S7

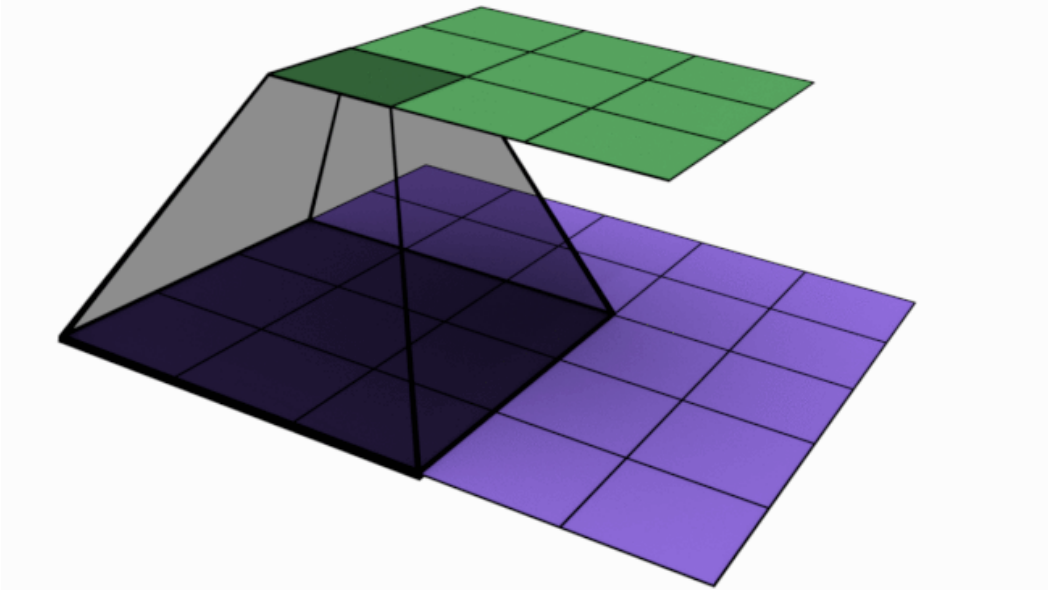
Due No Due Date **Points** None **Available** after Feb 26 at 6:30am

Advanced Convolutions

What is Convolution?

The process of extracting features from input data using kernels/filters. Filter moves discretely on top of the data-plane/channel, scales the input data equal to the size of its receptive field, sums these results and creates a new feature map.

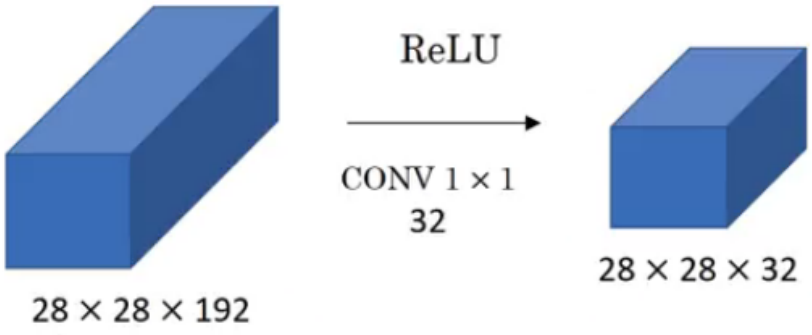
Normal Convolutions



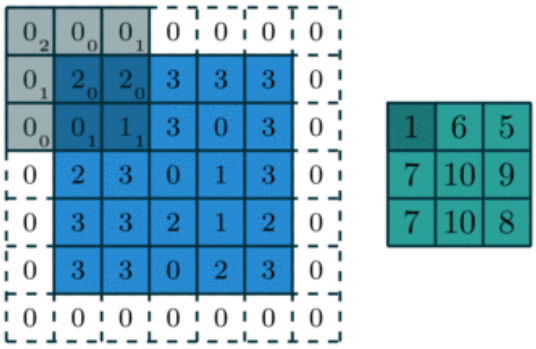
Pointwise Convolution

Already covered, so we will skit it.

