



Kindly download the paper from given link:

https://www.cv-foundation.org/openaccess/content_cvpr_2015/papers/Long_Fully_Convolutional_Networks_2015_CVPR_paper.pdf

This paper describes Fully Convolutional Neural Networks (FCNNs) for semantic segmentation. Read this paper and answer the following questions

1. The advantage of Fully Convolutional Networks over conventional CNNs

- a. Only fixed sized input data can be passed network.
- b. Any arbitrary sized input data can be passed through network.
- c. The size of the output of FCNNs depends on the input size.
- d. FCNN is a computationally cheaper way of achieving semantic segmentation compared to conventional CNNs

Which of the above statements are true?

- A. a,b & d
- B. a, c & d
- C RIGHT. b, c & d
- D. a,b, c & d

1. I dont know option-d is correct or not. But I know option-a is wrong.

2 One issue in this specific FCN is that by propagating through several alternated convolutional and pooling layers, the resolution of the output feature maps is down sampled.

Therefore, the direct predictions of FCN are typically in low resolution, resulting in relatively fuzzy object boundaries. What methods are suggested by authors to gain original resolution?

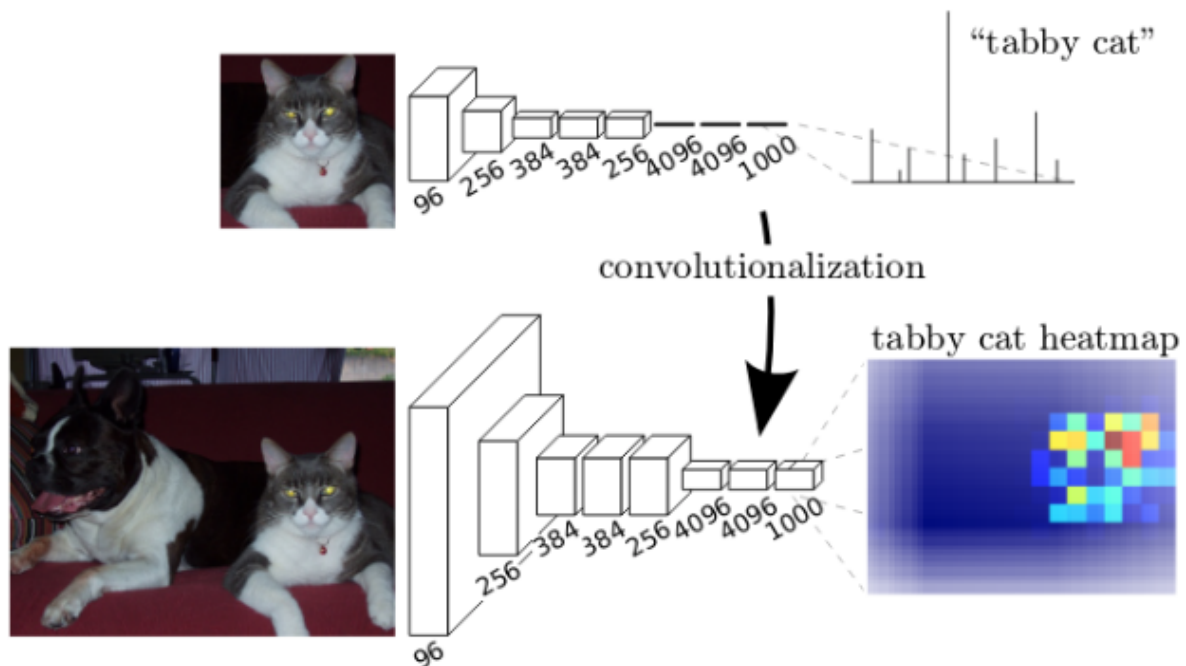
- A RIGHT. Bilinear Interpolation
- B RIGHT. Deconvolution
- C RIGHT. Backward convolution
- D. Forward convolution

Accepted Answers:

- Bilinear Interpolation
- Deconvolution
- Forward convolution

I think 'Forward convolution' is wrong. and it should be 'Backward convolution'

This figure has been taken from the paper mentioned above. It illustrates the transformation of a fully connected network into a fully convolutional network. Suppose the 5th layer of the FCN has feature maps of size 32x32.



What are the kernel sizes used in the final three (6th, 7th and 8th) convolutional layers of FCN?

- **A RIGHT.** 32x32x4096, 1x1x4096, 1x1x1000
- **B.** 16x16x8192, 2x2x2048, 2x2x2048
- **C.** 64x64x1024, 3x3x1000, 3x3x1000
- **D.** 28x28x4096, 3x3x1024, 2x2x1024

4. Which of following statement/s is/are True?

- **A RIGHT.** Decreasing subsampling within a net is a tradeoff: the filters see finer information, but have smaller receptive fields and take longer to compute.
- **B RIGHT.** The shift-and-stitch trick is another kind of tradeoff: the output is denser without decreasing the receptive field sizes of the filters, but the filters are prohibited from accessing information at a finer scale than their original design.
- **C RIGHT.** In this paper, network is trained with a per-pixel multinomial logistic loss.
- **D RIGHT.** Network is validated with the standard metric of mean pixel intersection over union, with the mean taken over all classes, including background.

5. According to the paper, optimization parameter/s used in training the FCN-VGG16 is/are:

- **A RIGHT.** Stochastic Gradient descent used as optimization algorithms
- **B RIGHT.** Learning rate is 10^{-4}
- **C.** For hyperparameter Grid Search has been used.
- **D RIGHT.** Dropout has been used in the network.
 - "Dropout was included where used in the original classifier nets.". Read the [VGG16 paper](#).

6. In the context of deep learning, transfer learning depends on which of the following variables:

- **A RIGHT.** Similarity of the data to original data used for pre-training model weights
- **B RIGHT.** Amount of data available
- **C RIGHT.** Computational power
- **D.** Complexity of data

Accepted Answers:

- Similarity of the data to original data used for pre-training model weights
- Amount of data available
- Computational power
- Complexity of data

7. In this context of Transfer Learning which of the following statements are True?

- **A RIGHT.** Transfer learning enables feature extraction with pre-trained deep learning models.
- **B RIGHT.** When target labels are scarce, the weights of pre-trained models are frozen (fix weights) so as to avoid overfitting.
- **C RIGHT.** Fine-tuning of pre-trained model weights is generally preferred when target task labels are plentiful.
- **D RIGHT.** Transfer learning works better when the tasks on which the networks are trained for are similar.

8. Choose the correct statement with regard to different approaches used for hyper-parameter optimisation.

- **A.** Grid search approach is usually very efficient, and each new guess is independent of previous guess.
- **B RIGHT.** Random search approach is usually more efficient than grid search, and each new guess is independent of previous guess.
- **C.** In Bayesian optimisation, each new guess is independent of the previous guess.
- **D.** Both A and B.

9

Consider two different models:

$$\text{Model 1 : } \hat{y} := w_1 x + w_0$$

$$\text{Model 2 : } \hat{y} := \sum_{i=1}^n w_i x^i + w_0$$

What should be the value of regularisation parameter (i.e., hyper parameter)?

- **A.** High for model 1, low for model 2

- **B RIGHT.** Low for model 1, high for model 2
 - **C.** Low for both models
 - **D.** High for both models
1. The last two options won't make any sense since the low and high are relative terms and will make sense only if that is comparing to something.
 2. Here Model-2 is complex model comparing to model-1. The regularisation parameter is for controlling the model complexity. So Model-2 need high value.