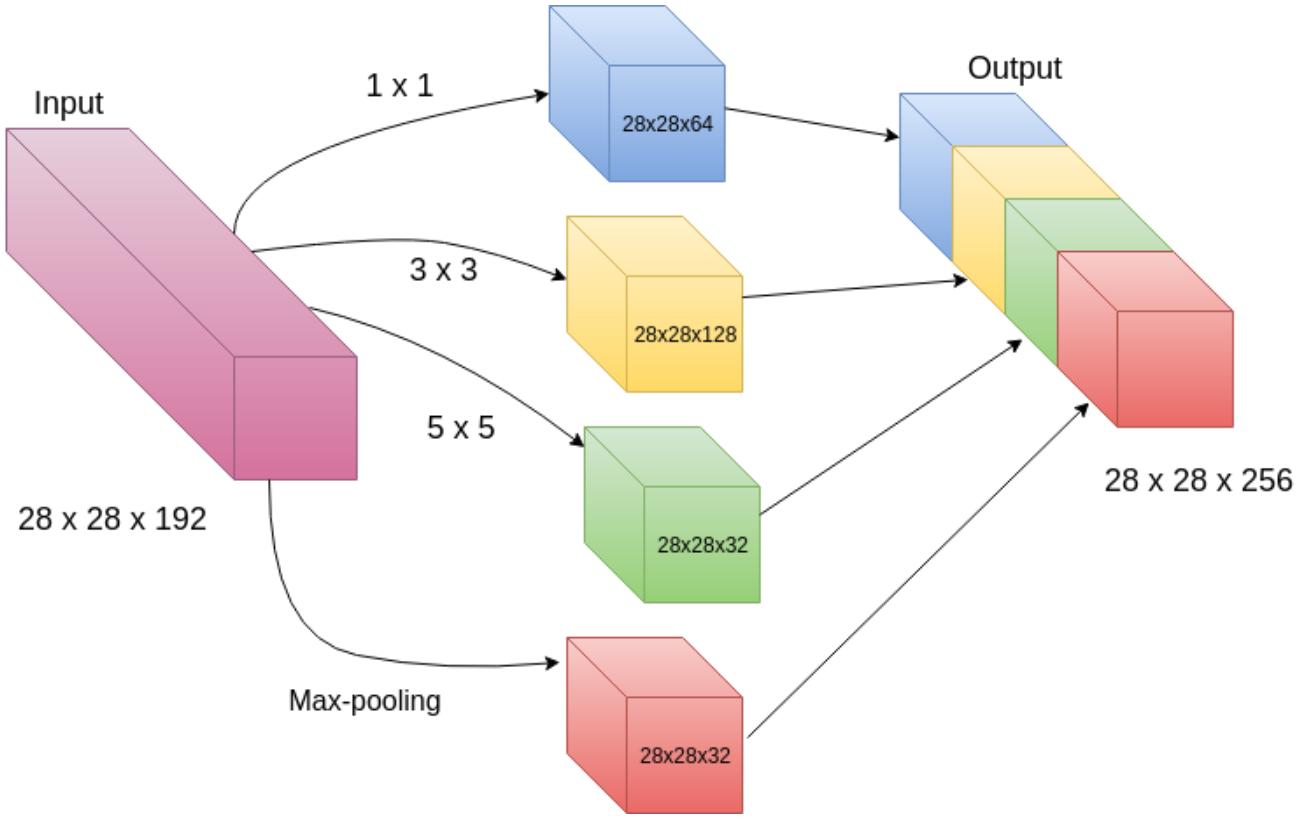
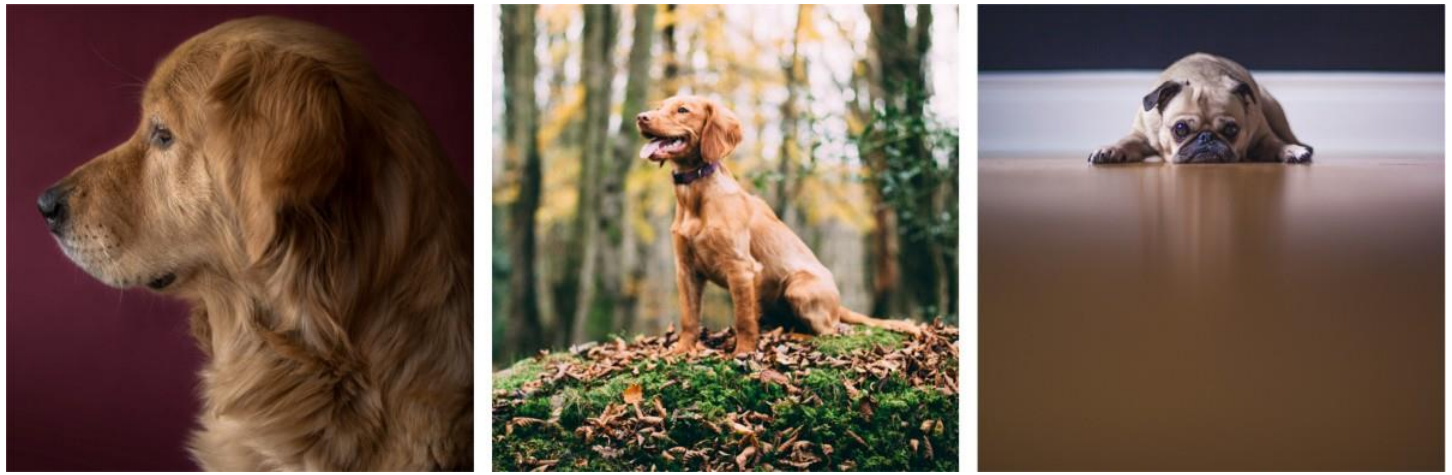
Case Study III: Inception



