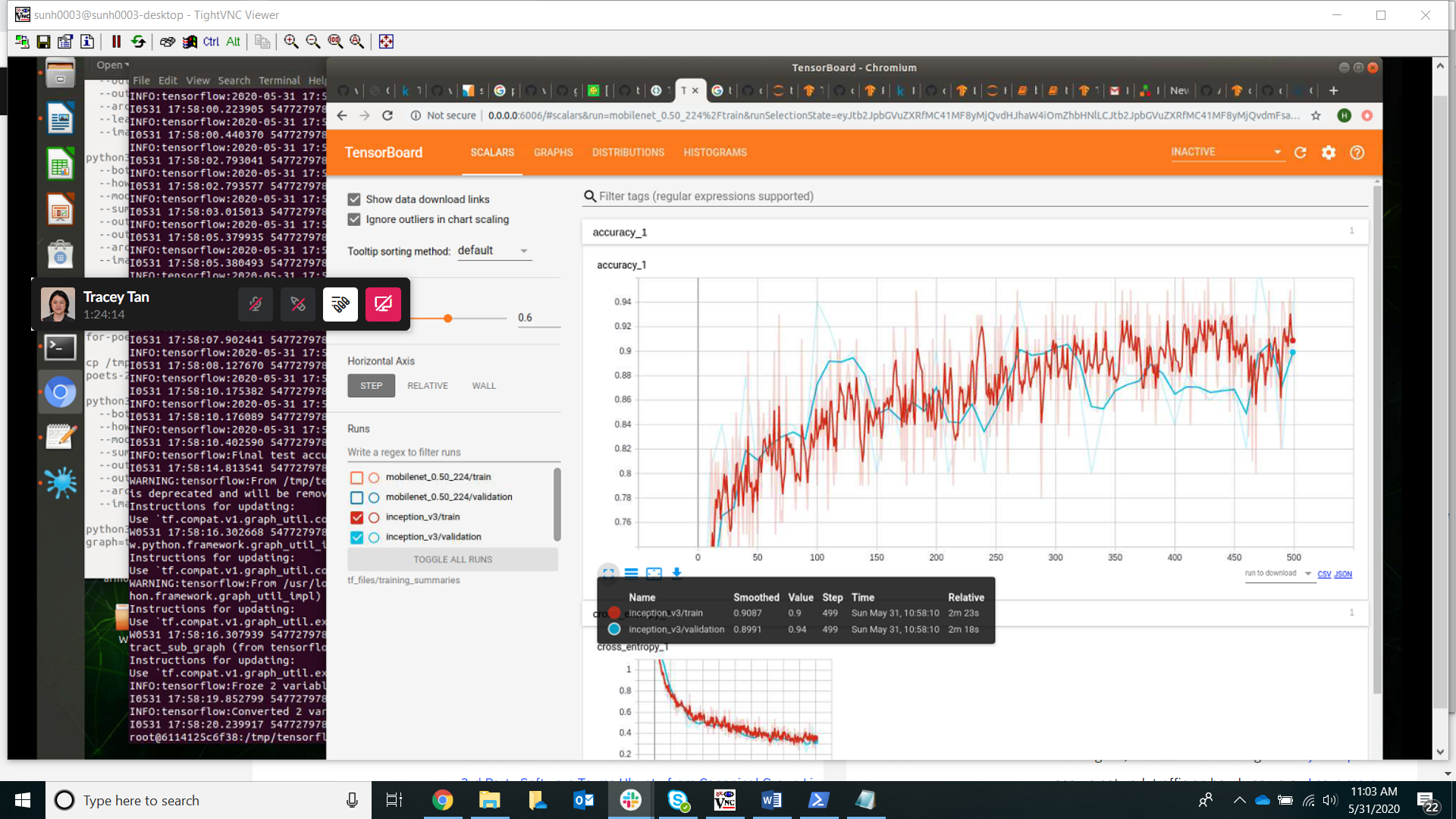
CPU mobilenet time=1m30sec, train accuracy=0.91, test accuracy=0.87



For mobileNet, it seems that the GPU and CPU time are comparable.

1. Try the training again, but this time do export ARCHITECTURE="inception\_v3" Are CPU and GPU training times different?

CPU with inception\_v3, train accuracy=0.91, test accuracy=0.89, time=2m30sec



GPU with inception\_v3, time= 2m 53sec

Train accuracy=0.90 test accuracy=0.86

1. Given the hints under the notes section, if we trained Inception\_v3, what do we need to pass to replace ??? below to the label\_image script? Can we also glean the answer from examining TensorBoard?

Input\_height=299

Input\_width-299

Input\_layer=Mul

python -m scripts.label\_image --input\_layer=Mul --input\_height=299 --input\_width=299 --graph=tf\_files/retrained\_graph.pb --image=tf\_files/flower\_photos/daisy/21652746\_cc379e0eea\_m.jpg