Аугментации и основные сверточные архитектуры

Марк Блуменау

Аугментации

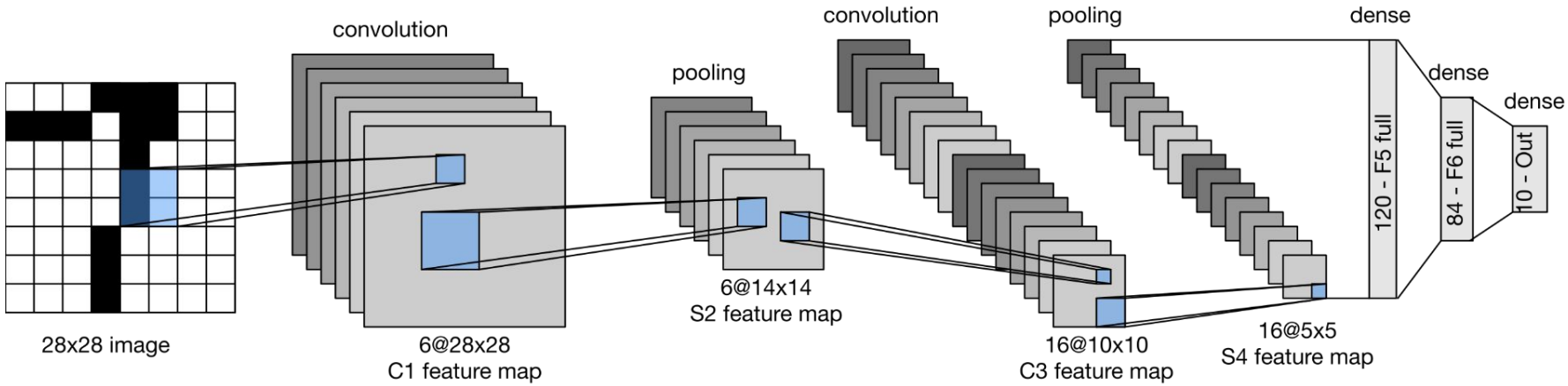
1) “Увеличиваем” датасет бесплатно
2) “Регуляризуем” модель

Применять к текущему батчу, либо нагенерить заранее (смотря что не жалко)

https://github.com/albumentations-team/albumentations

LeNet-5 (1995)

60K параметров + Аугментации = меньше 1% ошибок на MNIST



https://github.com/d2l-ai/d2l-en

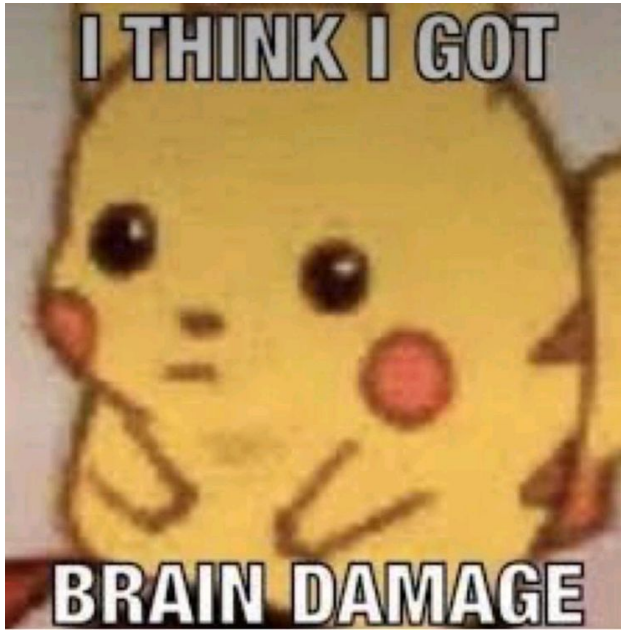
Мемы с Yann LeCun

Yann LeCun
@ylecun

Training with large minibatches is bad for your health.
More importantly, it's bad for your test error.
Friends dont let friends use minibatches larger than 32.
arxiv.org/abs/1804.07612

2:00 PM · Apr 26, 2018 · Facebook

Literally me:



