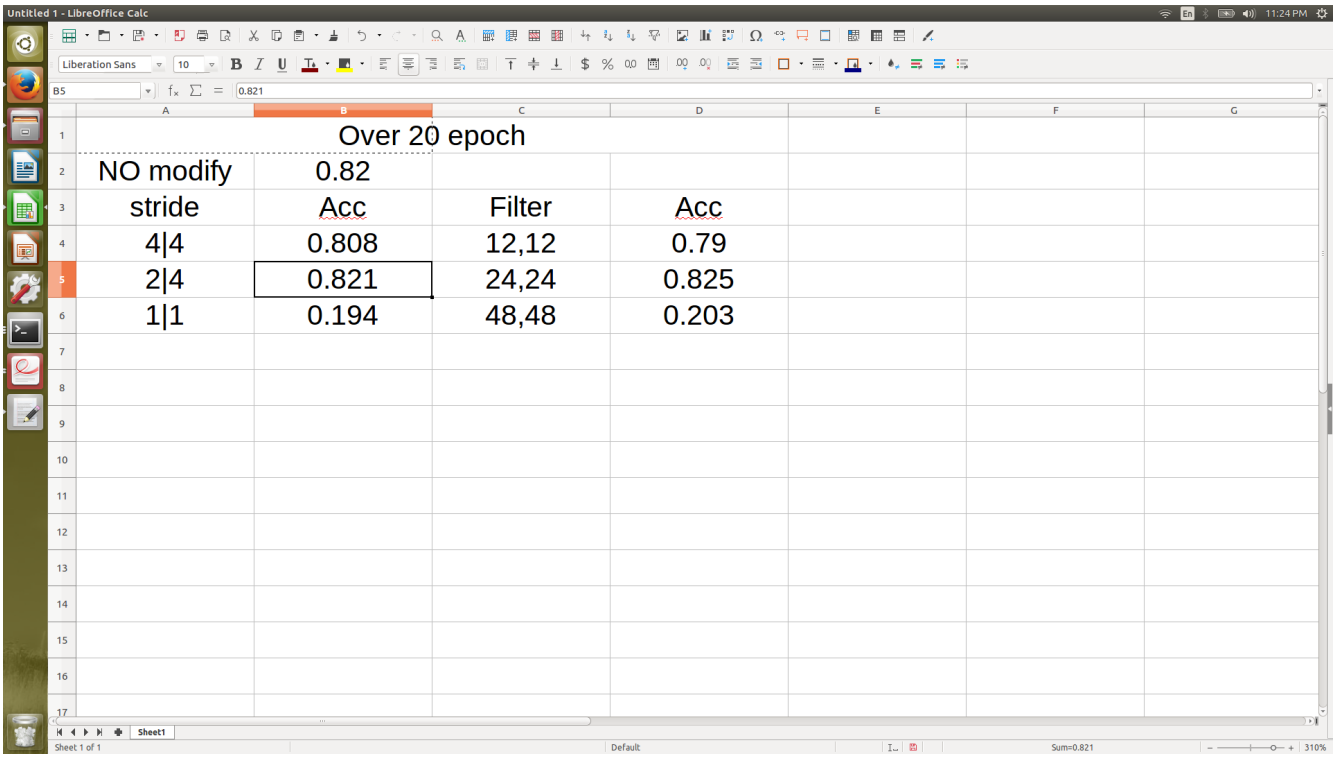
Summery:

Part1:
Using too many filter and small polling stride will waste more time on training and it cause a bad learning result.

Use small filter will provide not enough information of the picture. So It also will not provide good result.

