Inception Module in Deep Convolution Networks

CS8004: Deep Learning and Applications

Motivation…

- Standard CNN architecture until 2014 consisted of stacked convolutional layers with Max pooling and batch normalization.

- But standard CNN architecture has various limitations.
  - Large memory footprint
  - Large computation demand
  - Prone to overfitting
  - Vanishing and exploding gradients
  - AlexNet had ~60 million parameters

Motivation…

- Dense connections are expensive whereas Biological systems are sparse.
- Sparsity can be exploited by clustering correlated outputs (Theoretical work by Arora et al.)
- How to choose the filter size to capture features at different abstraction levels simultaneously.
- Wider filters have powerful ability of local abstraction.
- Levels of features can be enriched by the depth of the network (the number of stacked layers).
- One way is to integrate low/mid/high-level features and classifiers.

Main Issues

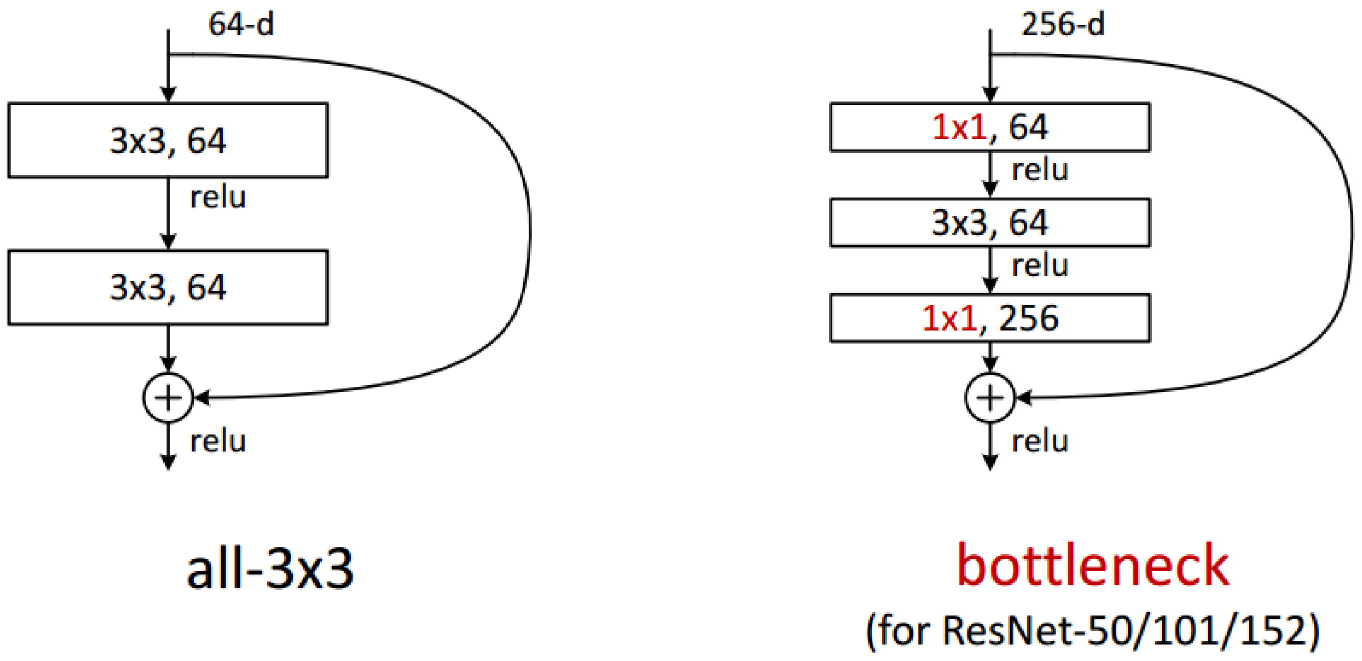
- Increasing model size and computational cost leads to immediate quality gains for most of the tasks.
- But computational efficiency decreases due to explosion in trainable parameters.
- What are the major factors in computational load : high dimensional convolutions.
- Higher dimensional representations are easier to process locally within a network.
- Increasing the activations in each layer in a convolutional network allows for more disentangled features. The resulting networks will train faster.
- Let us see how this can be achieved.

General Practice

- Avoid representational bottlenecks and maintain higher dimensional representations, particularly in the initial network layers.
- The representation size should gently decrease from the input layer to the output.
- Balance the width and depth of the network.

Example

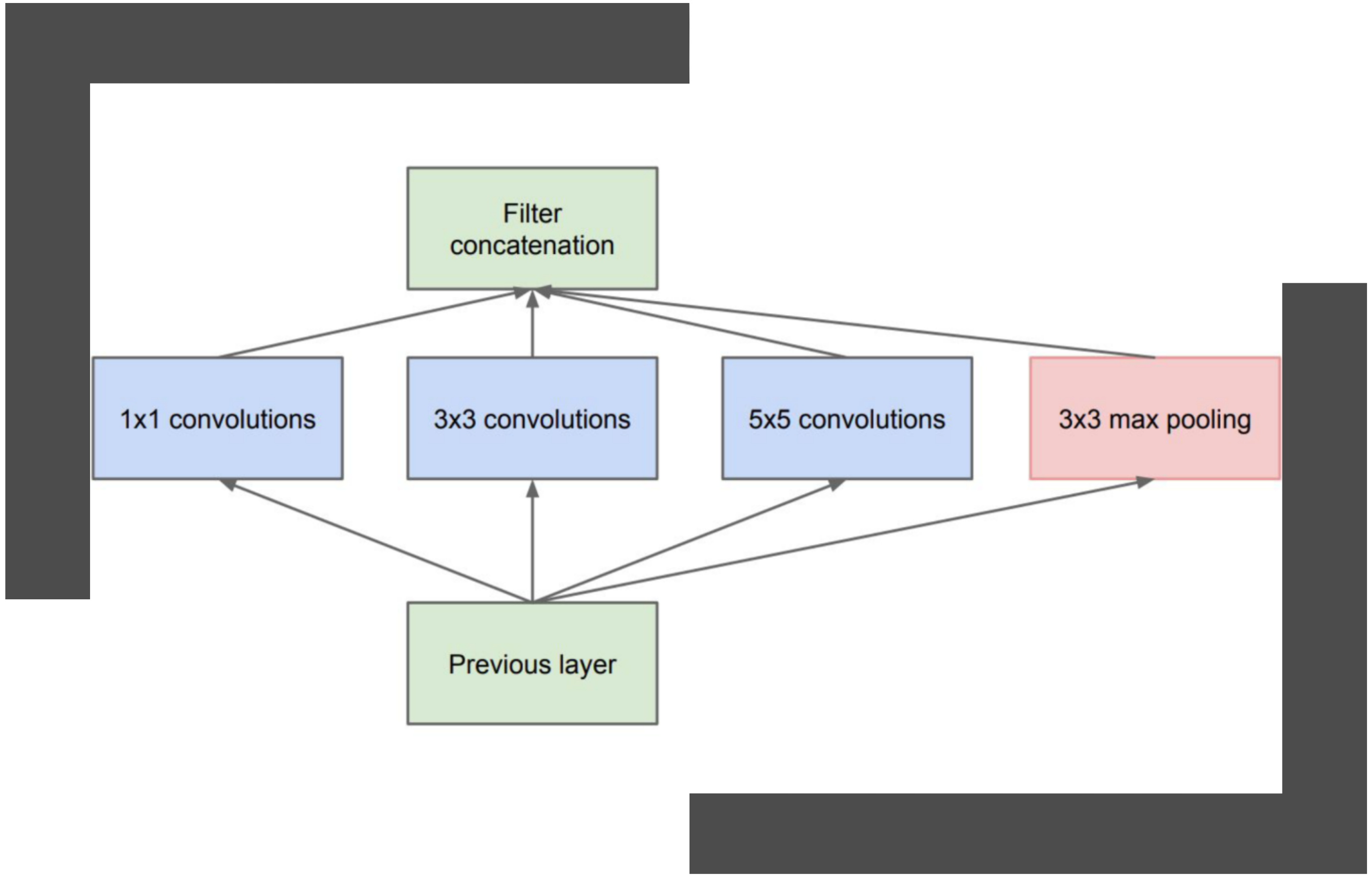
- Both the ResNet modules in a network have similar complexity.