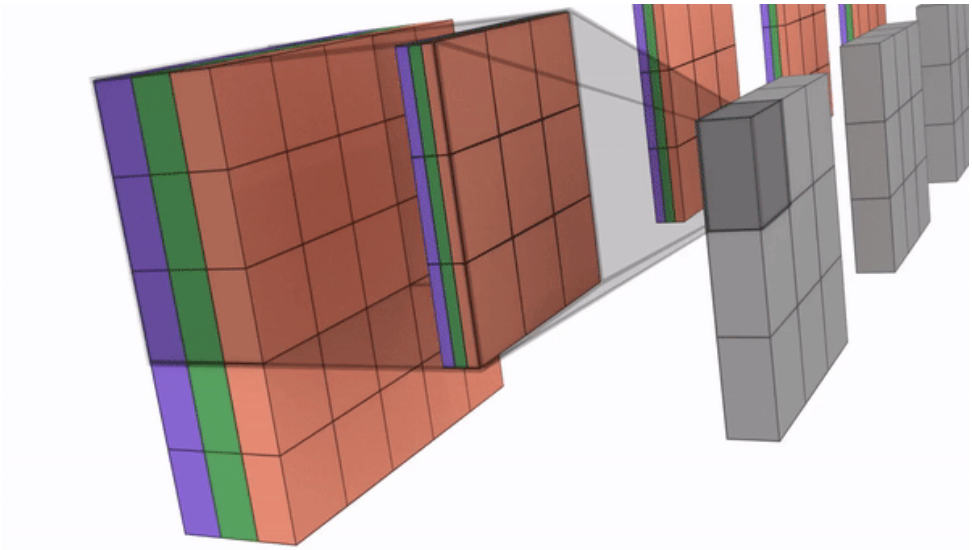
Basically



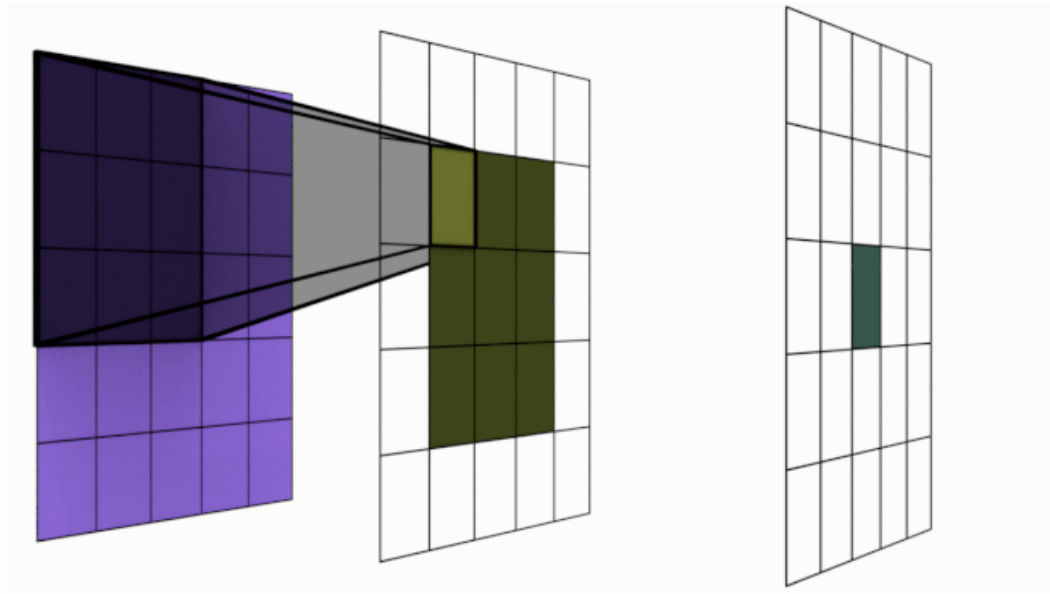
Behind the scenes:

