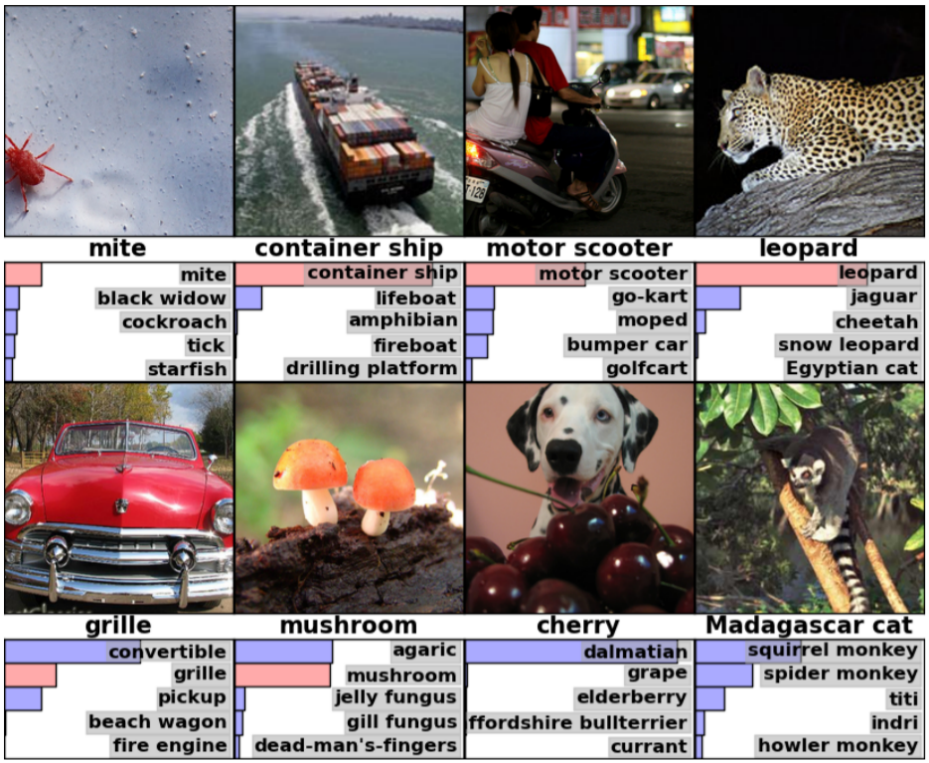
Image classification architectures

Victor Kitov

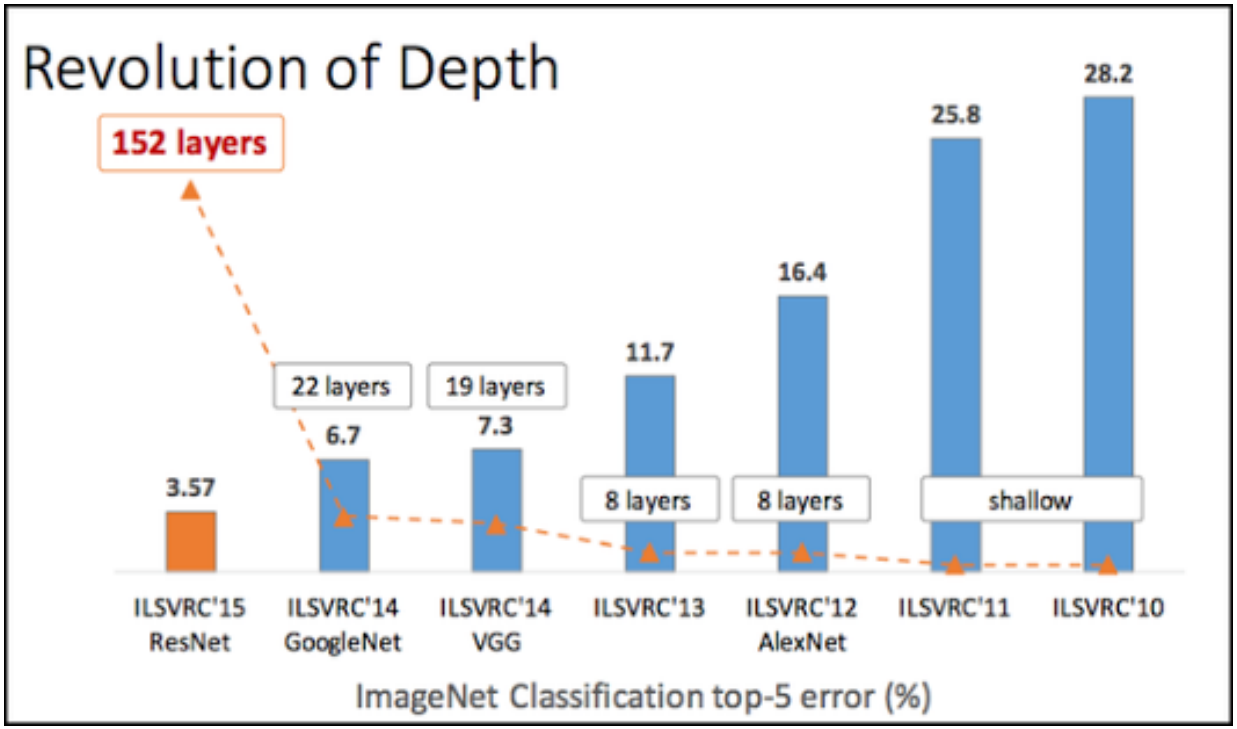
v.v.kitov@yandex.ru

ImageNet classification challenge

- 1000 unambiguous classes (including 120 dog breeds!).
- >1 million hand annotated images.
- Classifiers evaluated by top-5 accuracy