Inception Module

- Inception: Balance model size and computation cost.
- Use different size of filters altogether to capture different levels of feature abstractions simultaneously.
- Go deeper to enrich the feature abstraction.

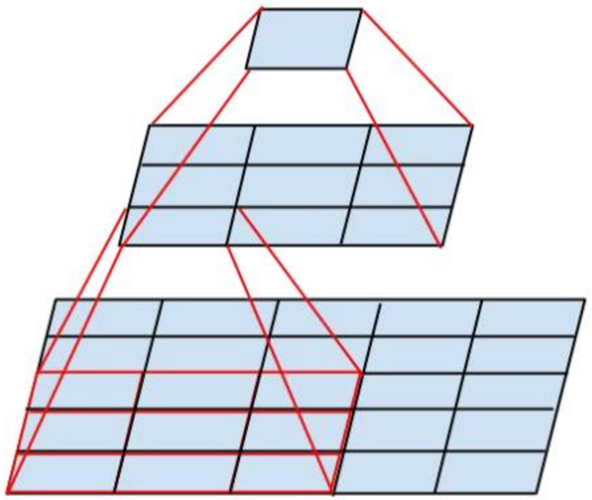
Inception Module (Naïve)



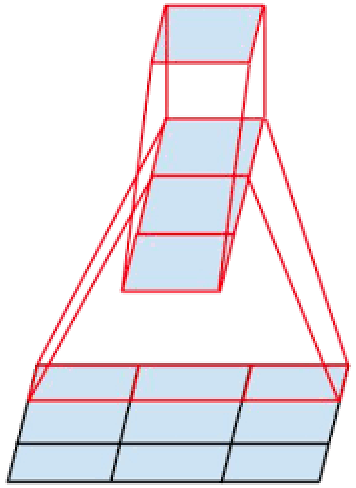
- Approximation of an optimal local sparse structure
- Process visual/spatial information at various scales and then aggregate
- But higher dimensional feature maps are computationally expensive.

Inception Module

- Let us see a computationally efficient \parameter reduction model.



