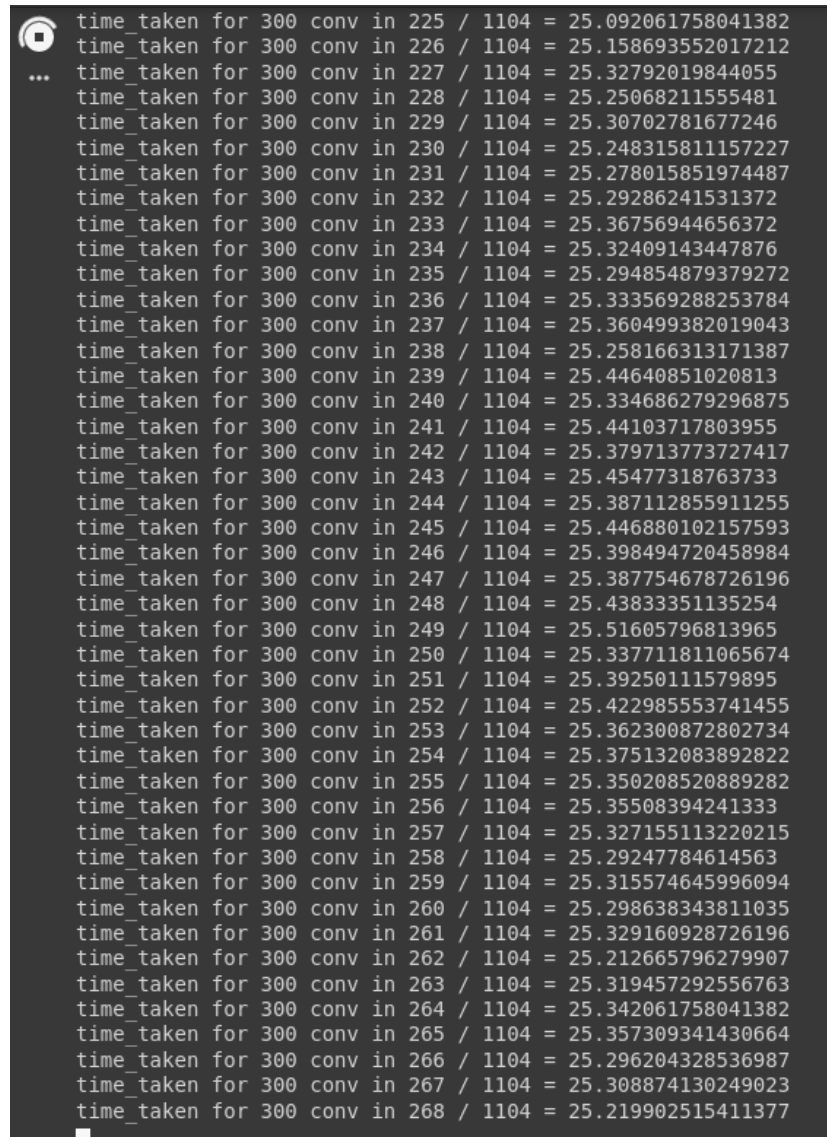


## DenseNet-161 Inference Update 16th March

- My work today involved optimizing the lookup part in convolution. As per last night's implementation, the time taken per convolution operation was 2 seconds which is now optimized to ~83ms per convolution. The alternate approach (implemented in *customconv\_v2.py*) takes ~25 seconds for 300 convolution operations on Google Colab (as per the img below). I have given a run for (conv2 layer, 1 test img, Google Colab) and it will take ~7.5



```
time_taken for 300 conv in 225 / 1104 = 25.092061758041382
time_taken for 300 conv in 226 / 1104 = 25.158693552017212
...
time_taken for 300 conv in 227 / 1104 = 25.32792019844055
time_taken for 300 conv in 228 / 1104 = 25.25068211555481
time_taken for 300 conv in 229 / 1104 = 25.30702781677246
time_taken for 300 conv in 230 / 1104 = 25.248315811157227
time_taken for 300 conv in 231 / 1104 = 25.278015851974487
time_taken for 300 conv in 232 / 1104 = 25.29286241531372
time_taken for 300 conv in 233 / 1104 = 25.36756944656372
time_taken for 300 conv in 234 / 1104 = 25.32409143447876
time_taken for 300 conv in 235 / 1104 = 25.294854879379272
time_taken for 300 conv in 236 / 1104 = 25.333569288253784
time_taken for 300 conv in 237 / 1104 = 25.360499382019043
time_taken for 300 conv in 238 / 1104 = 25.258166313171387
time_taken for 300 conv in 239 / 1104 = 25.44640851020813
time_taken for 300 conv in 240 / 1104 = 25.334686279296875
time_taken for 300 conv in 241 / 1104 = 25.44103717803955
time_taken for 300 conv in 242 / 1104 = 25.379713773727417
time_taken for 300 conv in 243 / 1104 = 25.45477318763733
time_taken for 300 conv in 244 / 1104 = 25.387112855911255
time_taken for 300 conv in 245 / 1104 = 25.446880102157593
time_taken for 300 conv in 246 / 1104 = 25.398494720458984
time_taken for 300 conv in 247 / 1104 = 25.387754678726196
time_taken for 300 conv in 248 / 1104 = 25.43833351135254
time_taken for 300 conv in 249 / 1104 = 25.51605796813965
time_taken for 300 conv in 250 / 1104 = 25.337711811065674
time_taken for 300 conv in 251 / 1104 = 25.39250111579895
time_taken for 300 conv in 252 / 1104 = 25.422985553741455
time_taken for 300 conv in 253 / 1104 = 25.362300872802734
time_taken for 300 conv in 254 / 1104 = 25.375132083892822
time_taken for 300 conv in 255 / 1104 = 25.350208520889282
time_taken for 300 conv in 256 / 1104 = 25.35508394241333