  - is the true class present among top-5 predictions?



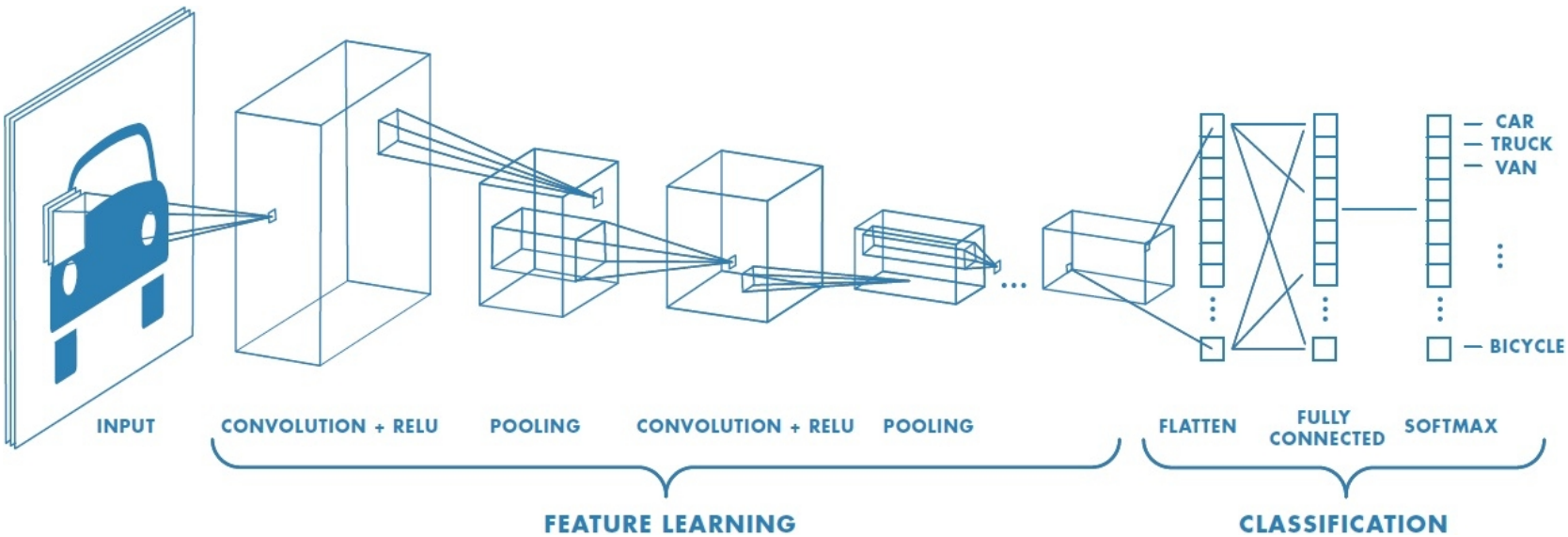
ImageNet challenge progress



- Starting from 2012 - triumph of deep convolutional networks.
- Human performance 3-15% (depending on acquaintance with the classes).[1]

[1] Andrew Karpathy human test.