Mini-network replacing the
5×5 convolutions



Mini-network replacing the
3×3 convolutions

fully-connected

convolution

Same
receptive field

3*3

3*3

$$\frac{9+9}{25}$$

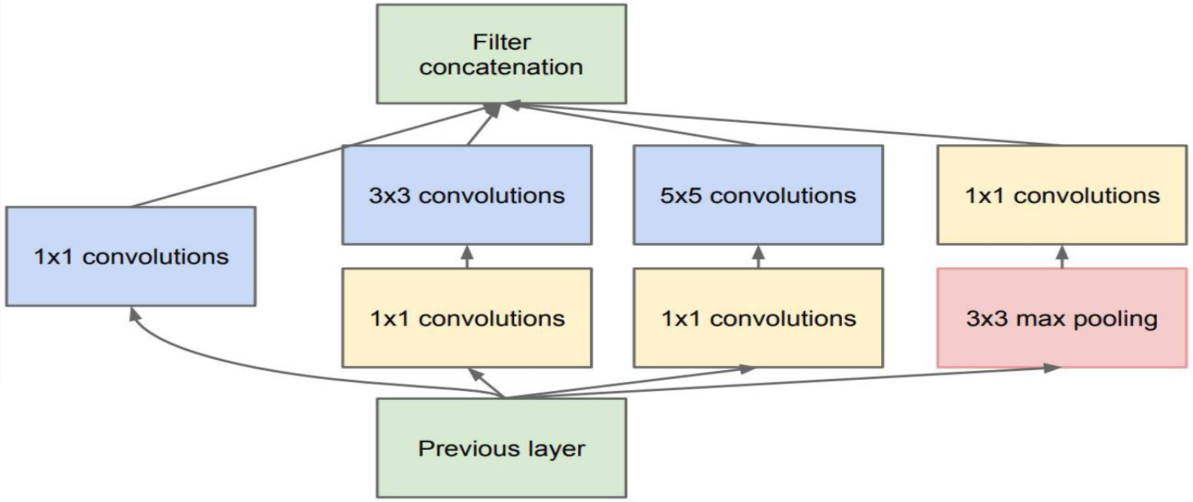
3*1

1*3

$$\frac{3+3}{9}$$

- Less parameters
- Same inputs and outputs

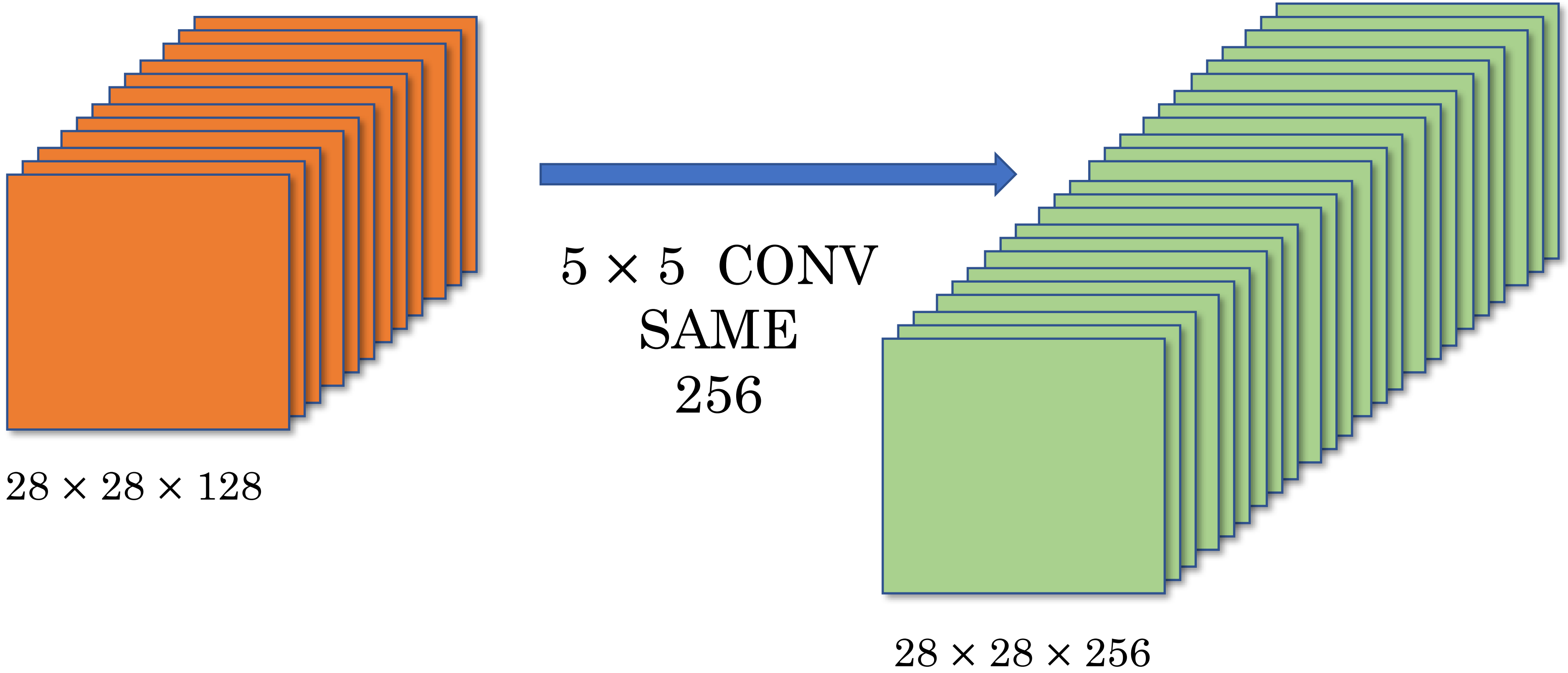
Inception Module with Dimension Reduction



- Computationally efficient.
- Dimension reduction
- Achieved with 1x1 convolutions.
- Pooling in depth instead of max/average pooling in height/width.

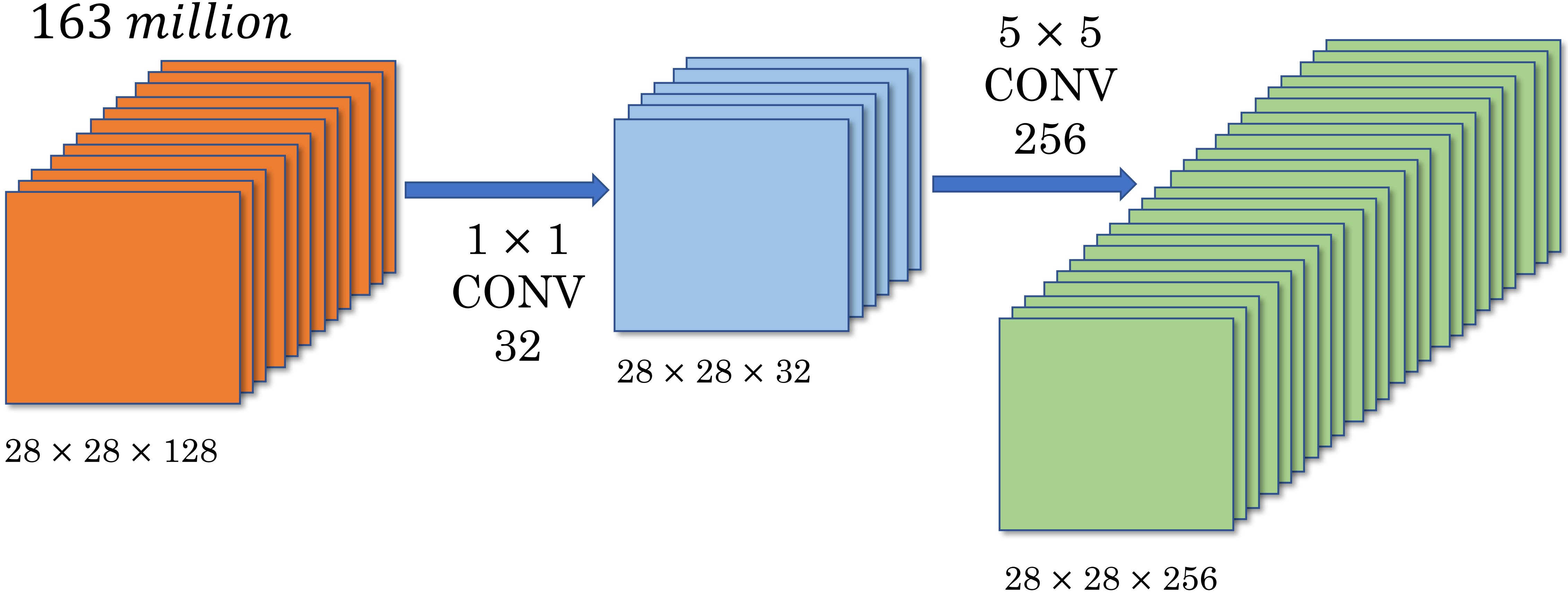
Computational Efficiency

- Number of multiplications: $(28 \times 28 \times 256\ ) \times (\ 5 \times 5 \times 128) = 642{,}252{,}800 \approx 642\ million$

