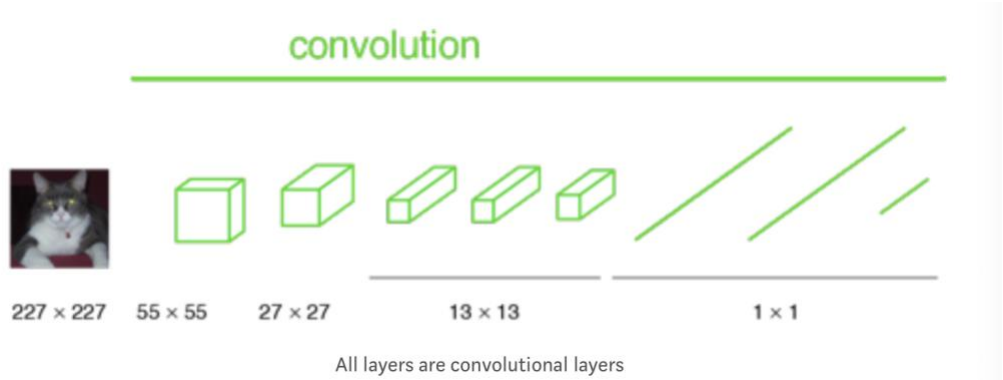
Team name: null

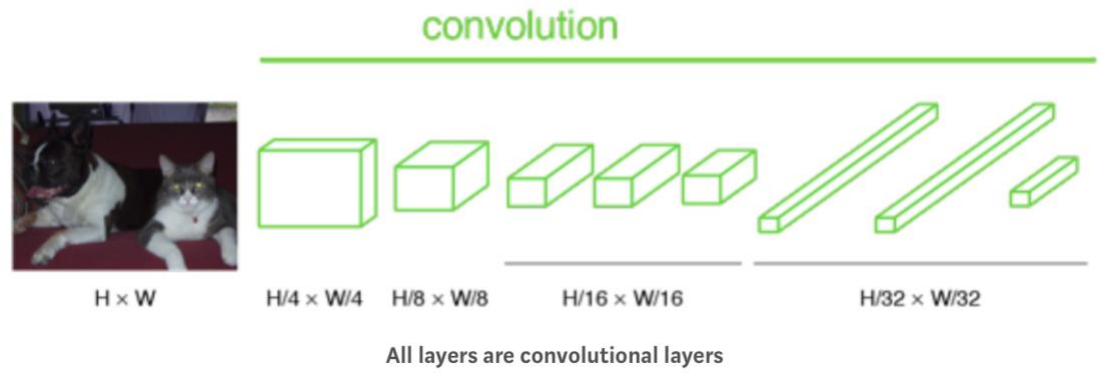
Team member: Miao Zhou, Kefei Wu

TASK3: Actor-Action Segmentation

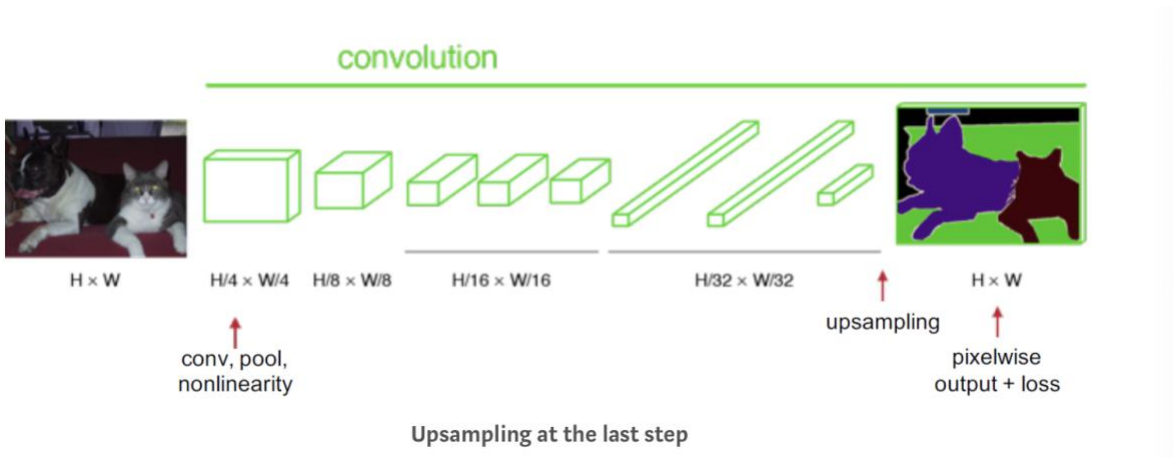
In this task, we choose to use FCN model as out baseline. FCN could take input of arbitrary size and produce pixel-wise predication. In classification, conventionally, an input image is downsized and goes through the convolution layers and fc layers, and output predicted label for the input image. In segmentation, we turn the fc layers into convolutional layers,

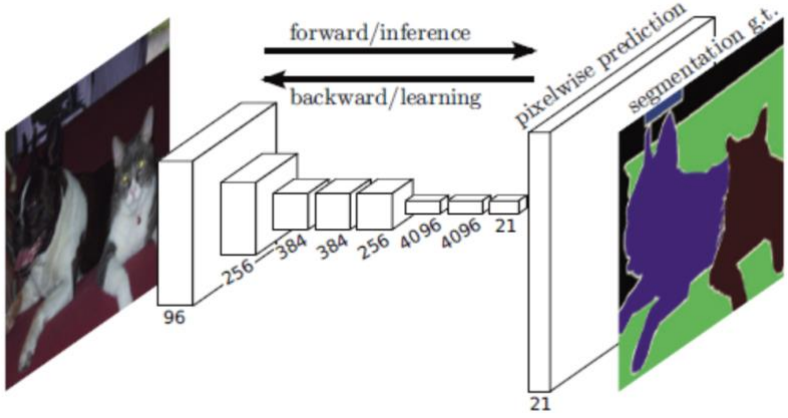


And if the image is not downsized, the output will not be a single label. Instead, the output has a size smaller than the input image(due to the max pooling) :



If we upsample the output above, then we can calculate the pixelwise output(label map) as below:





Feature Map / Filter Number Along Layers

After going through the conv, the output size is small, the output from pool5 is 2x unsampled and fuse with pool4 and perform 16x unsampling[1].

Our approaches does make use of pretrained model which is VGG. Loss function is cross_entropy2d which is based on the nn.NLLLoss, optimizer is SGD, the default epoch is 10 but we ran 50 epochs to train the model. the batch_size is 16. Initial lr is 1e-10. Due to the time limit, we didn't run for enough time to have the epoch for convergence.
The result of the val_set are as follows:

```
shuishui@pytorch-vm:~/csc249_final_proj/csc_249_final_proj_a2d_seg$ python eval.py --gt_label mask_gt.pkl --pred_label mask_FCN.pkl
    Accuracy Class: 0.3534882169078903
    Mean IoU: 0.25327972467648563
```

References:

[1] Review: FCN—Fully Convolutional Network (Semantic Segmentation): https://towardsdatascience.com/review-fcn-semantic-segmentation-eb8c9b50d2d1

[2] Jonathan Long, Evan Shelhamer, Trevor Darrell. Fully Convolutional Networks for Semantic Segmentation.2015 CVPR. 1411.4038v2.

[3] Tianxiang Pan, Bin Wang, Guiguang Ding, Jun-Hai Yong.Fully Convolutional Neural Networks with Full-Scale-Features for Semantic Segmentation.AAAiI-17.