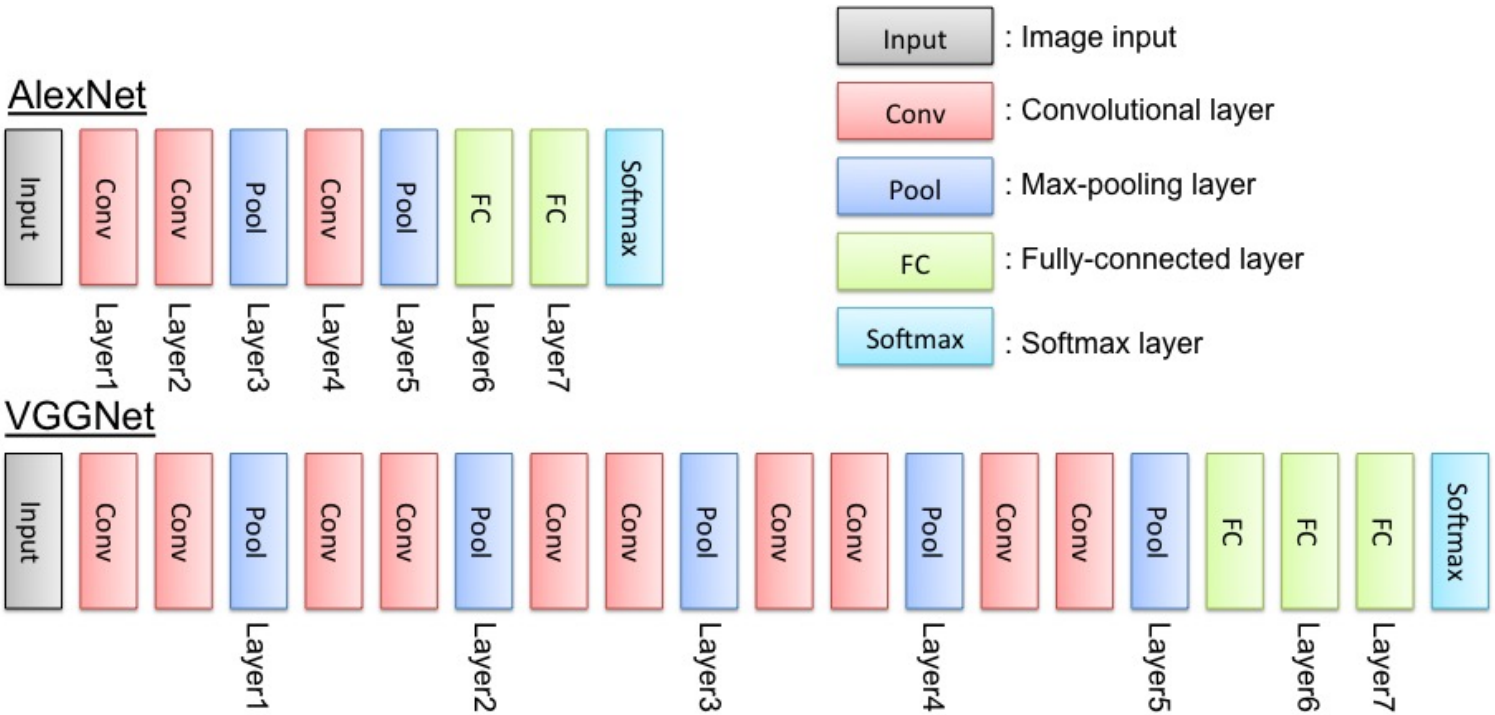
Convolution network



- Later layers learn more and more abstract features.
- Receptive field (in terms of original image) of neurons from deeper layers is wider.

