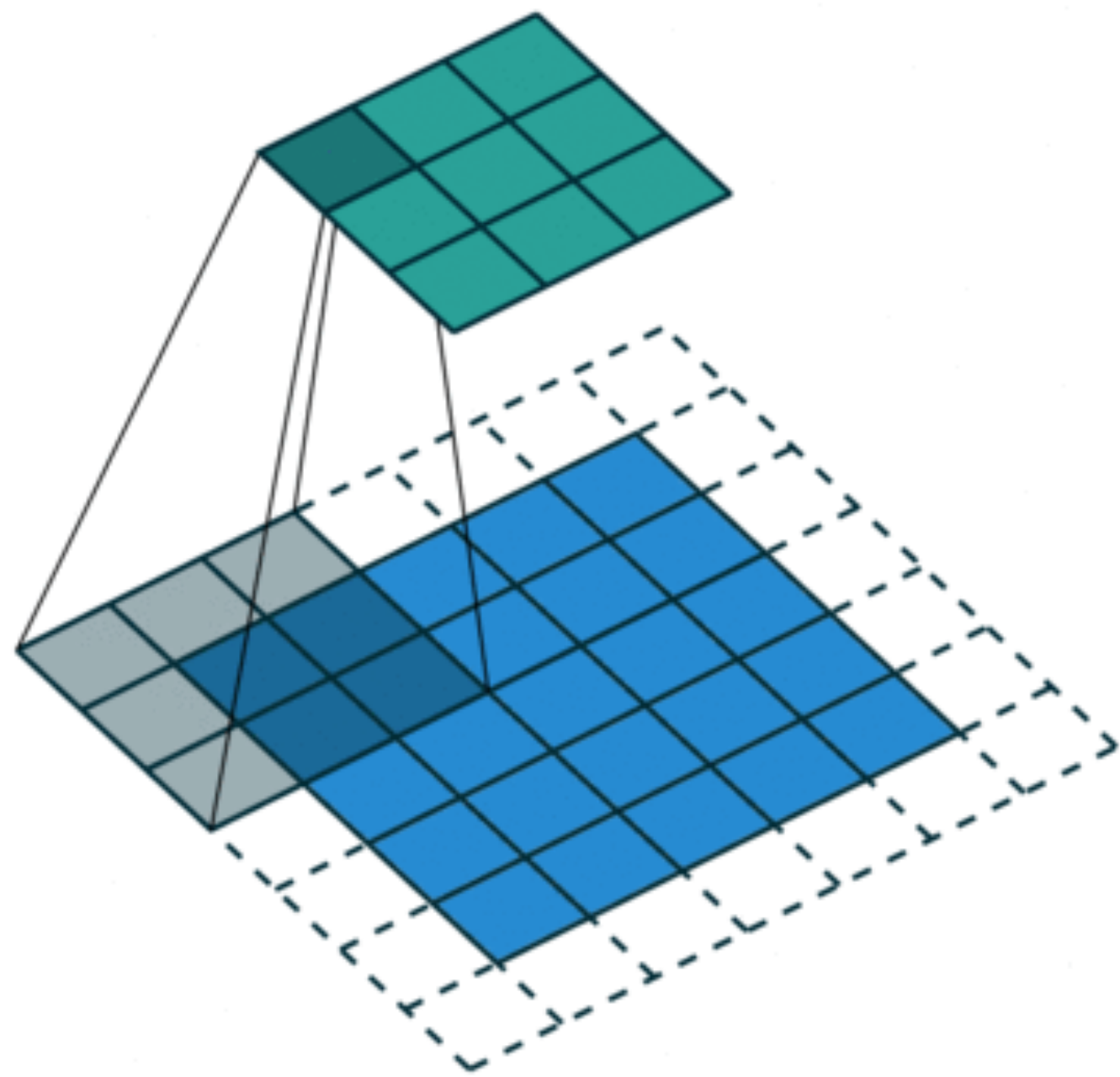


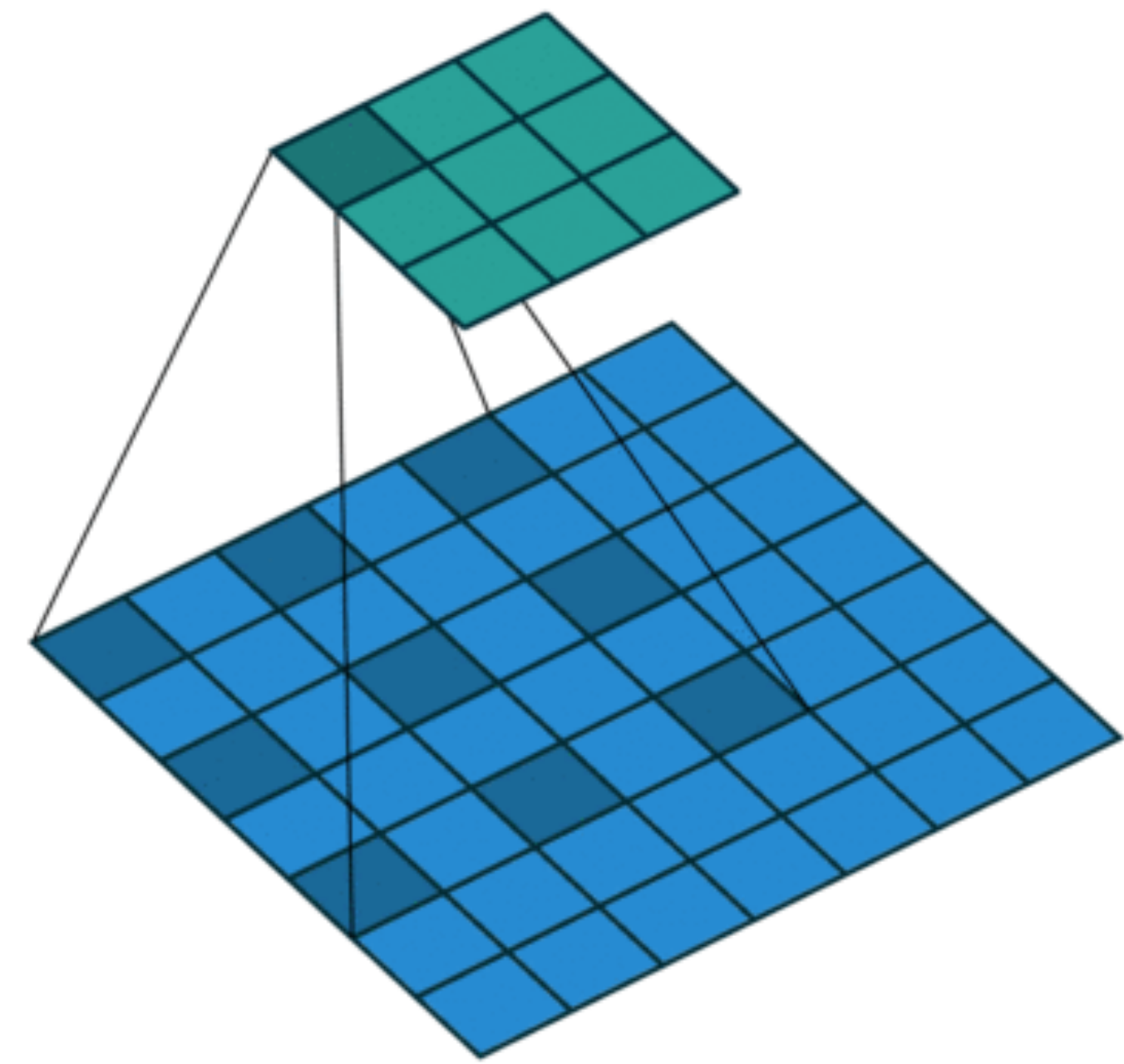
Comments

of Convolutional Hierarchical Attention Network for Query-Focused Video Summarization

- Encode video shots in parallel



Standard Convolution with a 3 x 3 kernel (and padding)



Dilated Convolution with a 3 x 3 kernel and dilation rate 2