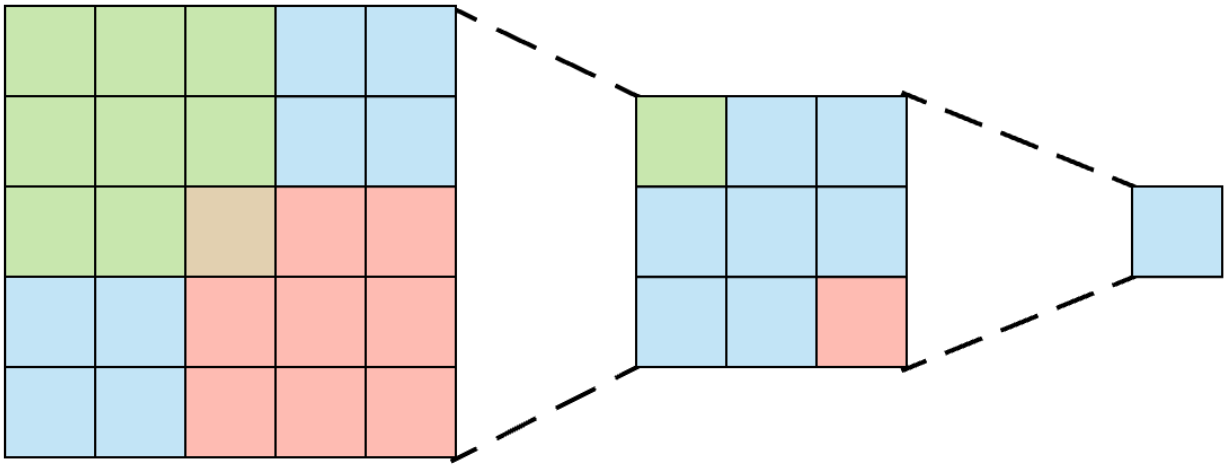


Concept of Channels



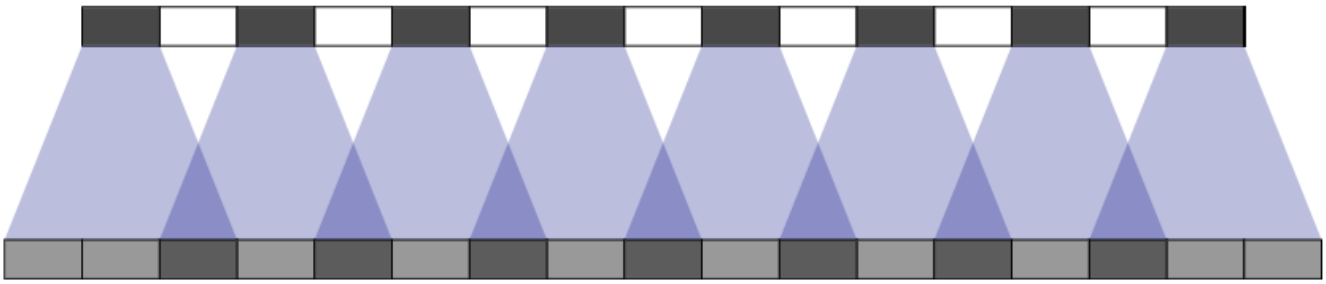
Receptive Field

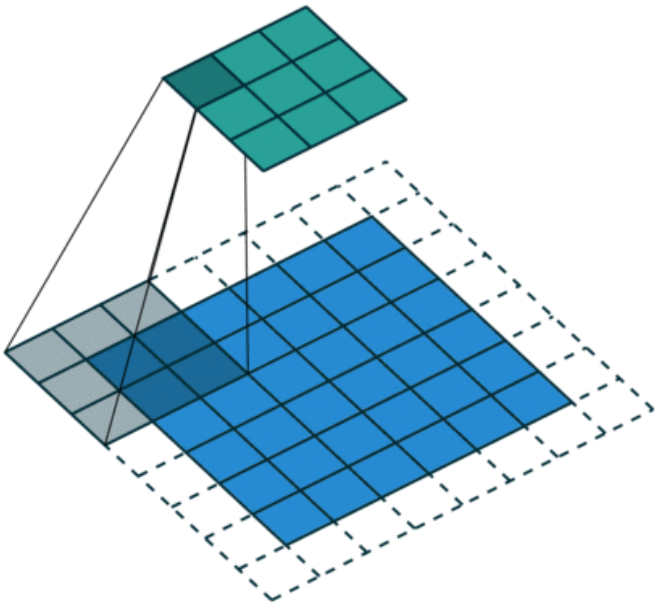




5x5 receptive field 3x3 receptive field

Checkboard Issue





We get a receptive field of 3x3 when we convolve by 3x3.

What should we do to get a receptive field of 5x5 when we convolve by 3x3?