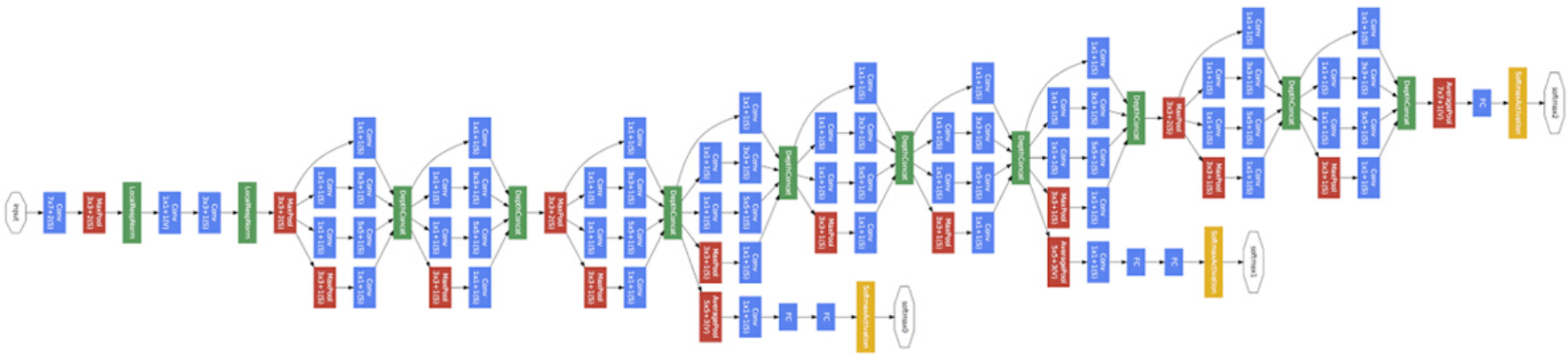
Major CNN architectures

AlexNet vs VGG



Each layer is followed by ReLu non-linearity.

GoogleNet

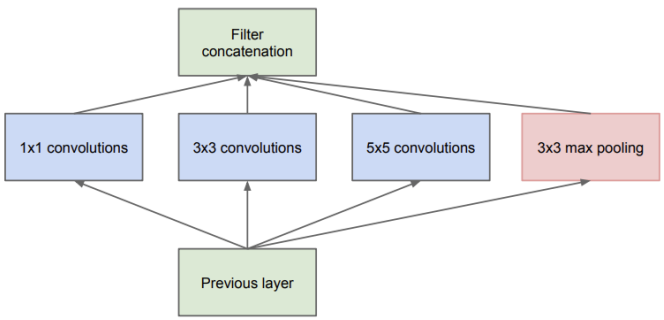


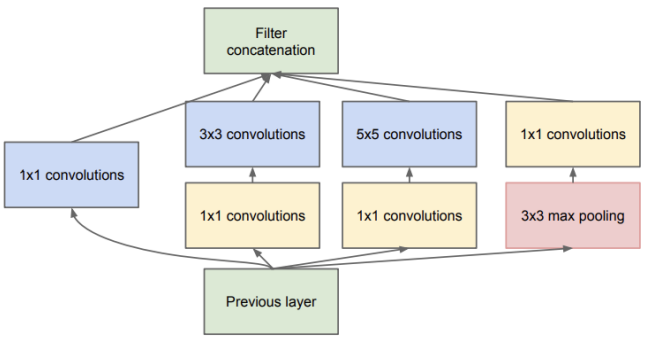
- add intermediate outputs during training
- reduce computation and # parameters by 1x1 convolutions

1x1 convolutions

- 1x1 convolutions provide dimensionality reduction
  - decrease computation and #parameters