time_taken for 300 conv in 257 / 1104 = 25.327155113220215
time_taken for 300 conv in 258 / 1104 = 25.29247784614563
time_taken for 300 conv in 259 / 1104 = 25.315574645996094
time_taken for 300 conv in 260 / 1104 = 25.298638343811035
time_taken for 300 conv in 261 / 1104 = 25.329160928726196
time_taken for 300 conv in 262 / 1104 = 25.212665796279907
time_taken for 300 conv in 263 / 1104 = 25.319457292556763
time_taken for 300 conv in 264 / 1104 = 25.342061758041382
time_taken for 300 conv in 265 / 1104 = 25.357309341430664
time_taken for 300 conv in 266 / 1104 = 25.296204328536987
time_taken for 300 conv in 267 / 1104 = 25.308874130249023
time_taken for 300 conv in 268 / 1104 = 25.219902515411377
```

hours to complete the inference over 1 image. Will give an update regarding this by tomorrow.

- The run that I gave for yesterday night - (conv2 layer, 1 test img, google colab) completed 41 iterations of **i**, that is  $41 \times 300 = 12300$  convolutions in ~7hrs => it took ~10 minutes for 300 convolutions. It was too slow, since in order to complete all the convolution operations it would take much more than 24 hrs. So I worked on an alternate approach to reduce lookup time.
- Will provide further updates regarding error metrics after the inference run gets over by tomorrow.
- Please refer to the codes present in shared Google Drive folder. Following are some commonly used variables to measure time taken for convolution and multiplication operation.
  - Input size = [1,2208,15,20]; kernel size = [1104,2208,1,1]
  - out\_size= [1,1104,15,20]
  - Layer applied on: **conv2** in Decoder
  - Total multiplications = 731289600; Total conv= 331200;
  - $i \rightarrow [0:1103]$ ;  $j \rightarrow [0:299]$ ;  $k \rightarrow [0:2207]$ ;
  - $i*j*k$  = total multiplications ;  $i*j$  = total convolutions
  - 1 iteration of  $i \Rightarrow$  300 iterations of  $j$ ; 1 iter of  $j \Rightarrow$  2208 multiplications
- I would like to go over the code implementation with you. Can we have a call tomorrow at a time of your convenience?

Google drive link:

[https://drive.google.com/drive/folders/1GqTDwZa4CxLfEdyPjWa\\_JdZ9B7kfzQim?usp=sharing](https://drive.google.com/drive/folders/1GqTDwZa4CxLfEdyPjWa_JdZ9B7kfzQim?usp=sharing)

Google colab sheet:

<https://colab.research.google.com/drive/1hR70s85isOyWzyo0-o8970rDvmBcUz6R?usp=sharing>

