Update_21st_March_2

1. A basic idea of the working of convolution operation in the decoder.conv2 layer.

In feature = [1,2208,15,20] -> [1,2208*1*1,15*20] = [1,2208,300]
kernel = [1104,2208,1,1] -> [1104,2208*1*1] = [1104,2208]

convolution = matmul(kernel, in_feature)

x = 1 batch of in_feature => x = [2208,300] ;
kernel = [1104,2208]
-->result = (kernel * x) {perform matmul}
result = [1104,300]

Standard matmul op pseudocode:

```
for i=0:1104:
    for j=0:300:
        for k=0:2208:
            result[i][j] += kernel[i][k] * x[k][j]
```

Time taken for inference:
12*1104 = 13248/3600 = 3.68 hrs

In_feature = the dimensions of the input feature map into the decoder.conv2 layer
Kernel = The weight values obtained from the pre-trained model *nyu_5000_e15.h5*

Matrix multiplication done over kernel and each batch of in_feature (x) ie [kernel * x] to compute convolution value.(result)

Each iteration consists of 2208*300 lookup operations ---> takes 12 seconds
Therefore, entire inference time = 1104*12/3600 = ~3.68 hrs (on avg)