Fully Convolutional Network (FCN)

Industrial AI Lab.

Prof. Seungchul Lee

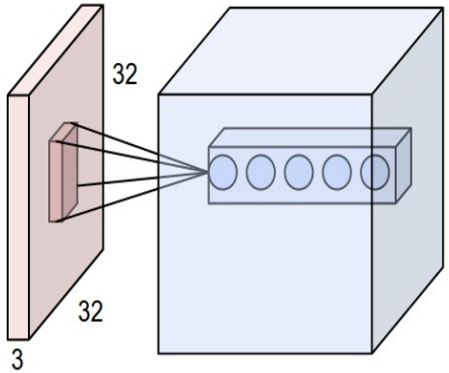
Deep Learning for Computer Vision: Review

Foundations

- Why computer vision?
- Representing images
- Convolutions for feature extraction

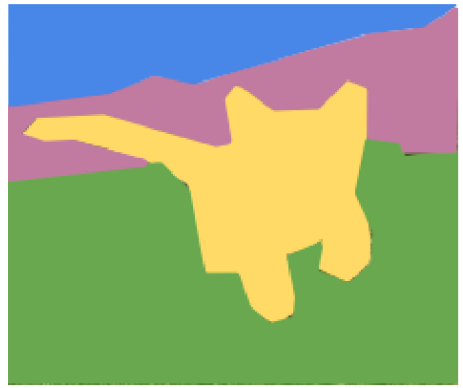
CNNs

- CNN architecture
- Application to classification: ImageNet



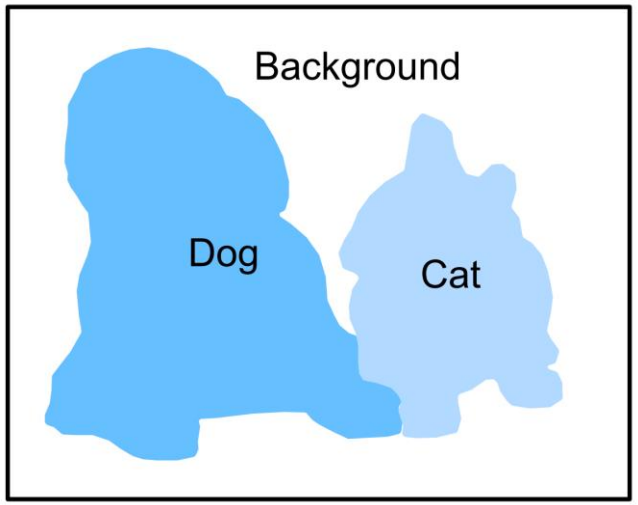
Applications

- Segmentation, object detection, image captioning
- Visualization

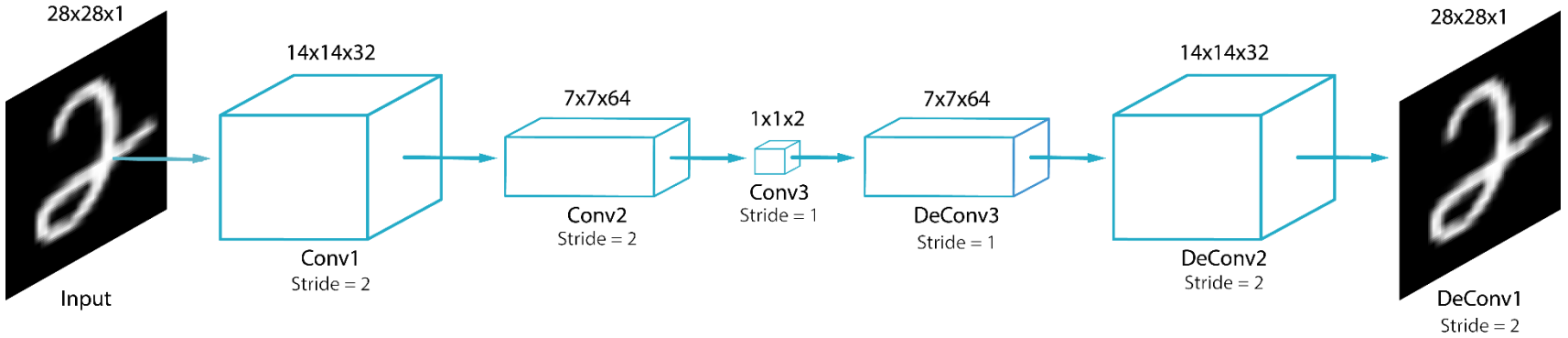


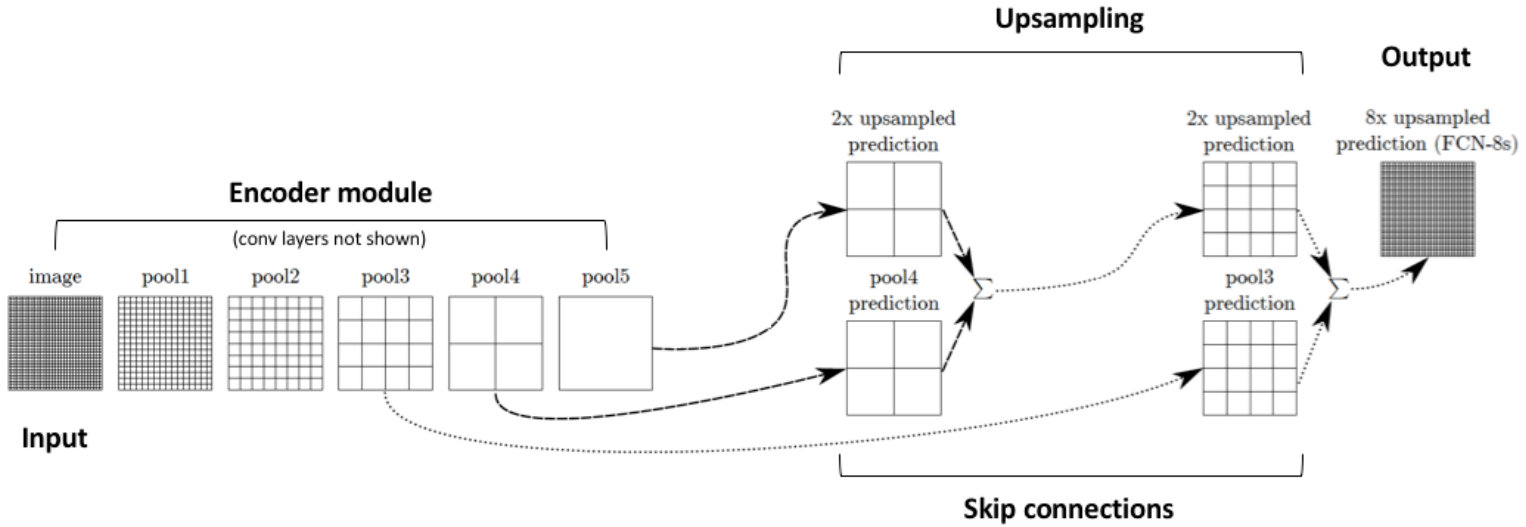
Segmentation

- Segmentation task is different from classification task because it requires predicting a class for each pixel of the input image, instead of only 1 class for the whole input.
- Classification needs to understand what is in the input (namely, the context).
- However, in order to predict what is in the input for each pixel, segmentation needs to recover not only what is in the input, but also where.
- Segment images into regions with different semantic categories. These semantic regions label and predict objects at the pixel level