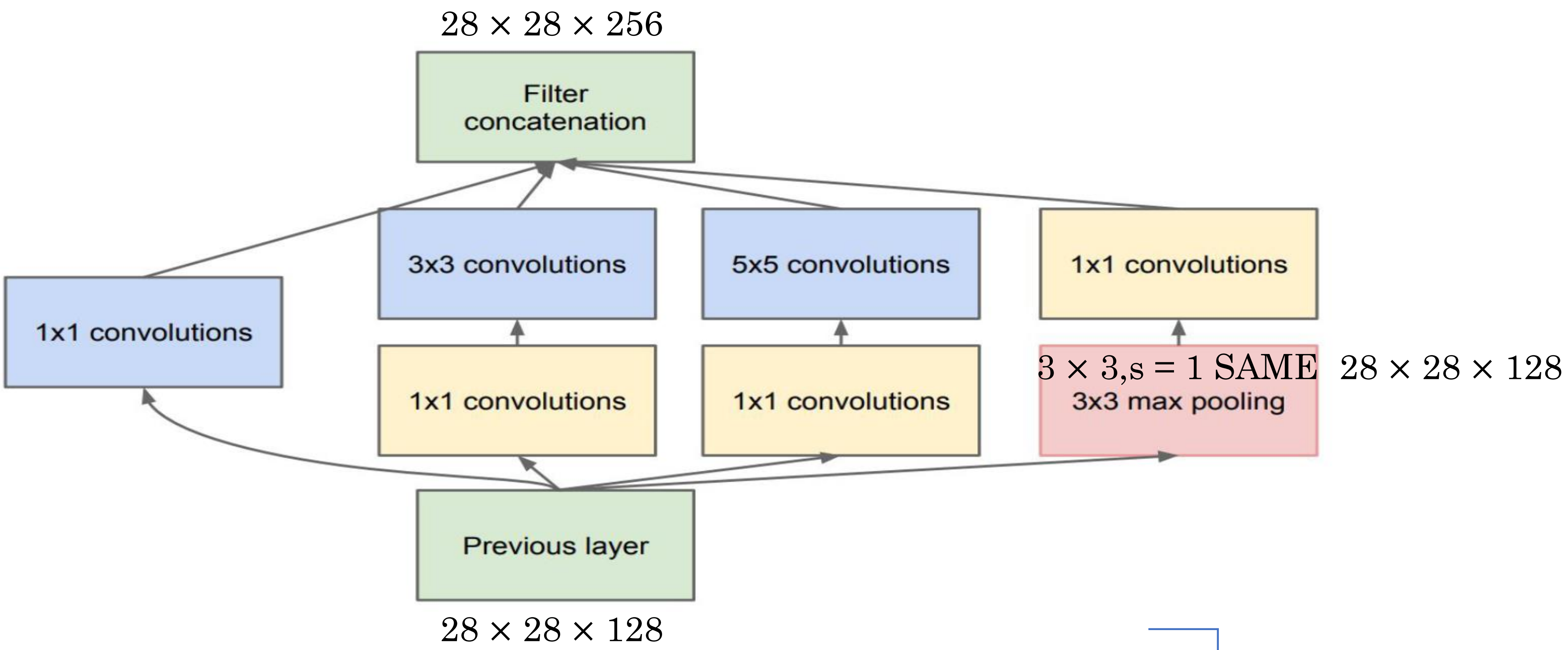


Computational Efficiency

- Number of multiplications: $(28 \times 28 \times 32\ ) \times 128 + (\ 28 \times 28 \times 256) \times (5 \times 5 \times 32) = 3{,}211{,}264 + 160{,}563{,}200 \approx 163\ million$