А кроме MNIST что-то есть вообще? CIFAR-10

60к картинок 32x32

10 классов

https://www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf

https://paperswithcode.com/dataset/cifar-10

ImageNet

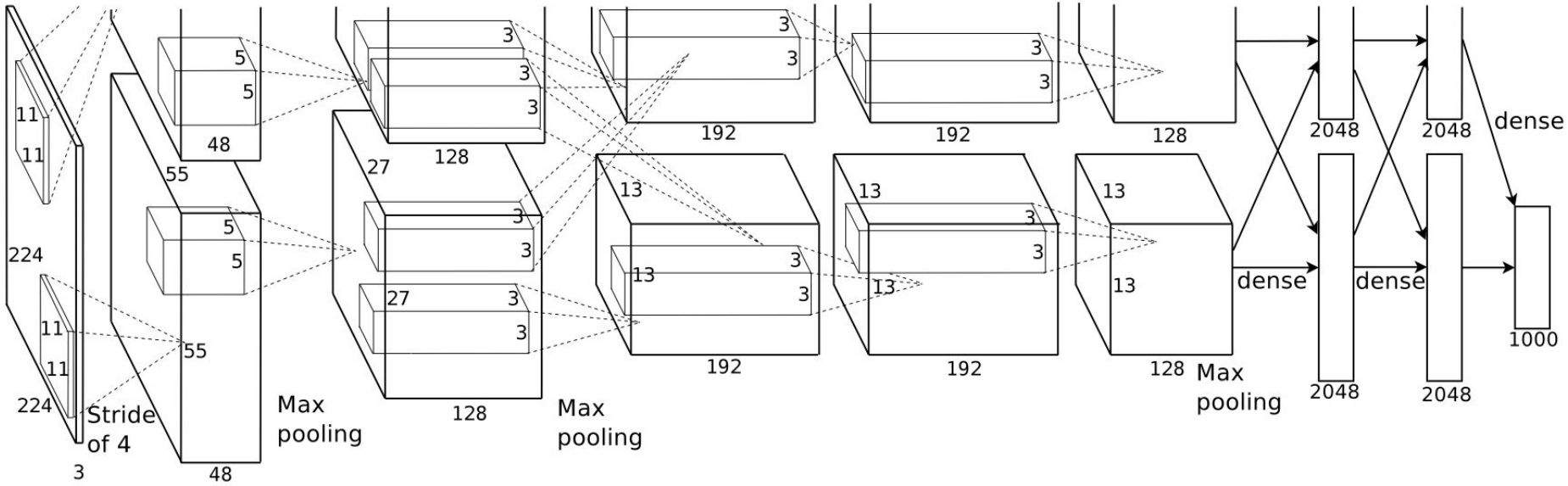
ImageNet Large Scale Visual Recognition Challenge (ILSVRC)

Изначально 1 млн картинок на 1000 классов, сейчас 14 млн.

https://image-net.org/challenges/LSVRC/

https://www.image-net.org/

AlexNet (2012)



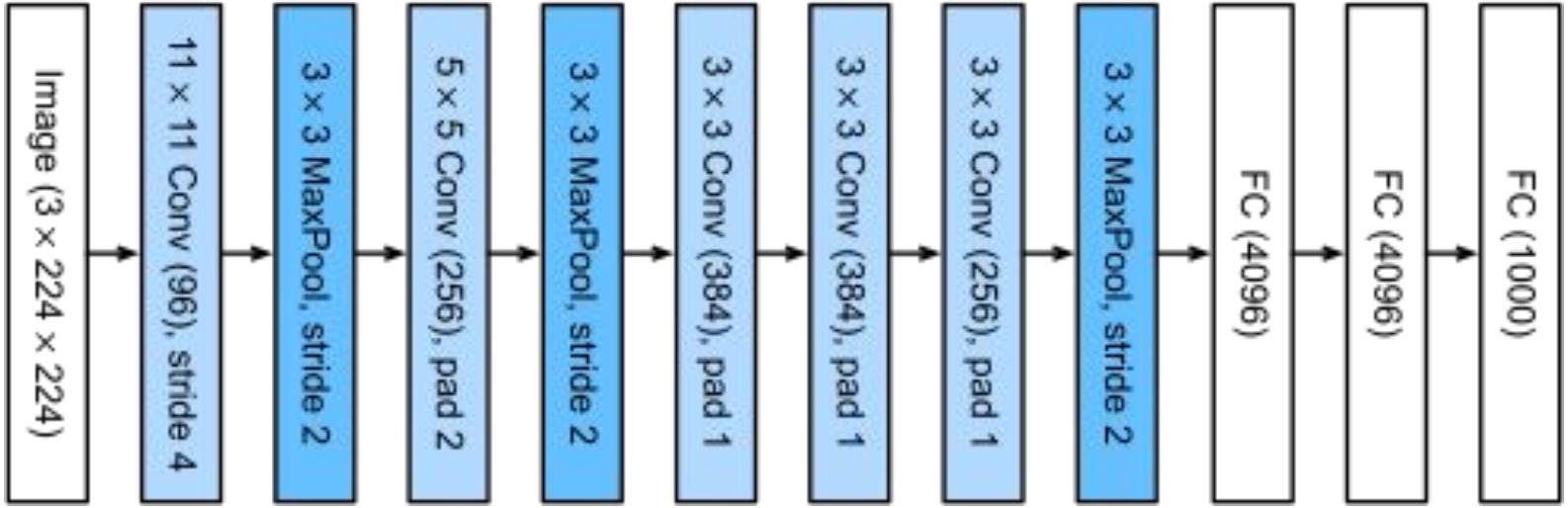
https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf

А что с архитектурой за раздвоение?

Figure 2: An illustration of the architecture of our CNN, explicitly showing the delineation of responsibilities between the two GPUs. One GPU runs the layer-parts at the top of the figure while the other runs the layer-parts at the bottom. The GPUs communicate only at certain layers. The network's input is 150,528-dimensional, and the number of neurons in the network's remaining layers is given by 253,440–186,624–64,896–64,896–43,264–4096–4096–1000.