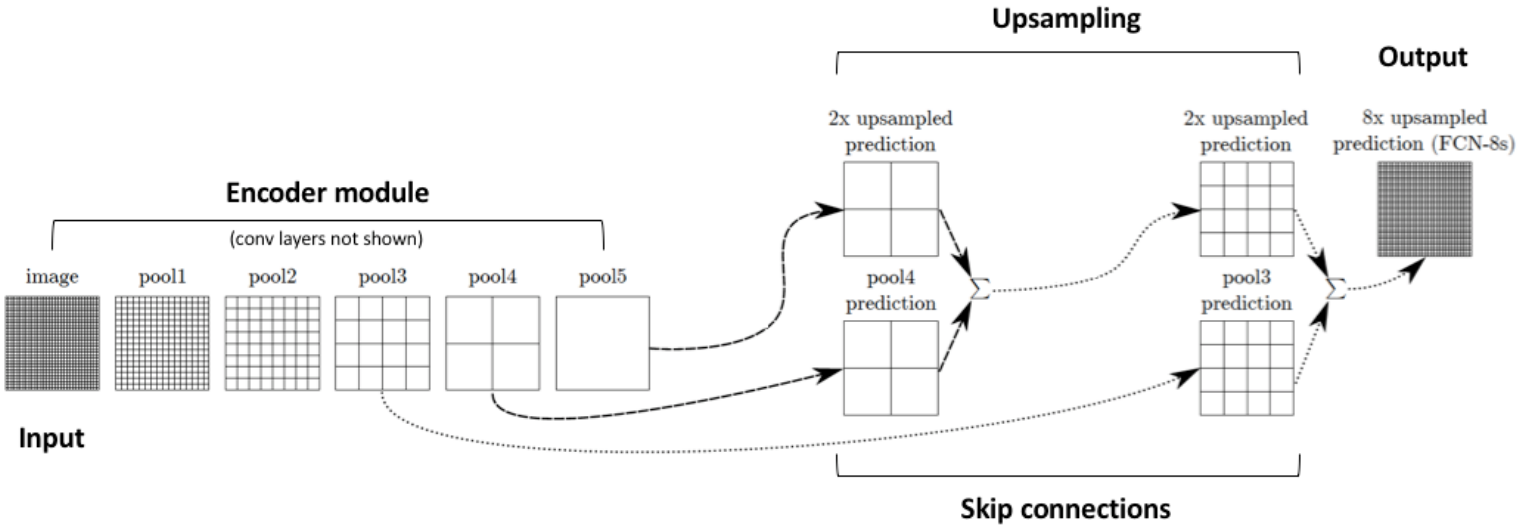
From CAE to FCN





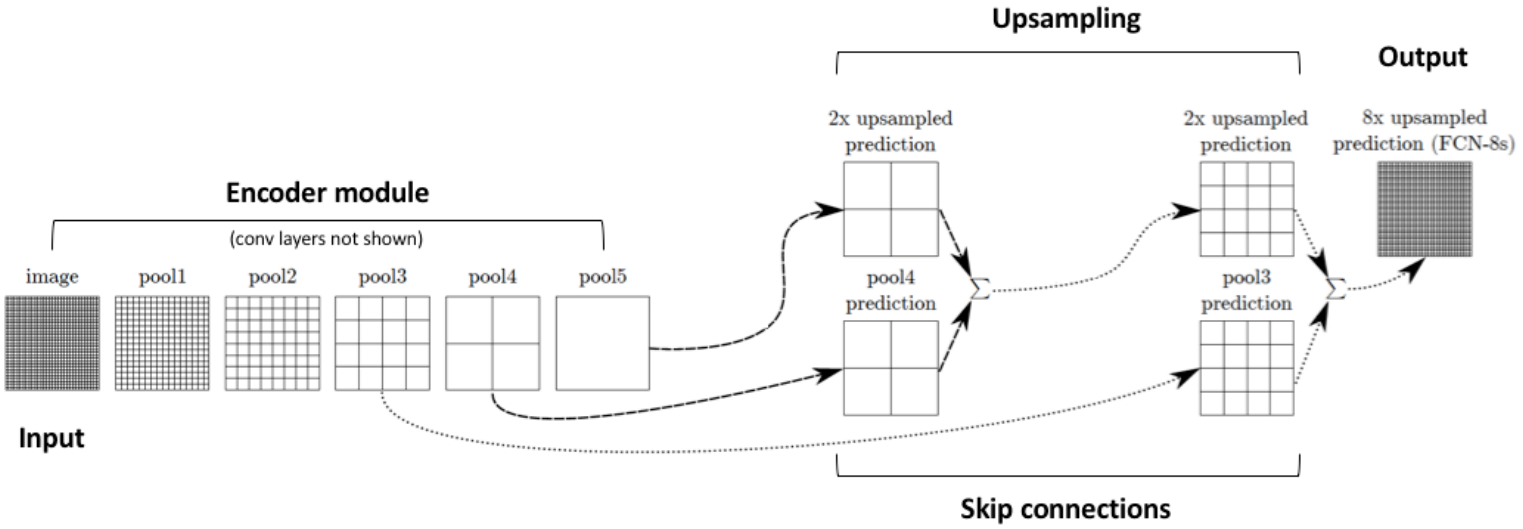
Fully Convolutional Networks (FCNs)

- To obtain a segmentation map (output), segmentation networks usually have 2 parts
  - Downsampling path: capture semantic/contextual information
  - Upsampling path: recover spatial information
- The downsampling path is used to extract and interpret the context (what), while the upsampling path is used to enable precise localization (where).
- Furthermore, to fully recover the fine-grained spatial information lost in the pooling or downsampling layers, we often use skip connections.



