Problem 2:

Data collected in coordination with Param Popat and Karthik Subramanian
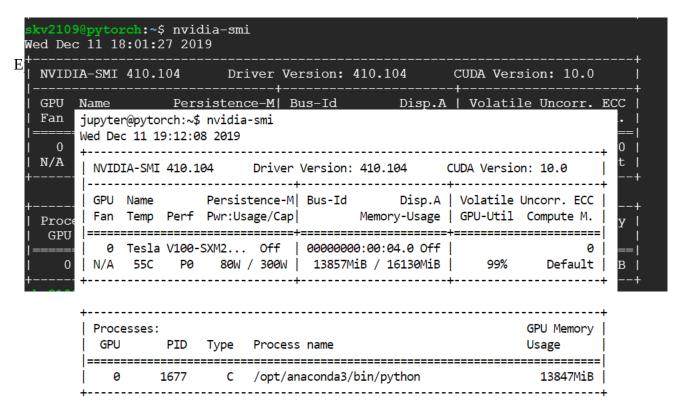
1,2. Coeffecient of Variation: Ratio of variance to the mean, and wild images accuracy

Run #	Machine	TTA (91%) (hrs)	Accuracy on Wild Images
1.	V100	4.581	92%
2.		4.033	90%
3.		3.833	88%
4.		4.433	90%
5.		3.667	98%
Mean		4.110	90
Std Dev		0.381	2.608
Coeff of Variance		9.47%	-
1.	TPU	3.012	88
2.		3.050	86
3.		3.117	92
4.		3.565	92
5.		3.667	90
Mean		92	90
Std Dev		3.85	2.608
Coeff of Variance		4.20	-

The coeffecient of variance values obtained are similar to that of the paper:

S.No	Machine Type	Our Value	ColeMan et al
1.	V100	9.47	11.2
2.	TPU	4.2	4.5

3. GPU utilization varied between 85% and 100% throughout the phases of training.
Epoch 10:

```
skv2109@pytorch:~$ nvidia-smi
Wed Dec 11 18:01:27 2019
+-----------------------------------------------------------------------------+
| NVIDIA-SMI 410.104      Driver Version: 410.104      CUDA Version: 10.0      |
|-------------------------------+----------------------+----------------------+
| GPU  Name        Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  jupyter@pytorch:~$ nvidia-smi                                           |
|===== Wed Dec 11 19:12:08 2019                                              0 |
|   0  +-----------------------------------------------------------------------------+
| N/A  | NVIDIA-SMI 410.104      Driver Version: 410.104      CUDA Version: 10.0    | t |
+-----  |-------------------------------+----------------------+----------------------+
        | GPU  Name        Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Proce | Fan  Temp  Perf  Pwr:Usage/Cap|         Memory-Usage | GPU-Util  Compute M. | y |
|  GPU  |===============================+======================+======================|
|=====  |   0  Tesla V100-SXM2...  Off  | 00000000:00:04.0 Off |                    0 |==|
|    0  | N/A   55C    P0    80W / 300W |  13857MiB / 16130MiB |     99%      Default | B |
+-----  +-------------------------------+----------------------+----------------------+

        +-----------------------------------------------------------------------------+
        | Processes:                                                       GPU Memory |
        |  GPU       PID   Type   Process name                             Usage      |
        |=============================================================================|
        |    0      1677      C   /opt/anaconda3/bin/python                  13847MiB |
        +-----------------------------------------------------------------------------+
```

Epoch 150:

```
jupyter@pytorch:~$ nvidia-smi
Wed Dec 11 20:37:27 2019
+-----------------------------------------------------------------------------+
| NVIDIA-SMI 410.104      Driver Version: 410.104      CUDA Version: 10.0      |
|-------------------------------+----------------------+----------------------+
| GPU  Name        Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp  Perf  Pwr:Usage/Cap|         Memory-Usage | GPU-Util  Compute M. |
|===============================+======================+======================|
|   0  Tesla V100-SXM2...  Off  | 00000000:00:04.0 Off |                    0 |
| N/A   56C    P0    90W / 300W |  13857MiB / 16130MiB |    100%      Default |
+-------------------------------+----------------------+----------------------+

+-----------------------------------------------------------------------------+
| Processes:                                                       GPU Memory |
|  GPU       PID   Type   Process name                             Usage      |
|=============================================================================|
|    0      1677      C   /opt/anaconda3/bin/python                  13847MiB |
+-----------------------------------------------------------------------------+
```

4. If the GPU utilization was indeed low, one measure we can take is to increase the batch size. This increases the amount of computation on the GPU. But we should be wary of the overall impact this has on our accuracy in itself.
If we do do so, we should also increase the learning rate as the batch size increases. This ensures the time to convergence will be in the same order of the initial setup.