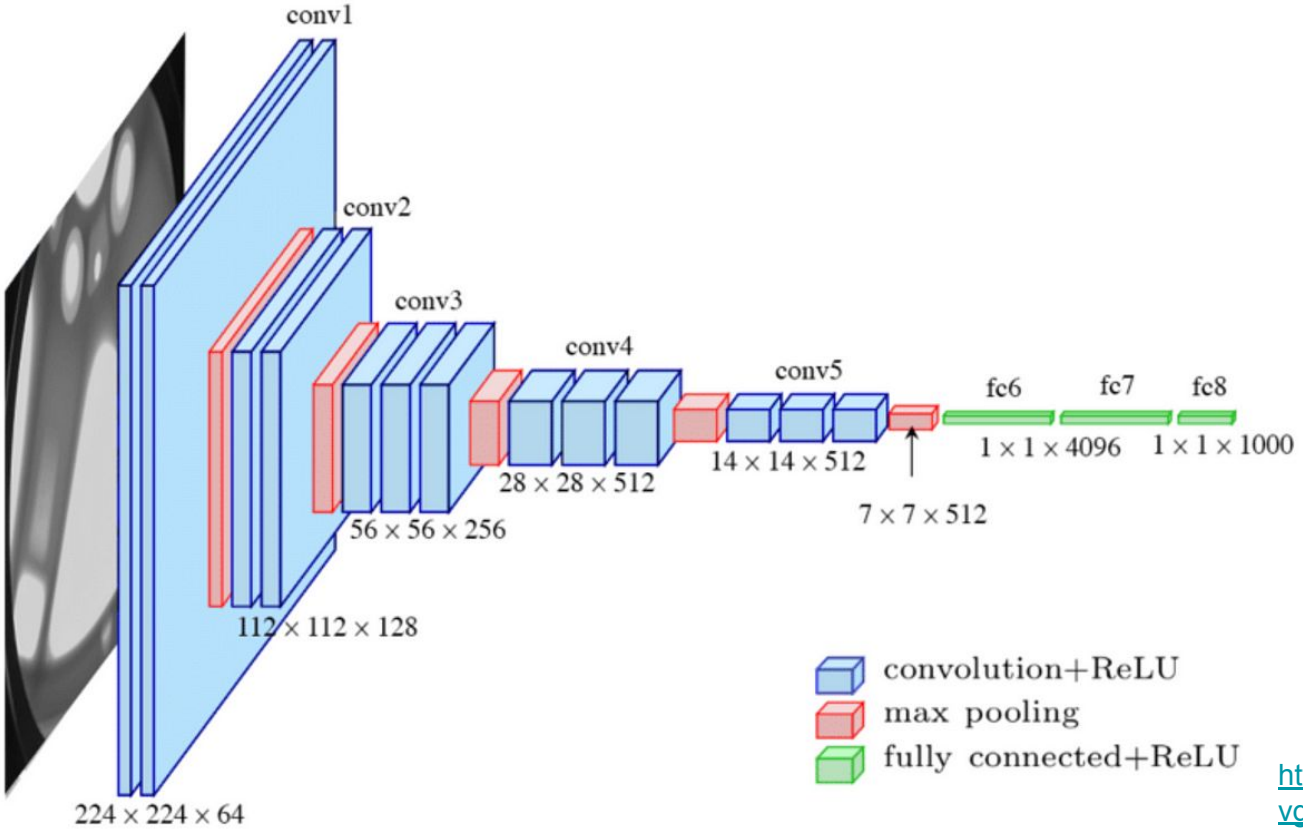
https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf

Более понятная картинка



https://github.com/d2l-ai/d2l-en

VGG (2014)



https://lekhuyen.medium.com/an-overview-of-vgg16-and-nin-models-96e4bf398484

VGG (2014)

Хитрость: у более старших сеток не обучаем с нуля, а берем веса из маленьких

Свертки все маленькие, нет больших фильтров

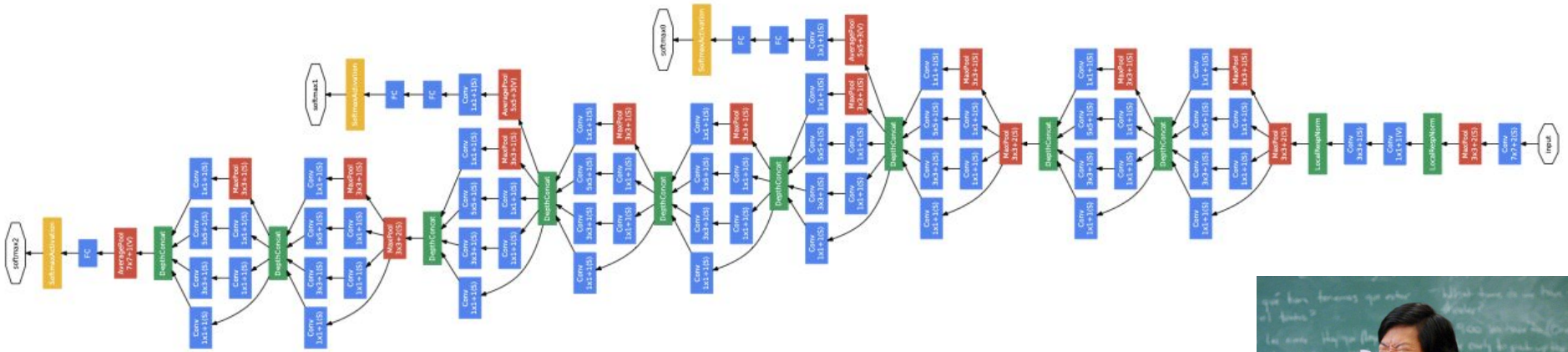
ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

Table 2: **Number of parameters** (in millions).