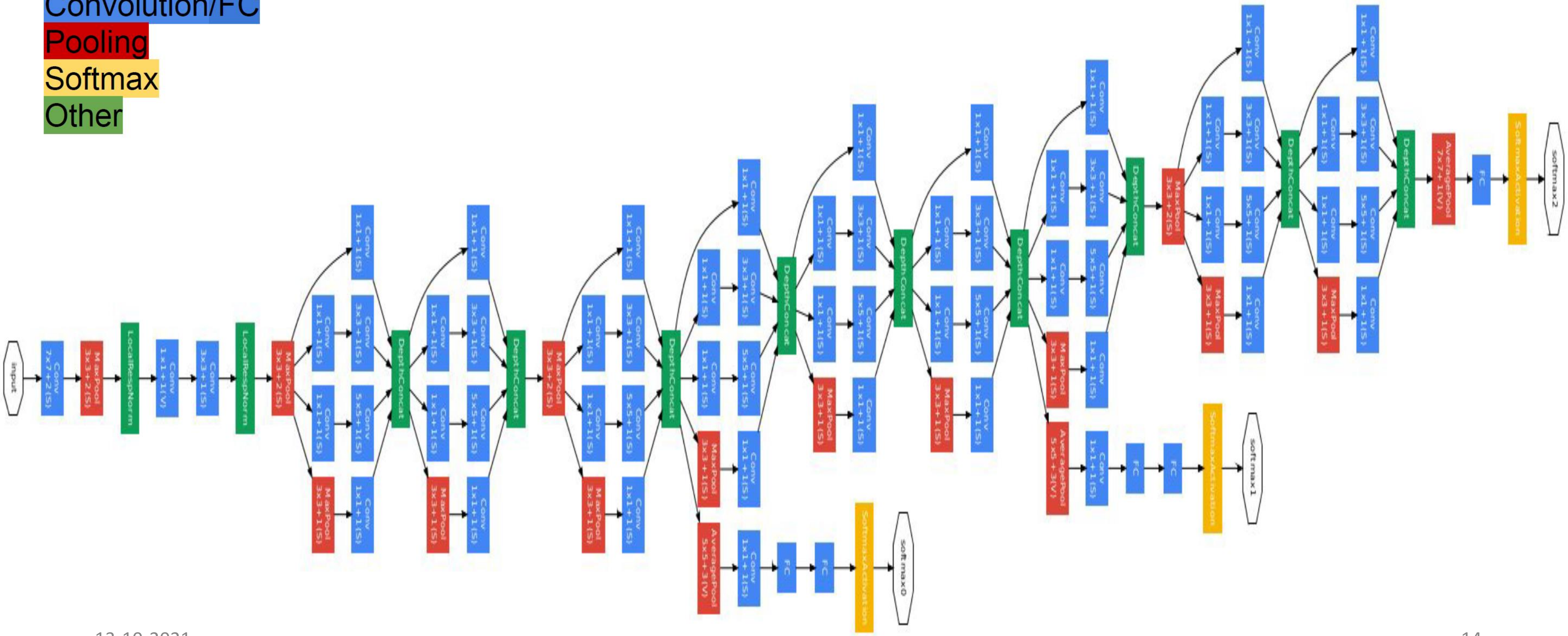


The Inception Module



Full Inception-v1

Convolution/FC
Pooling
Softmax
Other

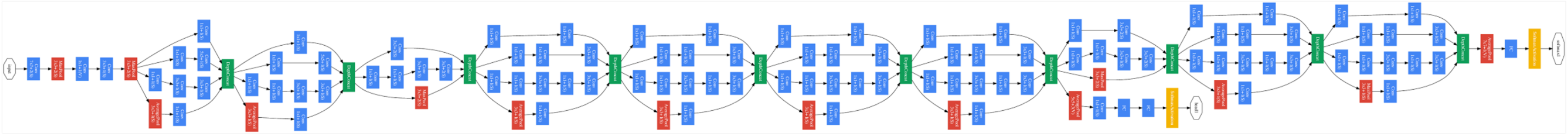


Inception Network v1: GoogLeNet

- Inception Modules: standard convolution, pooling, and normalization layers.
- Total Structure: 9 stacked inception modules.
- Final Classifier: average pooling and single fully connected (fc)-layer.
- Small Improvement over all fc-layers.
- Auxiliary classifiers
  - Helpful in discrimination
  - provide regularization
  - discarded at inference
- 22 layers total, only ~5 million parameters 12 times less than AlexNet.
- ILSVRC'14 classification winner (6.7% top 5 error)

Evolution of Inception

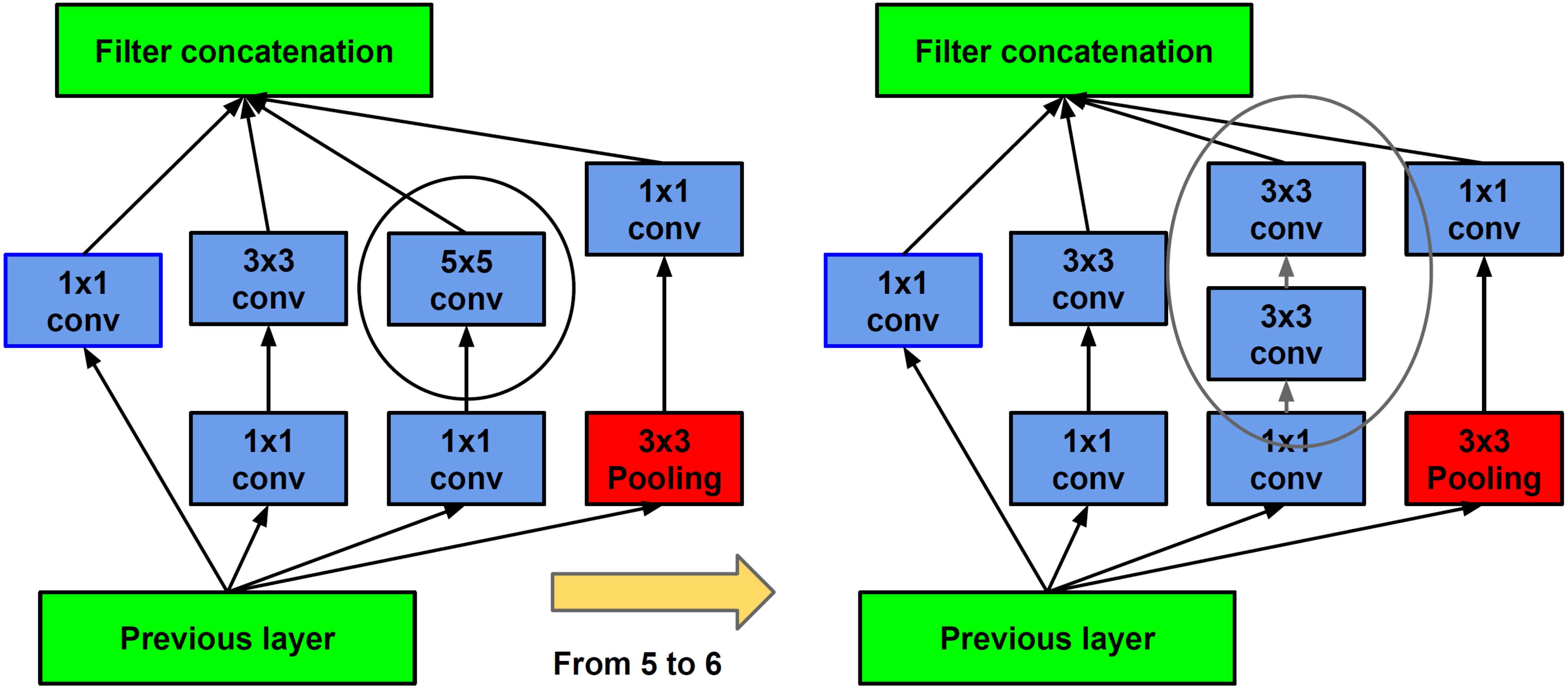
[1]Inception 5 (GoogLeNet)



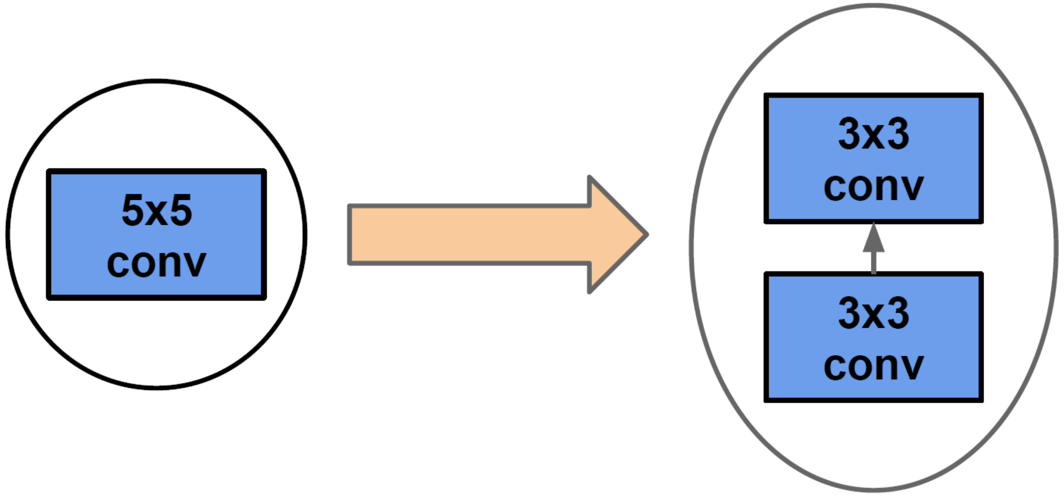
Inception 7a

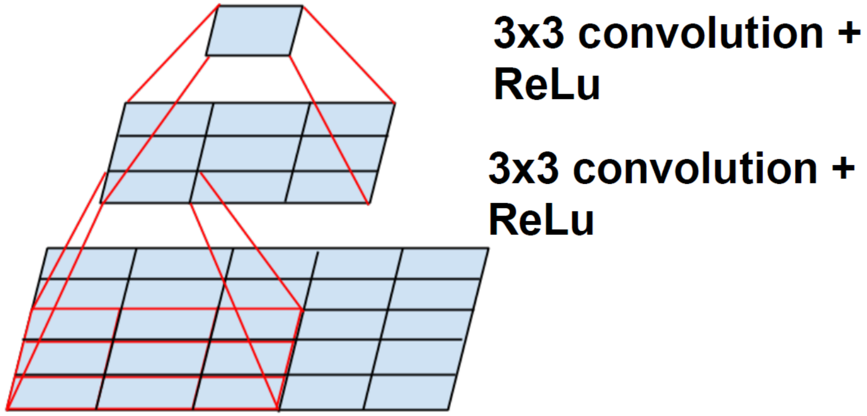
[1]Going Deeper with Convolutions, [C. Szegedy et al, CVPR 2015]

