Discussion Quiz #6

Ashutosh Tiwari (ashutiwa@iu.edu)

March 7, 2022

1 Papers

1.1 Paper 1

UNet

1.2 Paper 2

DenseNet

2 Questions

2.1 List four advantages of DenseNet discussed in paper 2. (2pts)

Paper mentions these four advantages:

1. they alleviate vanishing gradient problem
2. strengthen feature propagation
3. encourage feature reuse
4. substantially reduce the number of parameters

2.2 How DenseNet connections are different from ResNet connections? Why the author chose one over the other, explain. (3pts)

ResNet has skip connections from one layer to the next layer in a residual block. In contrast Dense has these connections from each layer to each layer in a Dense block.

Authors do this to maximize information flow between layers in the network. Also instead of summing these features before they are passed into a layer in case of Resnet, Densenet combine features by concatenating them with the non-skip flow.

2.3 Explain why DenseNet requires fewer parameters than a traditional CNN. What are the advantages of having fewer parameters in DenseNet? (3pts)

It has fewer parameters because of its layers are very narrow, which add only a small set of feature maps to the "collective knowledge" of the network and keep remaining feature maps unchanged. The final result is the combination of all feature maps.

Advantages of fewer parameters are reduced under-fitting and ability to scale the depth of the network.

2.4 Explain how the DenseNet architecture leads to implicit deep supervision? (3pts)

Densenet architecture leads to implicit supervision because each layer has direct access to gradients from loss function from the last layer and original input from the initial layers. Therefore they can choose the feature maps which they want to learn from give these two and feature maps from other layers.

2.5 What does the author mean by, "desired output should include localization" in the context of paper 1. (3pts)

Authors mean that there are some tasks in biomedical domain which require labels for spatially localized i.e. recognized at a pixel level or set of pixel level.

2.6 Why does U-net double the number of features when the resolution of features is reduced by half? (3pts)

This architectural feature is followed by many deep learning networks. There are many possible reasons for this.

1. First is that as you increase number of channels you increase the capacity of model to learn more complex features logarithmically
2. second is the fact that maxpooling in accordance with channel size increase forgets all spatial information and learns the semantics, i.e. learns complex features.

2.7 What are the different purposes served by the feature connection from contracting path to expanding path as discussed in paper 1? (3pts)

This feature connection is a typical architecture of a convolutional network. Authors mention that Dropout layers in this connection contribute to data augmentation. Additionally because it does maxpooling it tries to throw away spatial features and tries to learn more semantic features. It seems that they do cropping as well in the same network.

2.8 Explain why an architecture like U-net is suitable with smaller datasets? (3pts)

Unet is suitable for small datasets because of its heavy use of data augmentation techniques. Authors talk about their augmentation techniques in detail in the paper. On top of all these, Dropouts layers perform implicit data augmentation.

2.9 What are the differences between DenseNet and GoogLeNet architectures? (3pts)

Differences between GoogLeNet and DenseNet are

1. GoogLeNet attempts to widen the netweork using feature maps produced by filters of different sizes while DenseNet incorporates many number of layers to go deeper.
2. GoogLeNet has no ski[connections while DenseNet has.
3. GoogLeNet does intermediary supervision between input and output layers, while DenseNet doesn't do any explicit such thing.

2.10 Describe briefly. (a) dense blocks, (b) transition layers (paper 2), (c)growth rate (paper 2), (d) compression factor (paper 2) (4pts)

1. **Dense Blocks**

   Dense blocks are the residual blocks in DenseNet. They are different from all other previous networks in the sense that all of them are connected via skip connections.

2. **Transition Layers**

   These are the layers which separate two dense blocks. They are a combination of a convolution and pooling layers.

3. **Growth Rate**

   Growth rate is a hyper-parameter which authors use to scale how much information is added to network each layer. Since we are concatenating feature maps, channel dimension is increasing at every layer. If we make a layer to produce k feature maps every time, then we can generalize for the l-th layer as

   $$k_l = k_o + k \times (l - 1)$$

4. **Compression Factor**

   This is a hyperparameter, θ which is used to regulate model compactness. Authors define its behavior such that if a dense block has x feature maps then the next transition layer will generate $\lfloor \theta x \rfloor$ output feature maps.

References