Challenges in object recognition

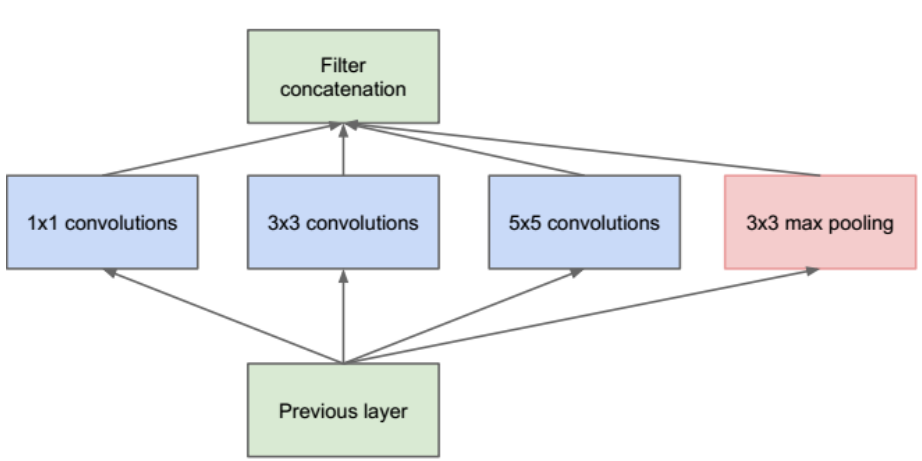
- Salient parts have great variation in sizes
  - Hence, the receptive fields should vary in size accordingly
- Intuitively, deeper models are preferred
  - But very deep nets are prone to overfitting



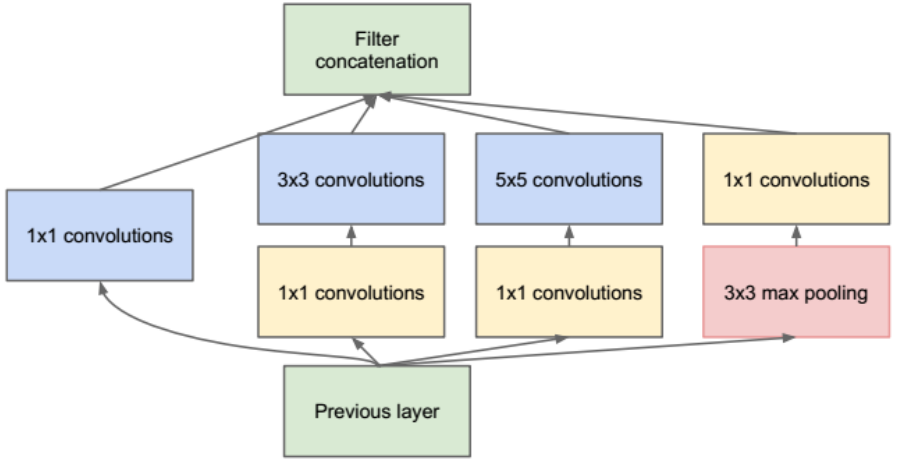
Picture credit: Bharath Raj

Inception module

- Multiple kernel filters of different sizes ($1 \times 1, 3 \times 3, 5 \times 5$)
  - Naïve version
  - Very expensive!
- Add intermediate 1×1 convolutions for compression