Network	A,A-LRN	B	C	D	E
Number of parameters	133	133	134	138	144

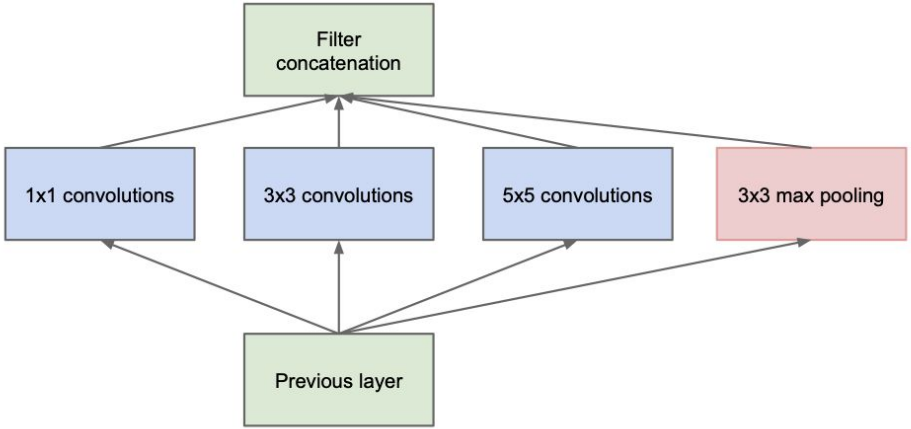
https://arxiv.org/pdf/1409.1556 The batch size was set to 256.

GoogLeNet aka Going deeper with convolutions (2014)

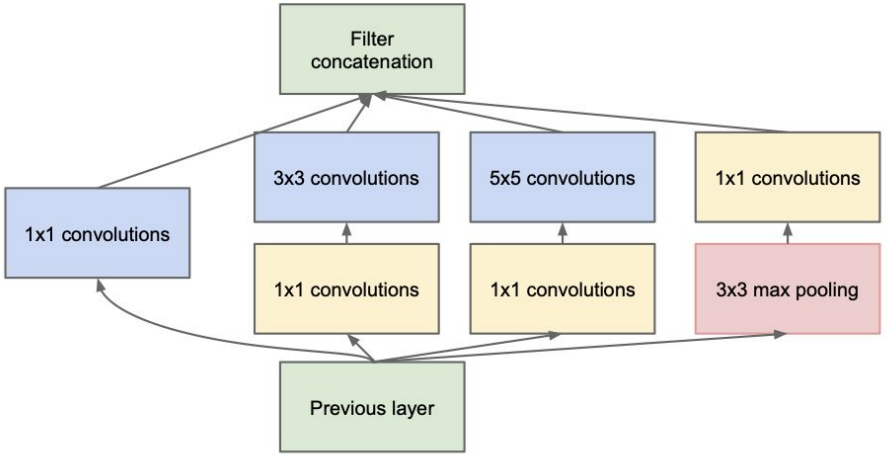


https://arxiv.org/pdf/1409.4842

Inception module



(a) Inception module, naïve version



(b) Inception module with dimension reductions

https://arxiv.org/pdf/1409.4842

ResNet (2015)

