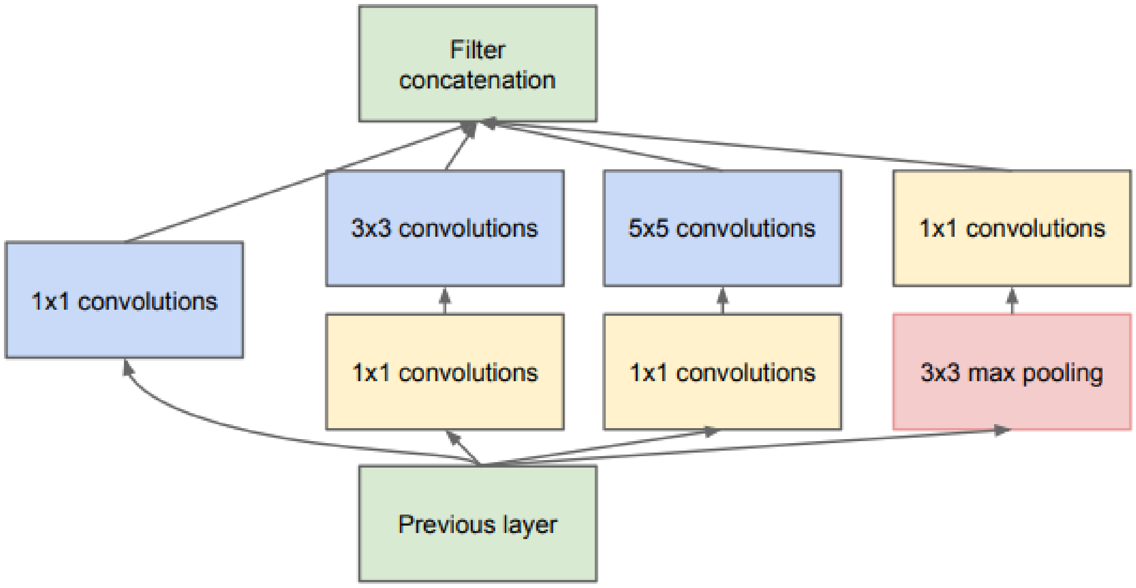
- #paper/to-read ~ Computer Vision
  - https://arxiv.org/abs/1409.4842
  - Sequel papers:
    - Inception-v2 and Inception-v3
    - Inception-v4 and Inception-ResNet
    - Xception
  - Mentioned papers:
    - Provable Bounds for Deep Representations
    - Network In Network
    - R-CNN
    - Scalable Object Detection
  - Mentioned topics:
    - Hebbian Theory
    - Jaccard Index
    - Polyak Averaging
    - Superpixel
    - Generalized Linear Model, GLM
  - [ ] https://towardsdatascience.com/a-simple-guide-to-the-versions-of-the-inception-network-7fc52b863202

- ## Summary
  - Since salient parts in an image can be very different in size, an **inception module** uses convolutions (and pooling filters) of different

sizes on the same level.



- 1×1 convolutions reduce the number of input channels to prevent the explosion of required compute in the higher inception modules.
  - Note that for Max Pooling, the 1×1 convolution goes after it. Otherwise the max pooling operation would be in vain.
- The main idea behind this pattern is based on finding how an optimal local sparse structure can be approximated with dense components.
  - Units from earliest layers correspond to some region of the input image and these units are grouped into Filter Banks.
    - Later on, features from these units are aggregated by 1×1 convolutions.
      - These convolutions also serve as Non-linearity providers.

- The second (and necessary) idea is to apply Dimensionality Reduction whenever computational Complexity would increase too much otherwise.
  - The paper mentions occasional Max Pooling layers with stride 2 that halve the resolution.
- **GoogLeNet** as the main example in the paper.
  - It employs **auxiliary Loss** after some of the inception modules in the middle of the architecture, that mitigates the Vanishing

Gradient Problem.

- ## Questions
  - May there be any obstacles for mixing this architectural approach with Depthwise Separable Convolutions from MobileNets?
- ## Ideas
  - What if we create sparse convolutions of the size $N \times N$ where N is big enough and each convolution is randomly masked with the Probability of p?
    - Seems that it was used widely for some time since this paper was published.
    - > This raises the question whether there is any hope for a next, intermediate step: an architecture that makes use of the extra sparsity, even at filter level, as suggested by the theory, but exploits our current hardware by utilizing computations on dense matrices.