(a) Inception module, naïve version



(b) Inception module with dimension reductions

Picture credit: Bharath Raj

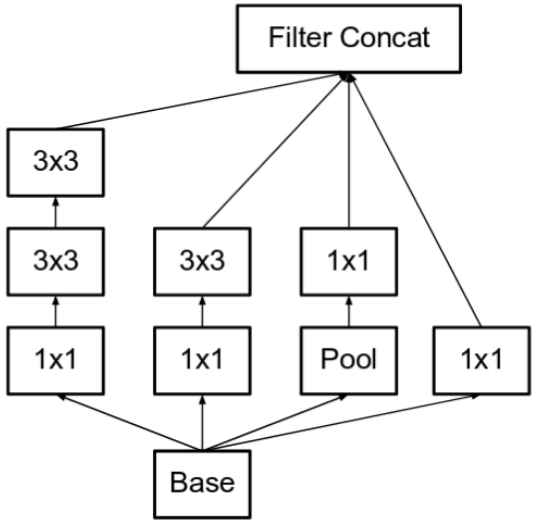
Architecture

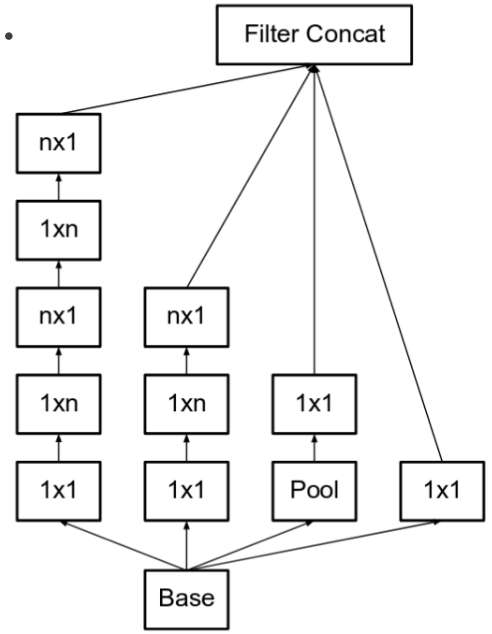
- 9 Inception Modules
  - 22 layers deep (27 with the pooling layers)
  - Global average pooling at the end of last Inception Module
- Because of the increased depth → **Vanishing gradients**
- Inception solution to vanishing gradients: **intermediate classifiers**
  - Intermediate classifiers removed after training
- 6.67% Imagenet error (Alexnet: 18.2%)

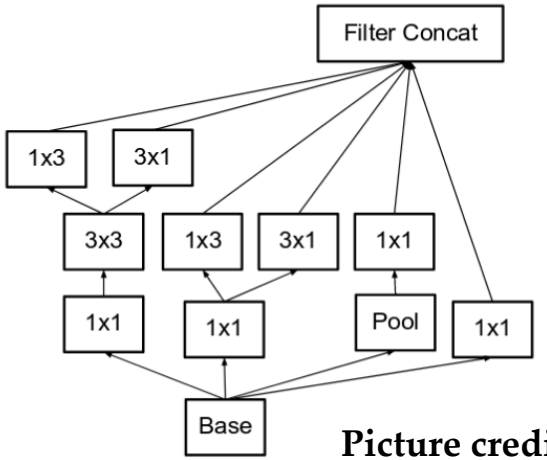
Picture credit: Bharath Raj

Inceptions v2, v3, v4, ….

- Factorize 5×5 in two 3×3 filters
- Factorize $n \times n$ in two $n \times 1$ and $1 \times n$ filters (quite a lot cheaper)
- Make nets wider
- RMSprop, BatchNorms, …







Picture credit: Bharath Raj