We have a problem

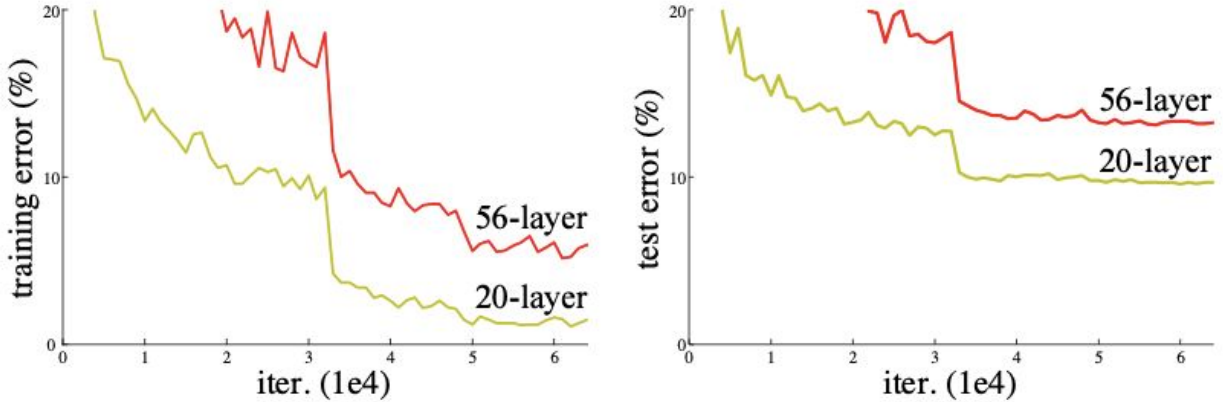
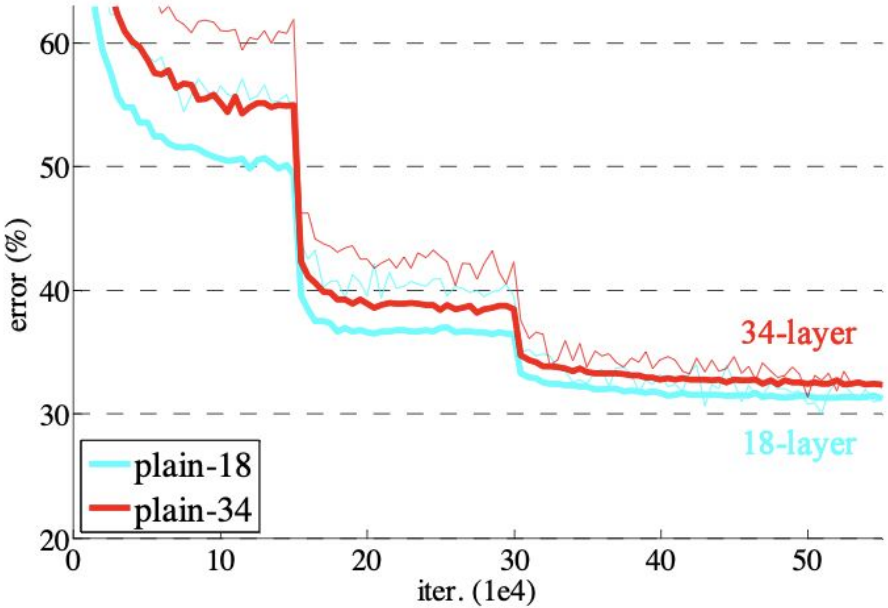
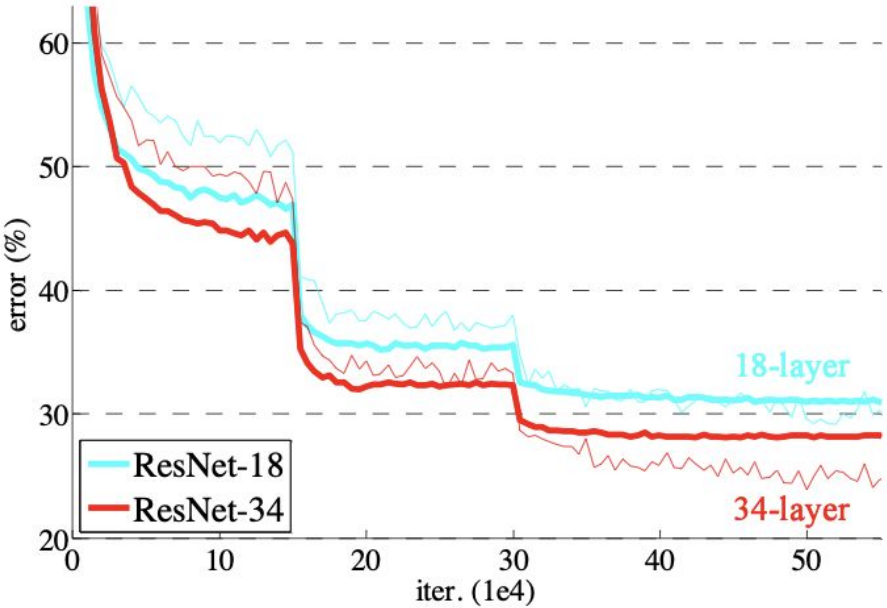


Figure 1. Training error (left) and test error (right) on CIFAR-10 with 20-layer and 56-layer "plain" networks. The deeper network has higher training error, and thus test error. Similar phenomena on ImageNet is presented in Fig. 4.

https://arxiv.org/abs/1512.03385

О, мы решили проблему





https://arxiv.org/abs/1512.03385

А как? (Основа всего CV)