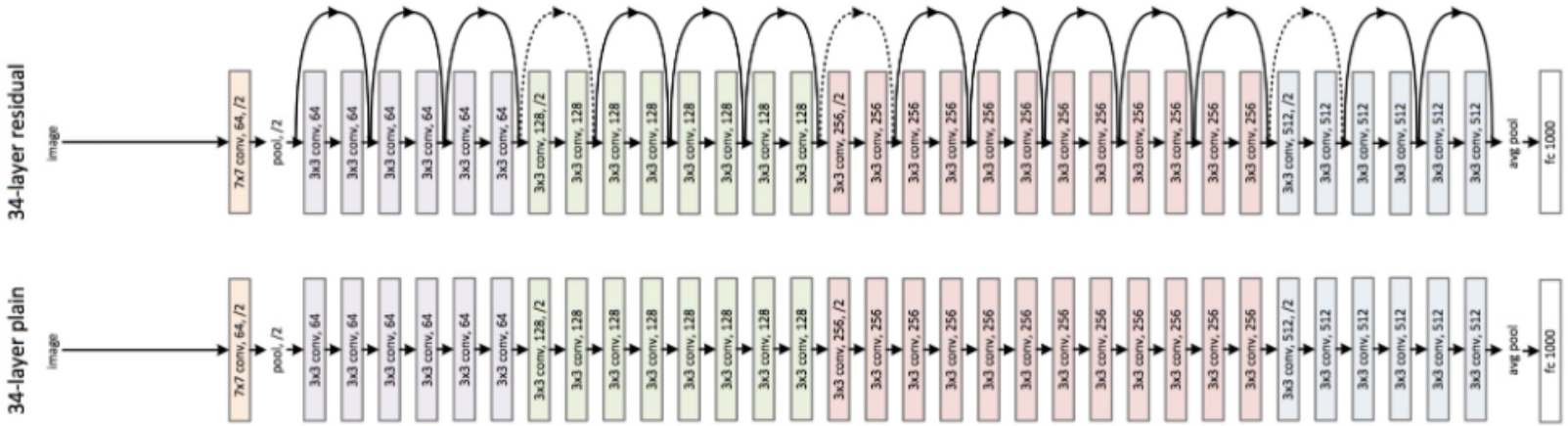
Naive and reduced dimensionality inception blocks:



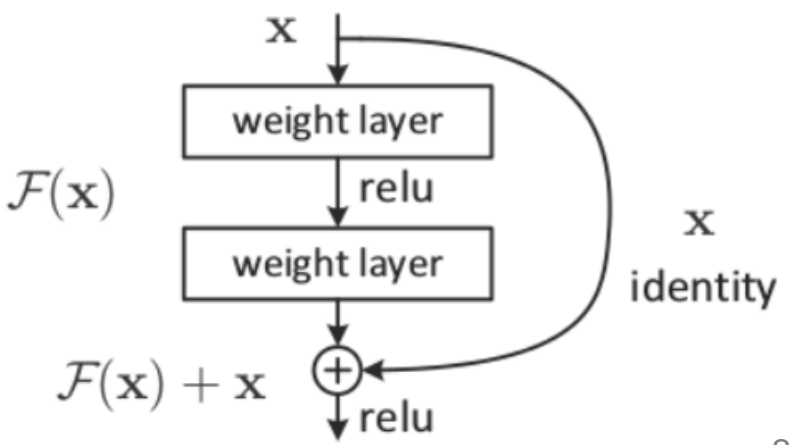


ResNet

ResNet vs. plain network



ResNet building block:



Skip identity connections allow:

- better propagate gradient backwards
- allow more natural initialization of weight layers
  - so that they are almost constant

VGG network[2]

Winners of ImageNet Challenge 2014!

- Data preprocessing - extract mean for R,G,B channels from all pixel intensities.
- Key idea: gradually reduce size and increase receptive field:
  - Filters with a very small receptive field: 3x3, 1x1.
  - Padding to keep original size (1 pixel for 3x3 conv)
  - Stride 1 for conv
  - Max-pooling is performed over a 2×2 pixel window, with stride 2.
- All hidden layers are followed by ReLU

[2] 2015 - Very deep convolutional networks for large-scale image recognition - Simonyan et al.

VGG details

- Optimization: SGD with momentum
  - learning rate decreased 3 times.
- Parallelization over minibatches
  - gradients are then averaged
- First 2 fully connected layers:
  - weight decay regularization
  - dropout regularization
- Train more shallow net, then with learned weights initialize deeper network.
- Dataset augmentation: random scaling, cropping, horizontal flipping and random RGB color shift.

Conclusion

- Train swallow net first, then use it as initialization to deep net.
    - or use intermediate outputs to make predictions to backprop to earlier layers.
- Use dataset augmentation.
- Use convolutions and poolings of small size.
- Use regularization (L_2, dropout, batchnorm)
- Use identity skip-connections