Atrous Convolutions

or

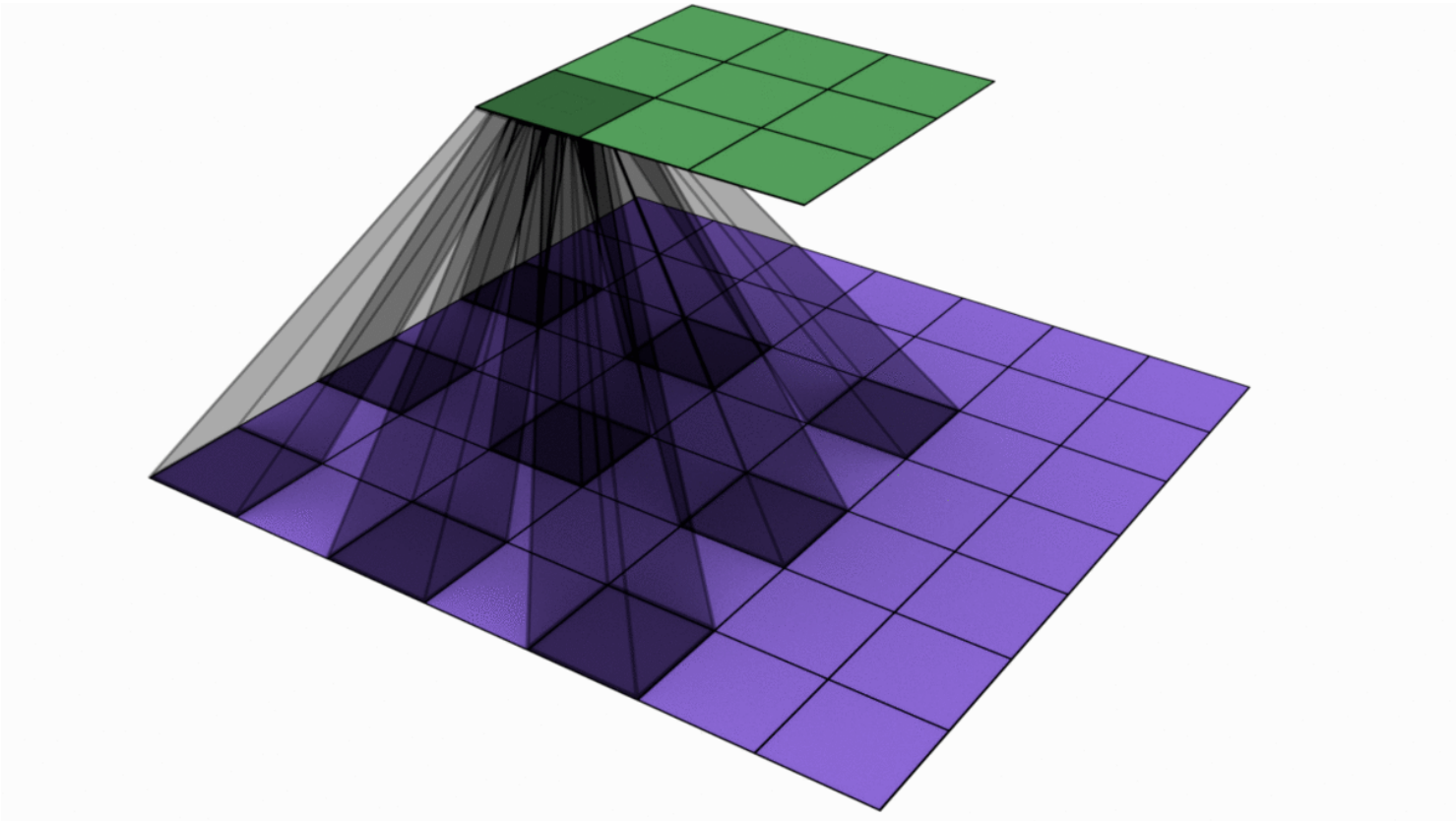
Dilated Convolutions

Dilated convolution is a way of increasing the receptive view (global view) of the network exponentially and linear parameter accretion. With this purpose, it finds usage in applications cares more about integrating the knowledge of the wider context with less cost.

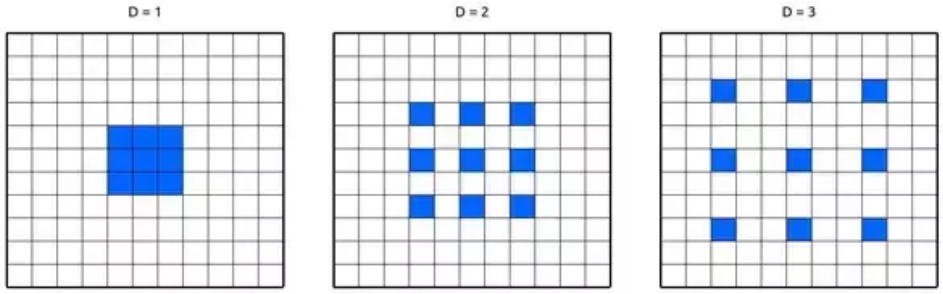
The key application the dilated convolution authors have in mind is dense prediction: vision applications where the predicted object that has similar size and structure to the input image. For example, semantic segmentation with one label per pixel; image super-resolution, denoising, demosaicing, bottom-up saliency, keypoint detection, etc.

> In many such applications one wants to integrate information from different spatial scales and balance two properties:

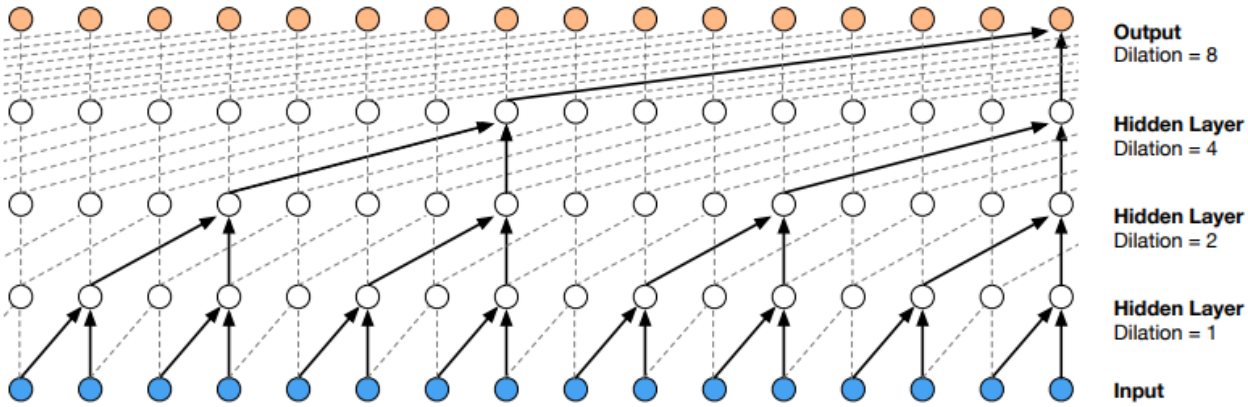
1. local, pixel-level accuracy, such as precise detection of edges, and
2. integrating the knowledge of the wider, global context