Structural Changes between Inception 5 and 6



Structural Changes between Inception 5 and 6



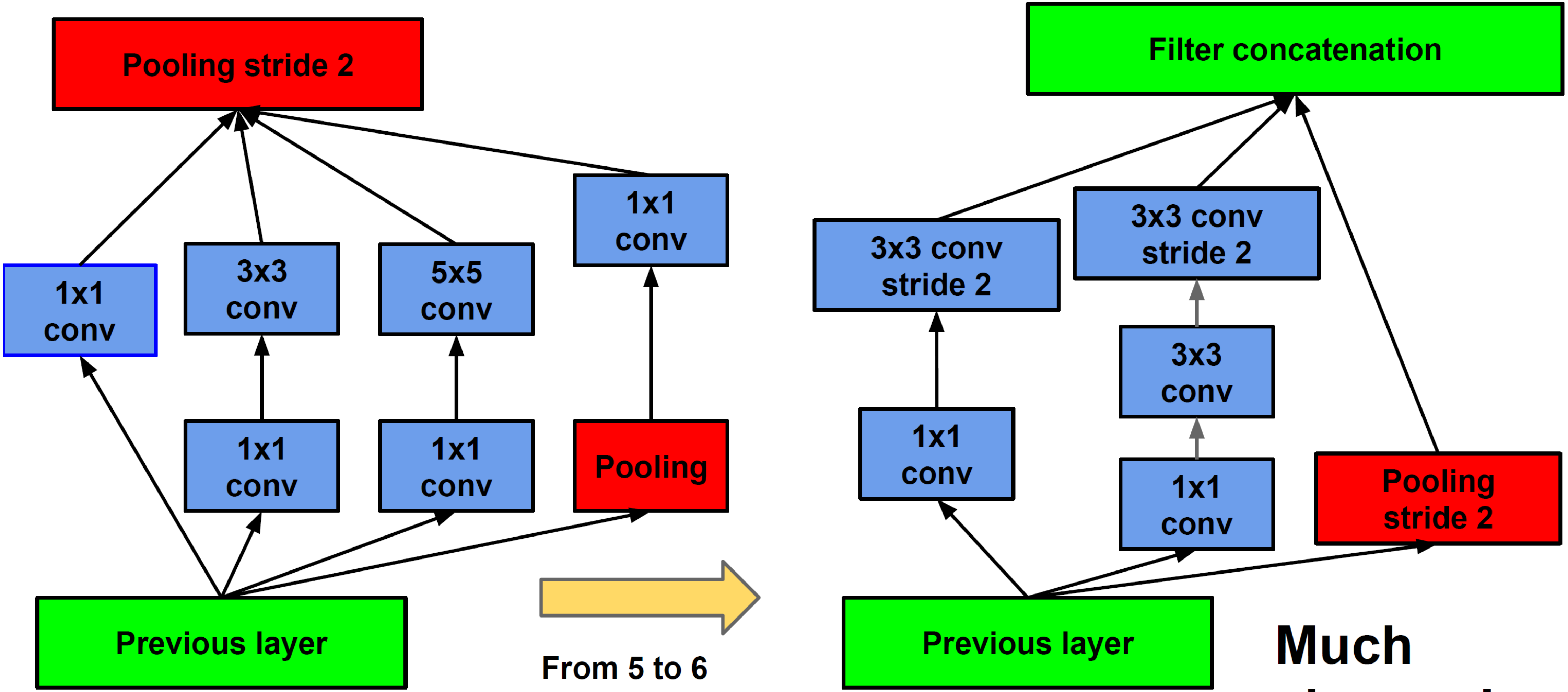


- **Each mini network has the same receptive field.**
- **Deeper: more expressive (ReLu on both layers).**
- **25 / 18 times (~28%) cheaper (due to feature sharing).**
- **Computation savings can be used to increase the number of filters.**