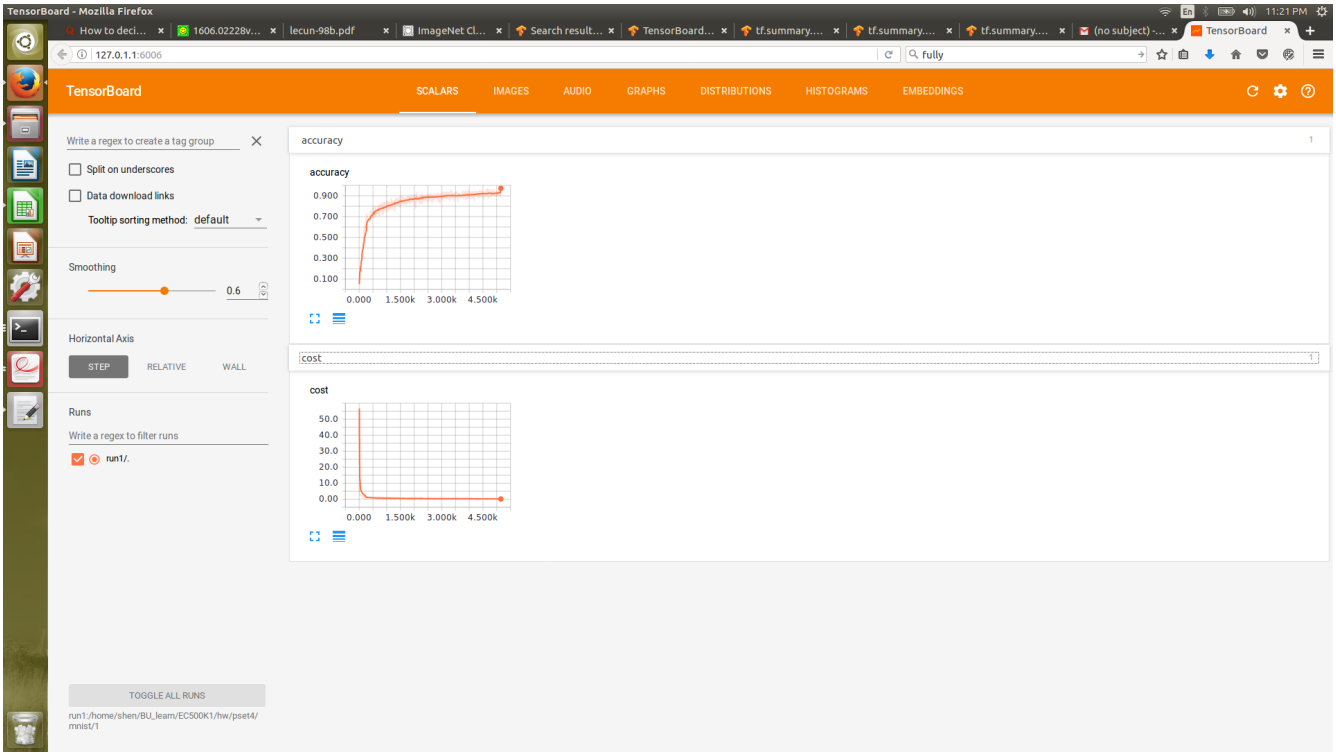


Part2:
The save and load training get a good performance quickly(85% at epoch 10, which is better than the cnn-map model, with 81% acc over 50 epoch).
The pre-training will generate a good initialize of the parameter at lower level. And the shape and structure of the higher lever would not affect the initialization too much.

Part3:
the normal training get 33% accuracy, the pre-load training get only 27.7%.
This result may reveal that the training parameter for different model is not compact.

Part4:

The Filters variable is hard to get.
I only plot the accuracy and cost.



Part 5:

I modify the model and the training file.
The model1 implement:
1. early stopping on training step
2. mini batch normalization, modify the batch size
3. two fully connected layer
4. dropped out on fully connected layer.
5. 2- layer pre-trainning 3 layer CNN

The model 2:
1. adjust the kernel and pooling unit size and number.
2. 4 layer CNN
3. MLP with more hidden unit

The original test accuracy is 82%, I have improve it to 88%.