Dilation:



Dilated Convolution in WaveNet



Source **(http://sergeiturukin.com/2017/03/02/wavenet.html)**

The Problem: Image Segmentation

What is the network output? "Dense"

The Problem: "convolution is a rather local operation"

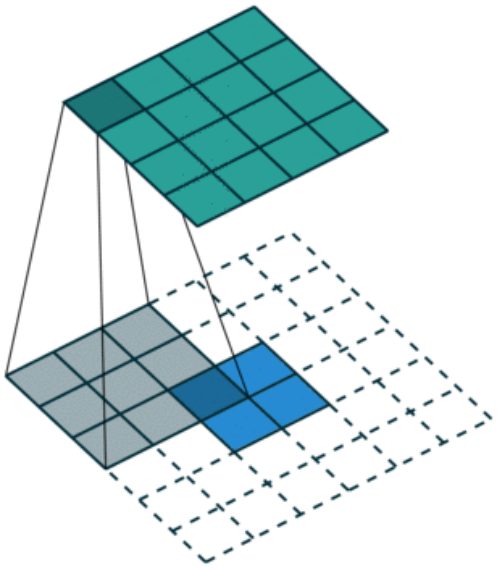
Improving the resolution of segmentation results

Dilated convolutions or atrous convolutions present a potential solution to this problem: they are capable of capturing global context without reducing the resolution of the segmentation map. Normal convolutional neural networks (CNNs) are based on the fact that each convolutional layer collects information from a larger neighborhood around each pixel/voxel. This means that in the upper layers of the network, each pixel of a feature map could potentially hold information about a large region of the image. However, experiments show that during learning this is usually not the case. Rather, the **information in each pixel stays localized** [source (https://arxiv.org/abs/1412.6856)] and the network “does not live up to its full potential”. Dilated convolutions change the rules of the game by using kernels of the same size as normal convolutional layers but spread out over a larger area on the image (see animation).

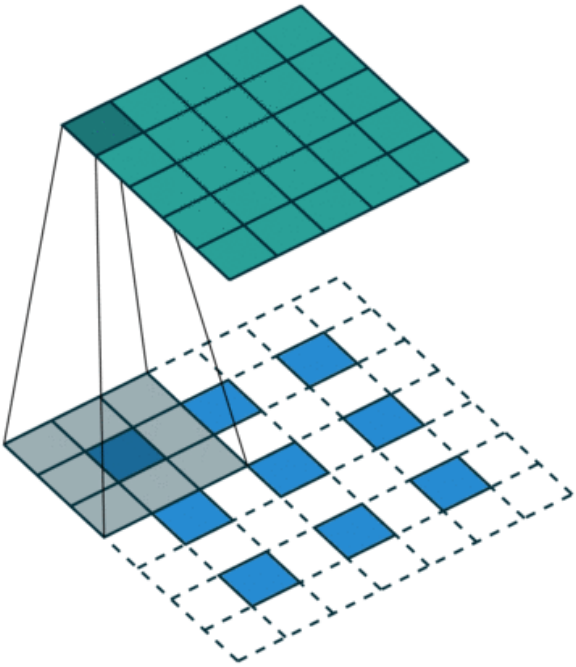
DECONVOLUTION or Fractionally Strided OR Transpose Convolution

We have a kernel of size 3x3. If we convolve on 5x5 we get 3x3. What do we do to get 7x7? Or get a larger channel size than we started with?

A simple approach, but inefficient.



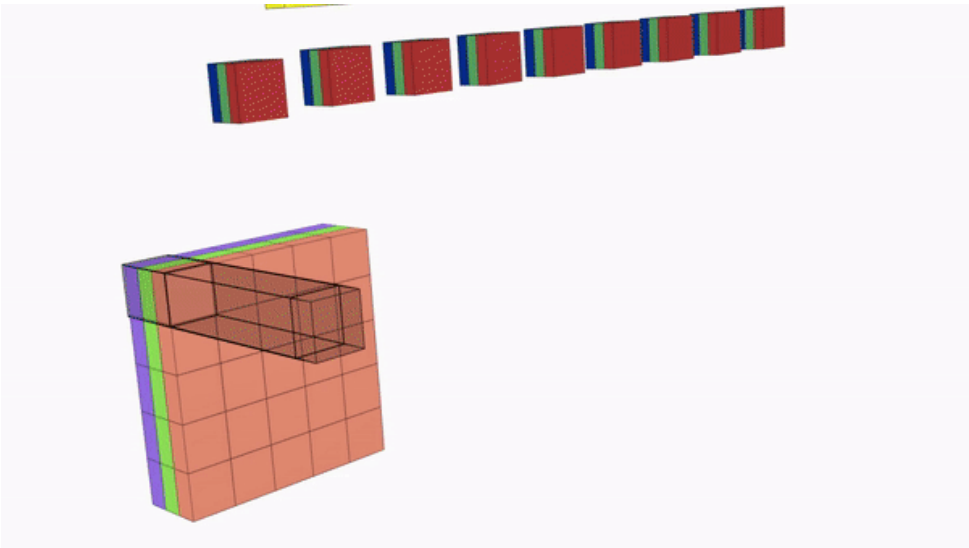
A Better approach, but..



