TTIC 31230, Fundamentals of Deep Learning

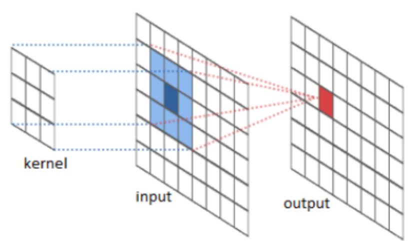
David McAllester, Winter 2020

Dilation, Hypercolumns, and Grouping

Dilation

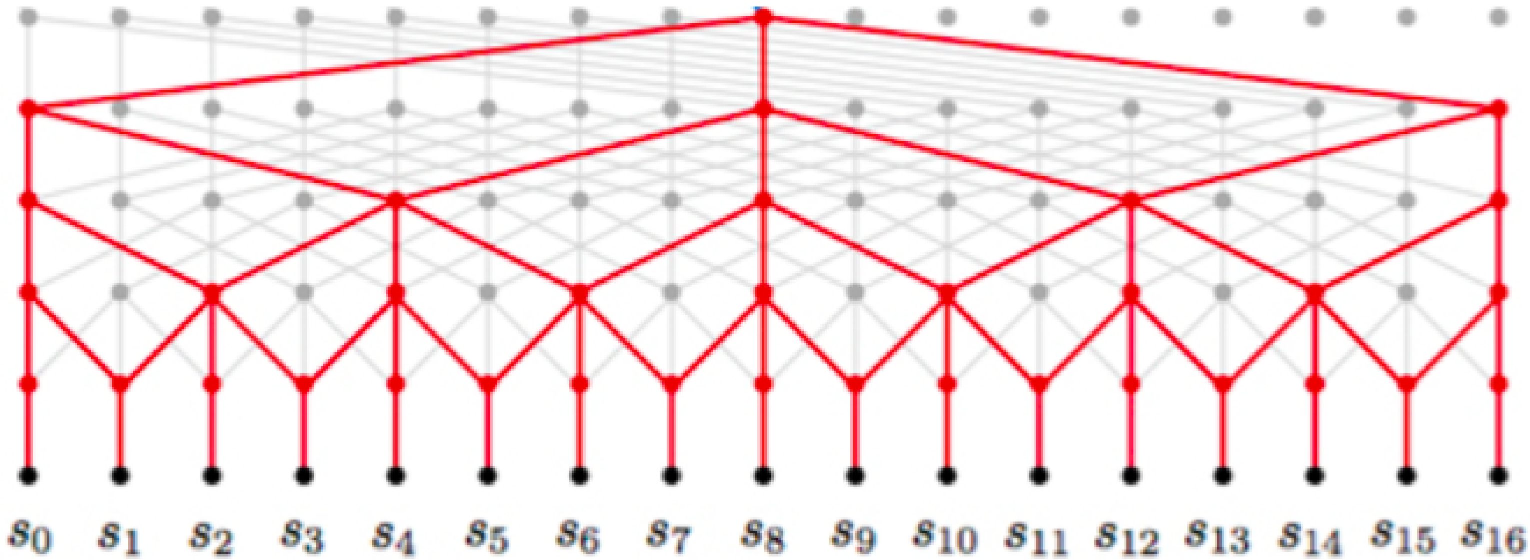
A CNN for image classification typically reduces an $N \times N$ image to a single feature vector.

Dilation is a trick for treating the whole CNN as a "filter" that can be passed over an $M \times M$ image with $M > N$.



An output tensor with full spatial dimension can be useful in, for example, image segmentation.

Dilation



This is called a “fully convolutional” CNN.

Dilation

To implement a fully convolutional CNN we can "dilate" the filters by a dilation parameter d.

$$L_{\ell+1}[b, x, y, j]$$

$$= \sigma(W[\Delta X, \Delta Y, I, j] L_\ell[b, x + d * \Delta X, y + d * \Delta Y, I] + B[j])$$

Vector Concatenation

We will write

$$L[b, x, y, J_1 + J_2] = L_1[b, x, y, J_1] \ ; L[b, x, y, J_2]$$

To mean that the vector $L[b, x, y, J_1 + J_2]$ is the concatenation of the vectors $L_1[b, x, y, J_1]$ and $L_2[b, x, y, J_2]$.

Hypercolumns

For a given image location $\langle x, y\rangle$ we concatenate all the feature vectors of all layers above the point $\langle x, y\rangle$.

$$
\begin{array}{rl}
& L\left[b, x, y, \sum_{\ell} J_\ell\right] \\
= & L_0\,[b, x, y, J_0] \\
& \vdots \\
& ; L_\ell\left[b, \left\lfloor x\left(\frac{X_\ell}{X_1}\right)\right\rfloor, \left\lfloor y\left(\frac{Y_\ell}{Y_0}\right)\right\rfloor, J_\ell\right] \\
& \vdots \\
& ; L_{\mathcal{L}-1}[b, J_{\mathcal{L}-1}]
\end{array}
$$

Grouping

The input features and the output features are each divided into G groups.

$$L_{\ell+1}[b,x,y,J] = L^0_{\ell+1}[b,x,y,J/G];\cdots;L^{G-1}_{\ell+1}[b,x,y,J/G]$$

where we have G filters $W^g[\Delta X, \Delta Y, I/G, J/G]$ with

$$L^g_{\ell+1}[b,x,y,j]$$

$$= \sigma(W^g[\Delta X, \Delta Y, I/G, j] L^g_\ell[x+\Delta X, y+\Delta Y, I/G, j] - B^g[j])$$

This uses a factor of G fewer weights.

END