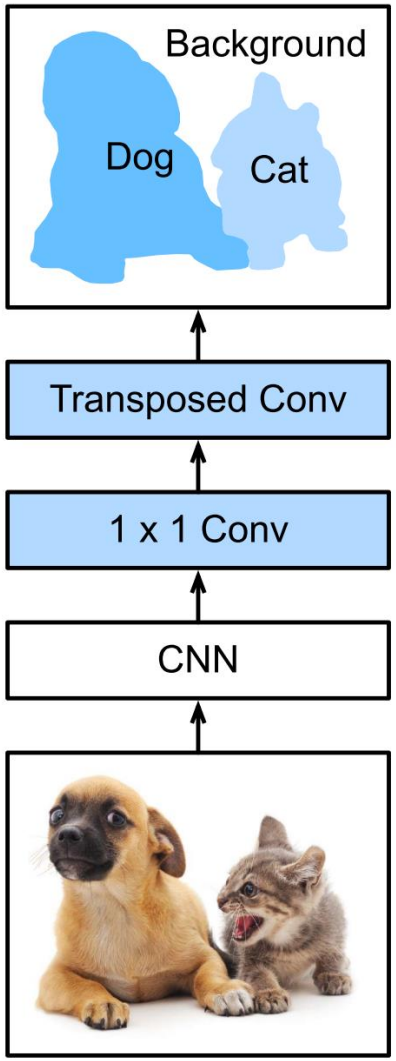
Fully Convolutional Networks (FCNs)

- FCN is built only from locally connected layers, such as convolution, pooling and upsampling.
- Note that no dense layer is used in this kind of architecture.
- Network can work regardless of the original image size, without requiring any fixed number of units at any stage.



FCN

- FCN uses a convolutional neural network to transform image pixels to pixel categories.
- FCN transforms the height and width of the intermediate layer feature map back to the size of input image through the transposed convolution layer, so that the predictions have a one-to-one correspondence with input image in spatial dimension
- Given a position on the spatial dimension, the output of the channel dimension will be a category prediction of the pixel corresponding to the location.

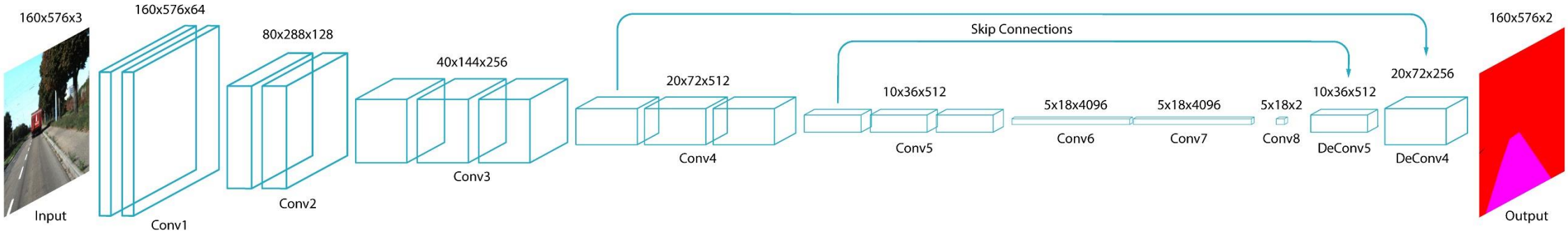


Skip Connection

- A skip connection is a connection that bypasses at least one layer.
- Here, it is often used to transfer local information by concatenating or summing feature maps from the downsampling path with feature maps from the upsampling path.
- Merging features from various resolution levels helps combining context information with spatial information.

Segmented (Labeled) Images

FCN Architecture



Segmentation Result