it introduces the checkerboard issue!

Perceptual Losses for Real-Time Style Transfer and Super-Resolution

New Convolution strategy! - Pixel Shuffle

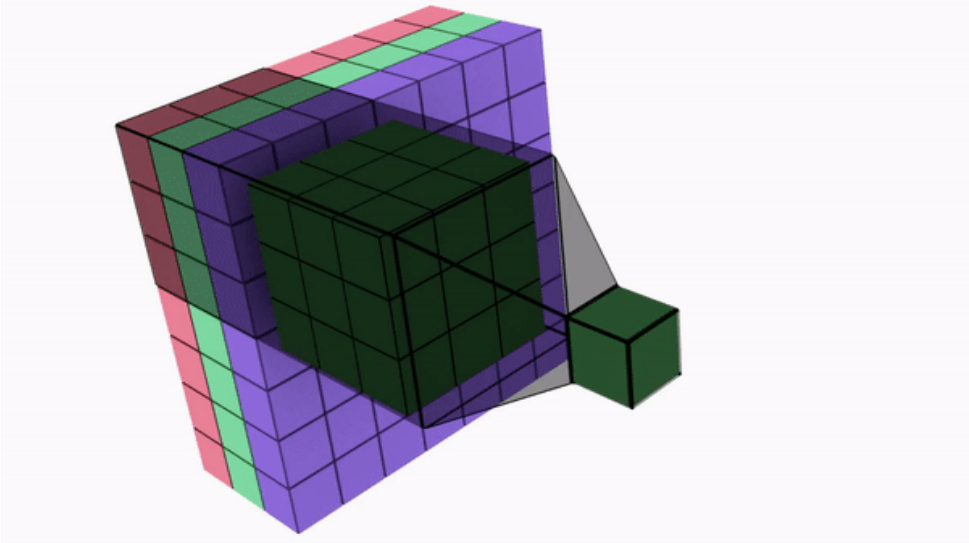


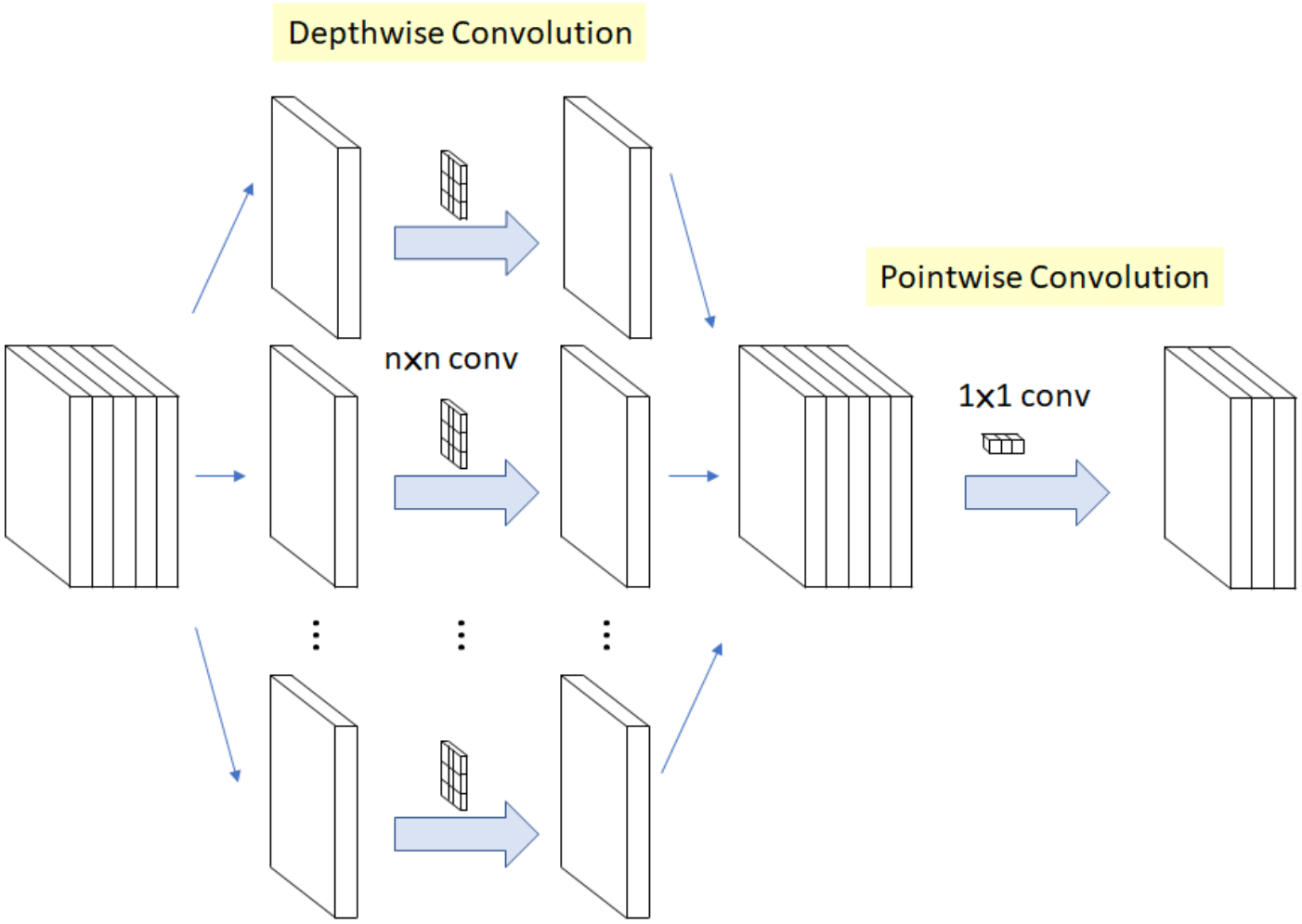
[Jeremy Howard's Explanation](https://www.youtube.com/embed/nG3tT31nPmQ?start=2128)