<u>Downside:</u>

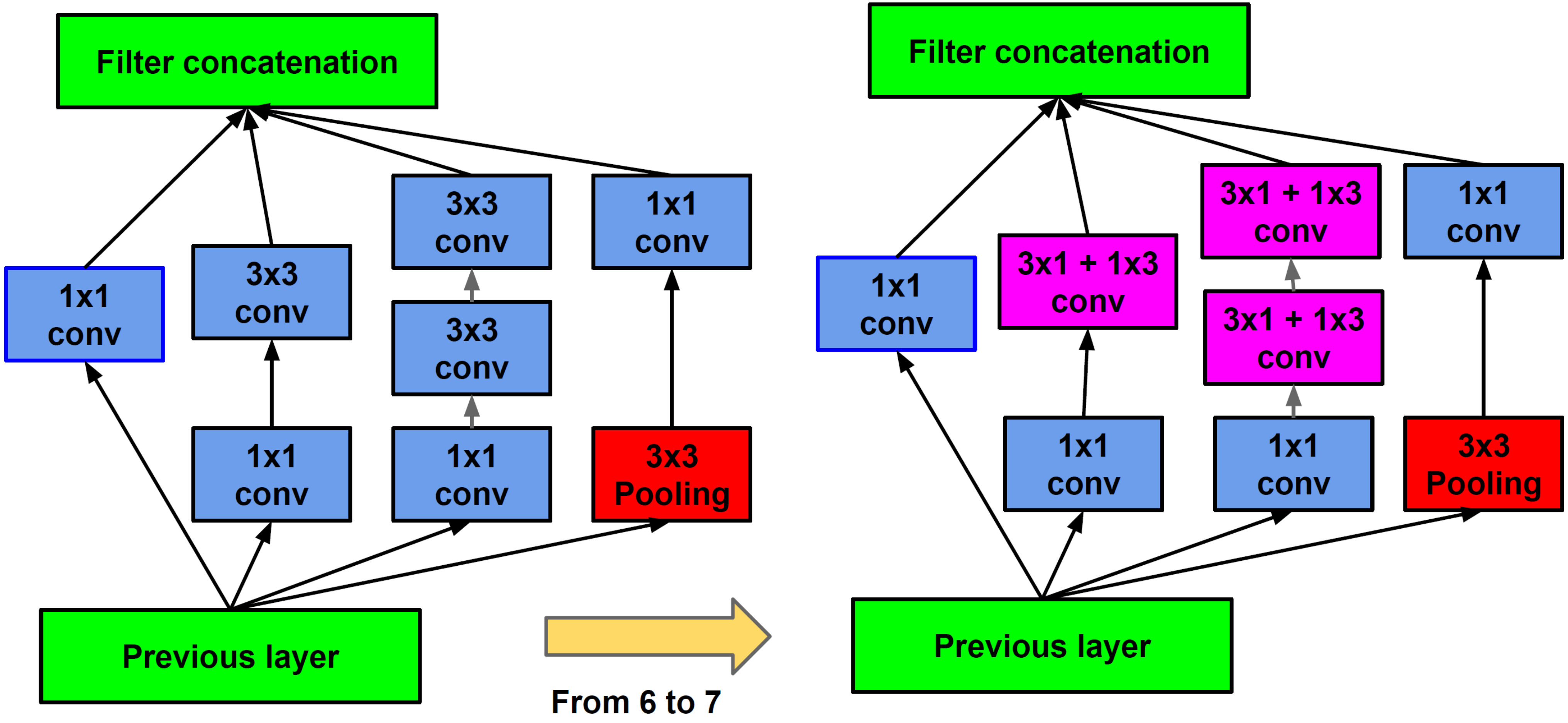
Needs more memory at training time

Grid Size Reduction from Inception 5 to 6

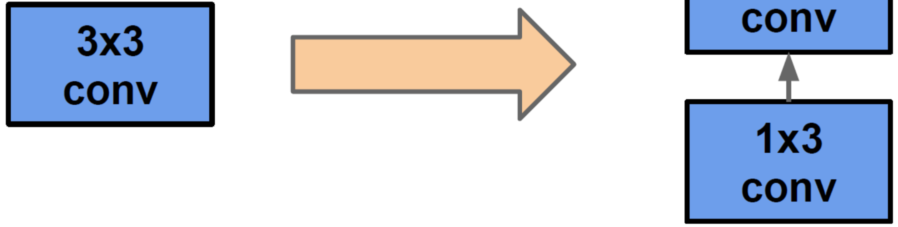


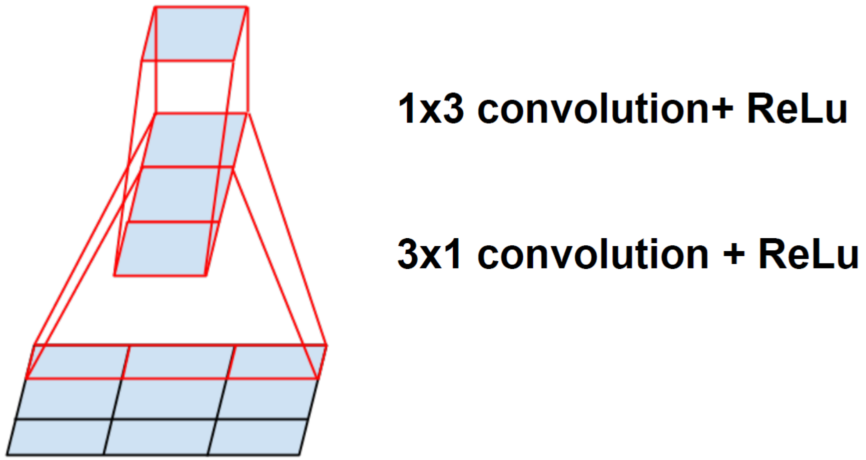
Much cheaper!

Structural Changes between Inception 6 and 7



Structural Changes between Inception 6 and 7





- **Each mini network has the same receptive field.**
- **Deeper: more expressive (ReLu on both layers).**
- **9 / 6 times (~33%) cheaper (due to feature sharing).**
- **Computation savings can be used to increase the number of filters.**

<u>Downside:</u>

Needs more memory at training time

Padding in Inception Modules

Inception 6: **SAME** padding throughout:

SAME padding

Input grid size	Patch size	Stride	Output grid size
8x8	3x3	1	8x8
8x8	5x5	1	8x8
8x8	3x3	2	4x4
8x8	3x3	4	2x2

- Output size is independent of patch size
- Padding with zero values

VALID padding

Input grid size	Patch size	Stride	Output grid size
7x7	3x3	1	5x5
7x7	5x5	1	3x3
7x7	3x3	2	3x3
7x7	3x3	4	2x2

- Output size depends on the patch size
- No padding: each patch is fully contained