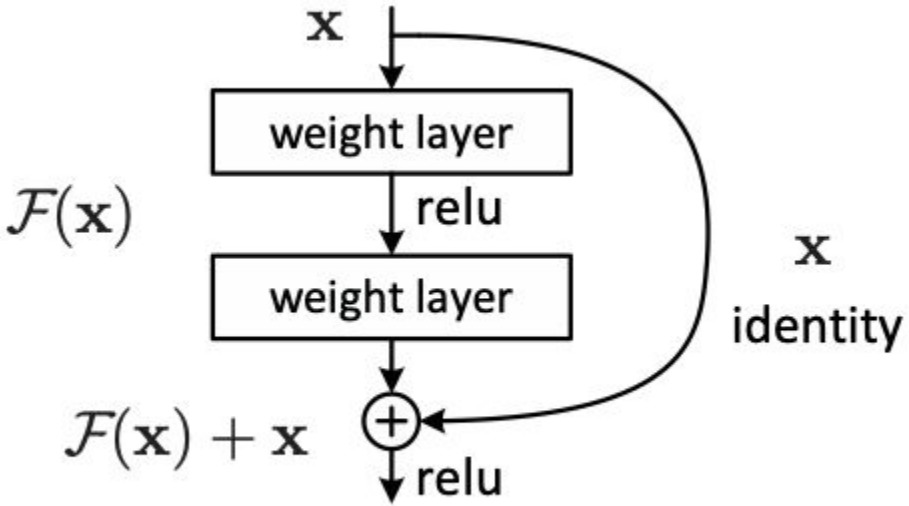
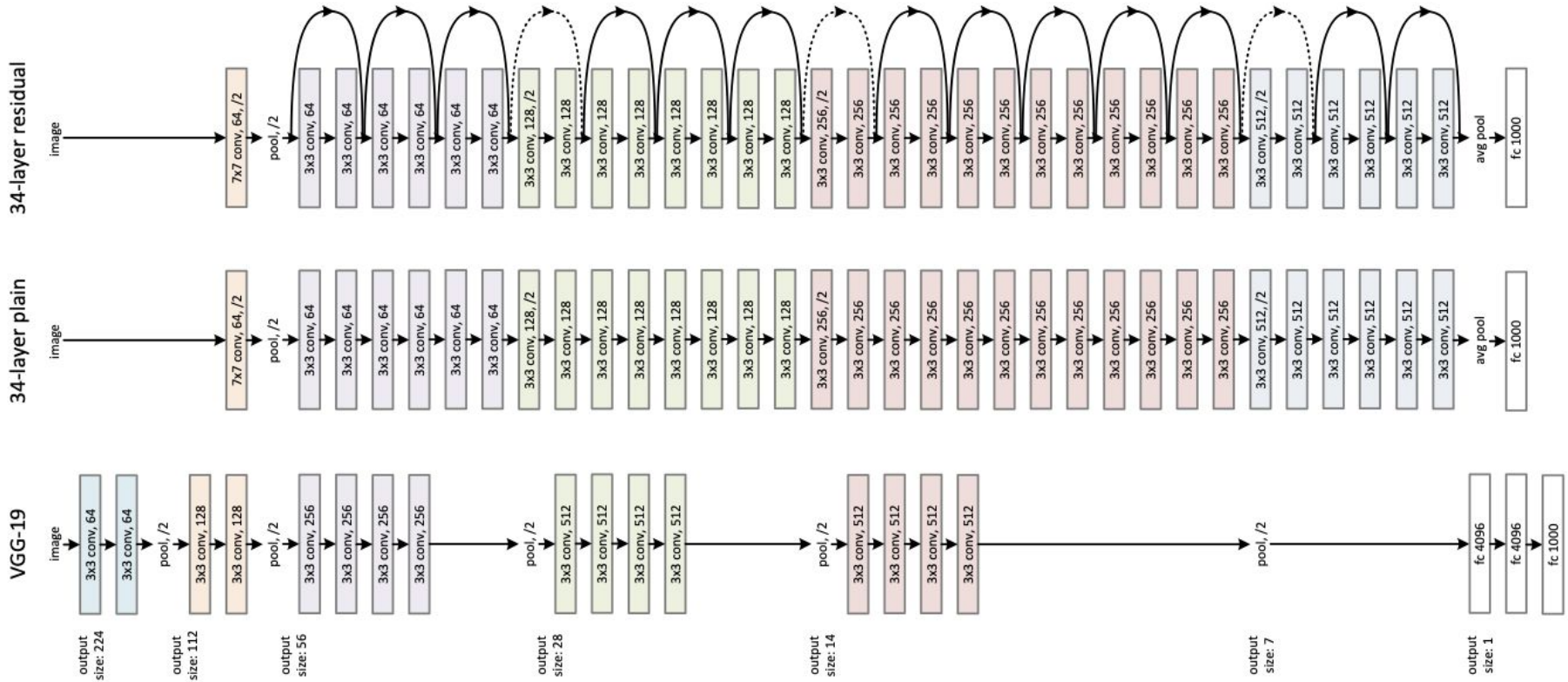


Figure 2. Residual learning: a building block.