[Pixel Shuffle Paper](https://arxiv.org/pdf/1609.05158.pdf)

Depthwise Separable Convolution

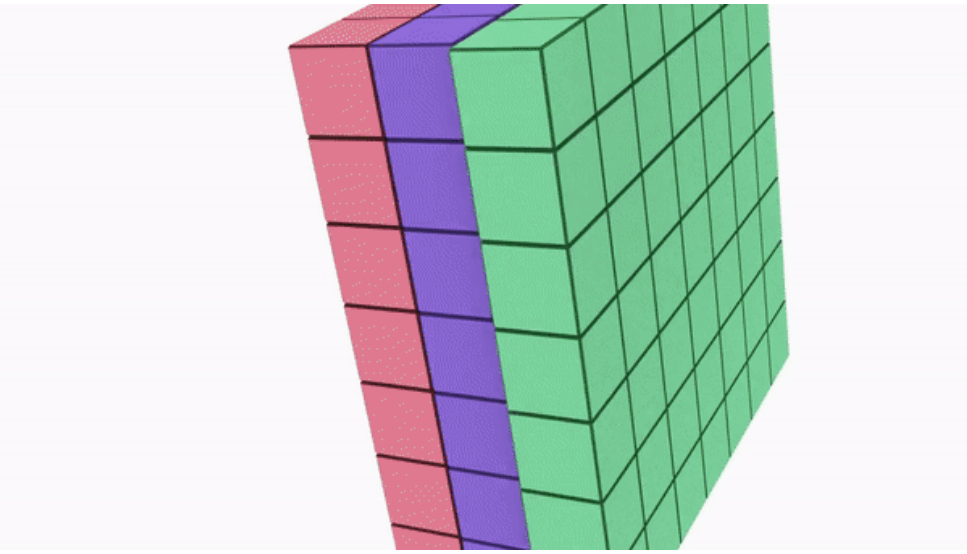
What is wrong in this convolution?





Used by XCEPTION, MOBILENET and newer Mobile Architectures a lot

Depthwise Convolution

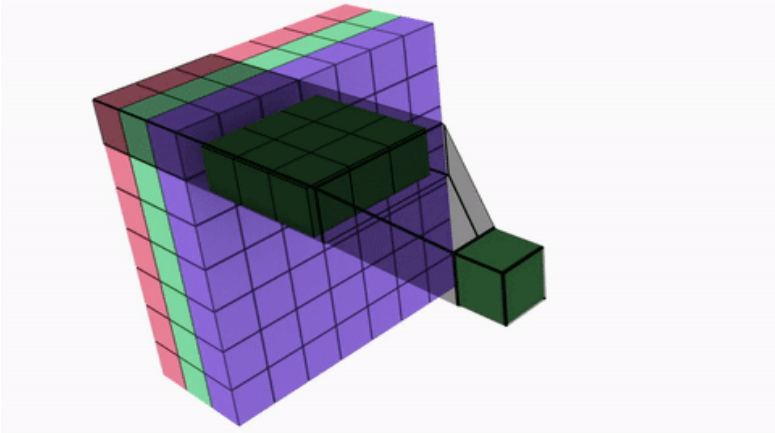


Reference **(https://eli.thegreenplace.net/2018/depthwise-separable-convolutions-for-machine-learning/)**

For a depthwise separable convolution on the same example, we traverse the 16 channels with 1 3x3 kernel each, giving us 16 feature maps. Now, before merging anything, we traverse these 16 feature maps with 32 1x1 convolutions each and only then start to them add together. This results in 656 (16x3x3 + 16x32x1x1) parameters opposed to the 4608 (16x32x3x3) parameters from above.