Padding in Inception Modules v6 and v7

Advantages of padding methods

SAME padding

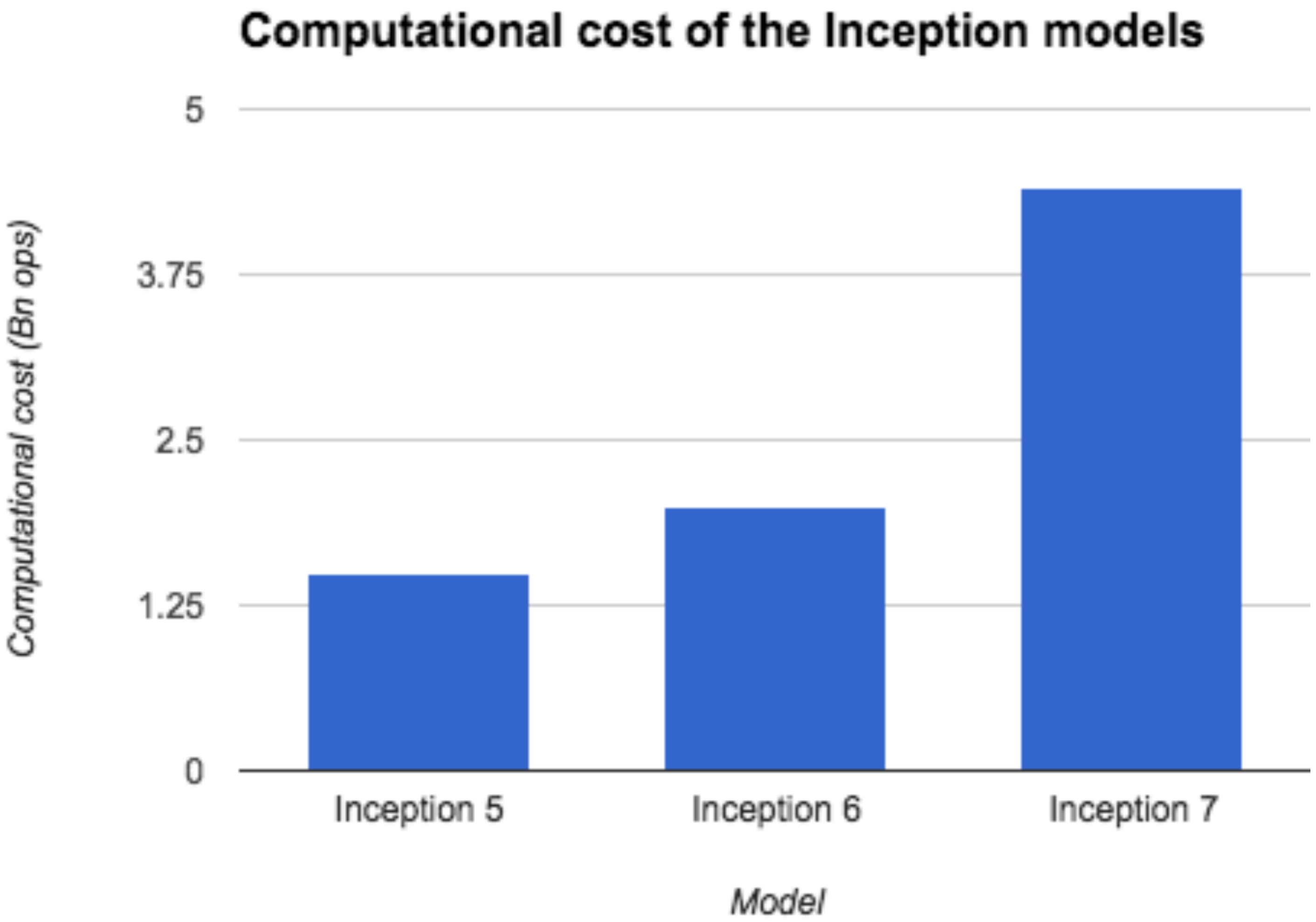
- More equal distribution of gradients
- Less boundary effects
- No tunnel vision (sensitivity drop at the border)

VALID padding

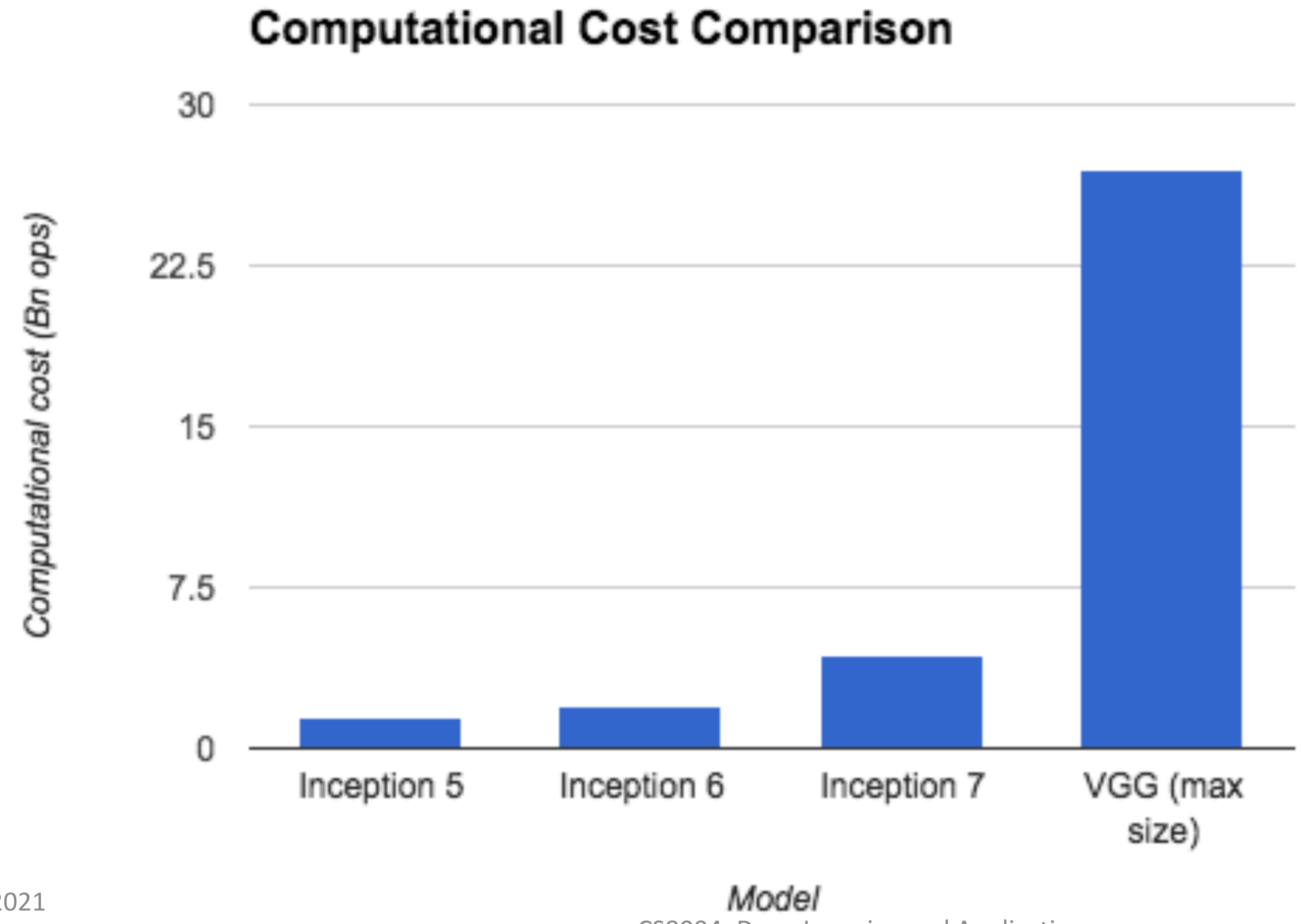
- More refined: higher grid sizes at the same computational cost

Stride	Inception 6 padding	Inception 7 padding
1	**SAME**	**SAME (VALID on first few layers)**
2	**SAME**	**VALID**

Computational Cost



Comparison with VGG



References

- Slides on inception Networks by
  - Wenchi Ma, Computer Vision Group, EECS,KU.
  - Christian Szegedy, Julian Ibarz and Vincent Vanhoucke’s presentation slides on Inception Networks, Google.
  - David White’s lecture slides on Inception Networks, Colorado State University