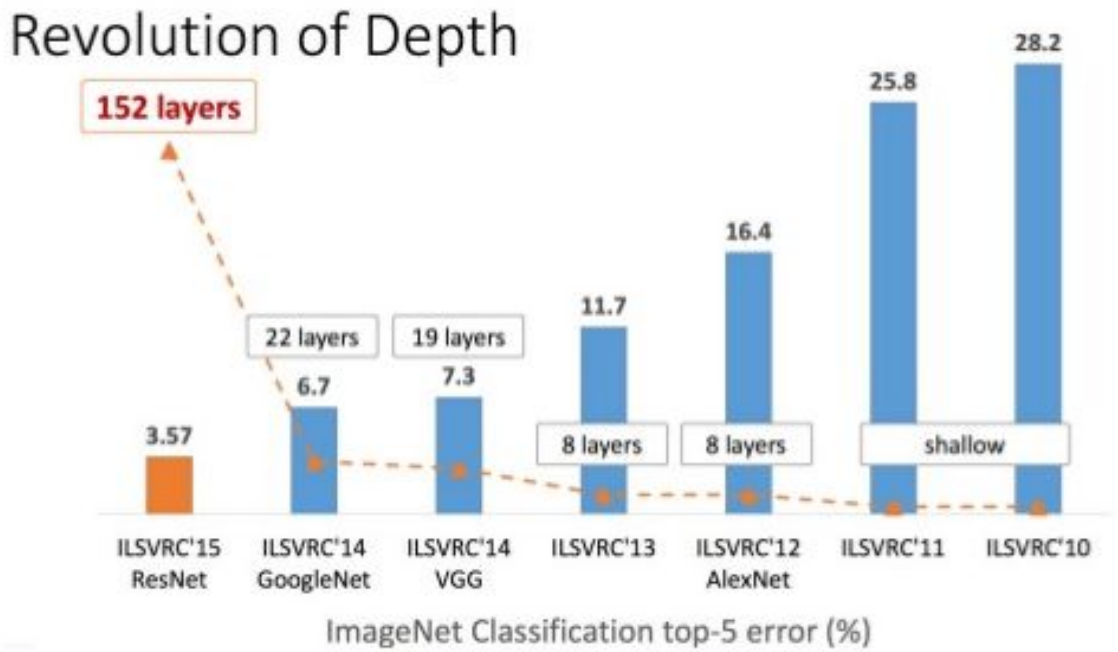
https://arxiv.org/abs/1512.03385

Как выглядит сетка



https://arxiv.org/abs/1512.03385

Что мы рассмотрели:



https://www.researchgate.net/figure/Performance-of-different-approaches-in-ImageNet-2015-competition_fig2_309392322

Что есть дальше? (курс CV)

Squeeze-and-Excite

MBConv (see MobileNet/EfficientNet)

Fused MBConv (see EfficientNetV2)