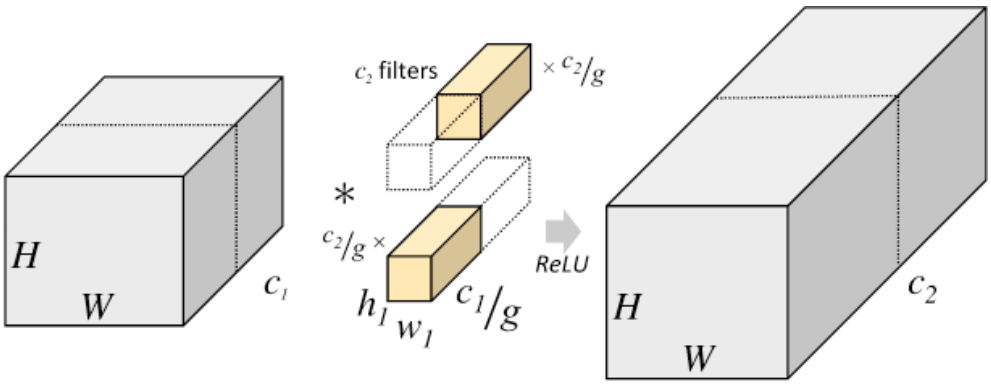
Spatially Separable Convolution

Was used immensely in different variants of Xception-Incepeption Networks as well as the mobile net.



Grouped Convolution

Grouped convolutions were initially mentioned in AlexNet and later reused in ResNext. The main motivation of such convolutions is to reduce computational complexity while dividing features into groups.



Go Crazy!!

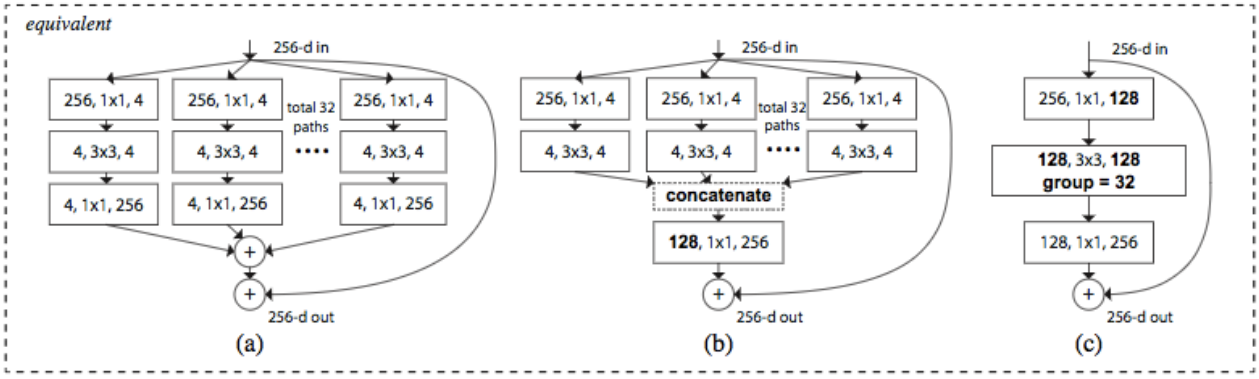


Figure 3. Equivalent building blocks of ResNeXt. **(a)**: Aggregated residual transformations, the same as Fig. 1 right. **(b)**: A block equivalent to (a), implemented as early concatenation. **(c)**: A block equivalent to (a,b), implemented as grouped convolutions [24]. Notations in **bold** text highlight the reformulation changes. A layer is denoted as (# input channels, filter size, # output channels).

Assignment 7

1. Run this **network** (https://colab.research.google.com/drive/1qlewMtxcAJT6fIJdmMh8pSf2e-dh51Rw) .
2. Fix the network above:
    1. change the code such that it uses GPU
    2. change the architecture to C1C2C3C40 (basically 3 MPs)
    3. total RF must be more than 44
    4. one of the layers must use Depthwise Separable Convolution
    5. one of the layers must use Dilated Convolution
    6. use GAP (compulsory):- add FC after GAP to target #of classes (optional)
    7. achieve 80% accuracy, as many epochs as you want. Total Params to be less than 1M.
    8. upload to Github
    9. Attempt S7-Assignment Solution

Session Video:


EVA 4 - Session 7 Wednesday