Пирамидки

ConvNext

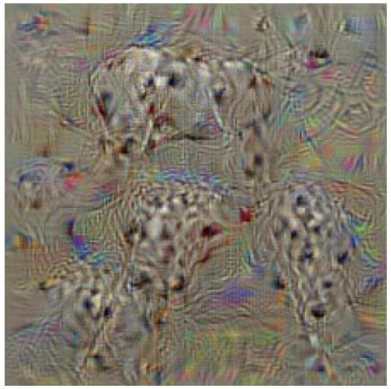
ViT

…

Интерпретируемость

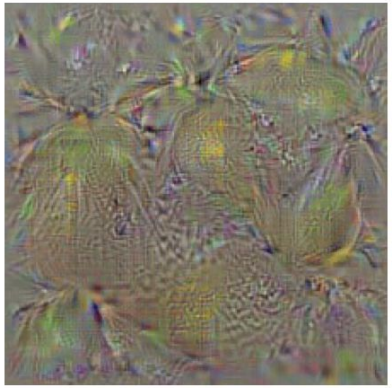
dumbbell

cup

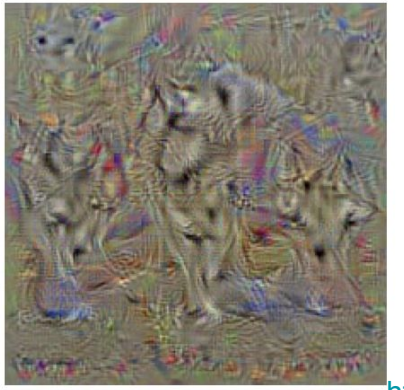


dalmatian

bell pepper

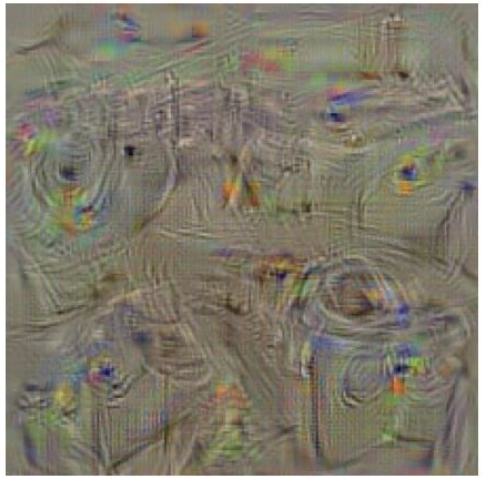


lemon

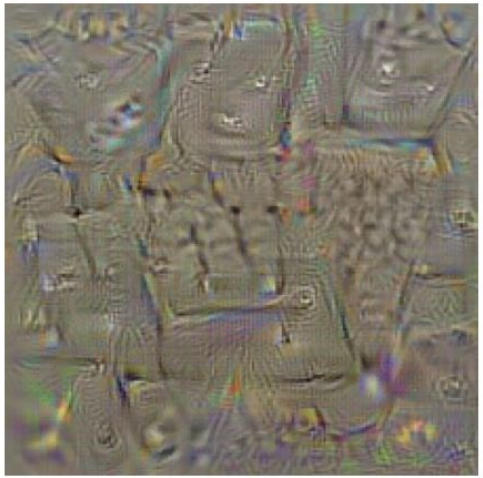


husky

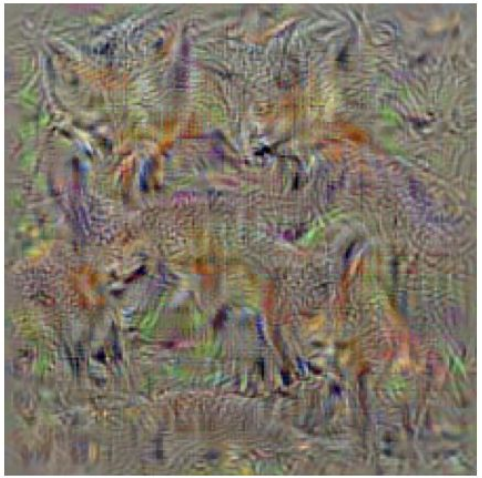
https://arxiv.org/pdf/1312.6034



washing machine



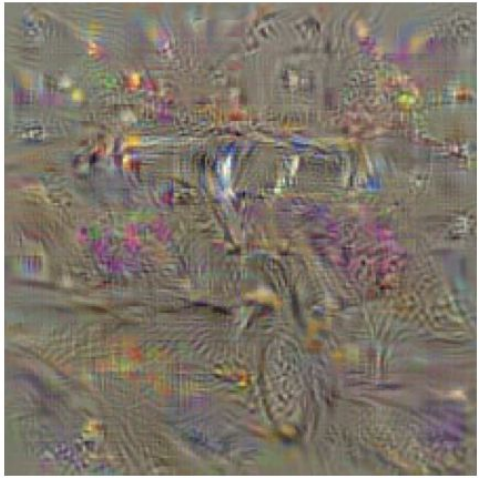
computer keyboard



kit fox

goose

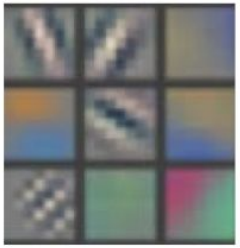
ostrich



limousine

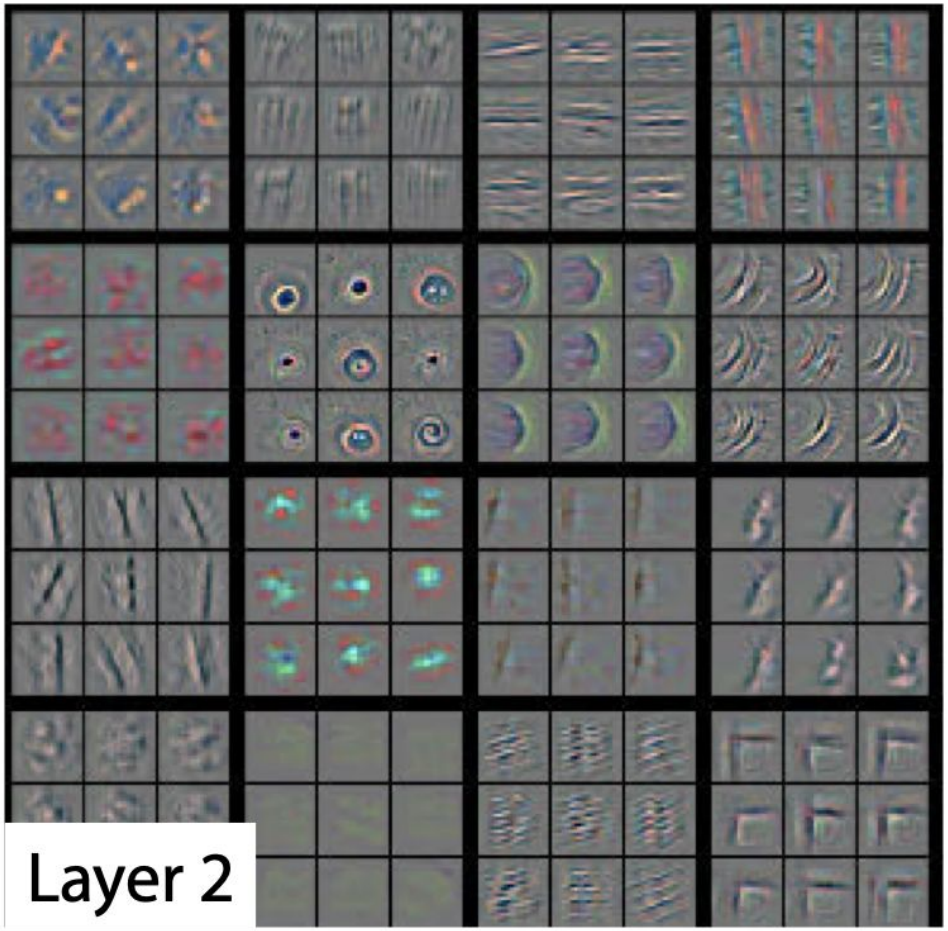
https://arxiv.org/pdf/1312.6034

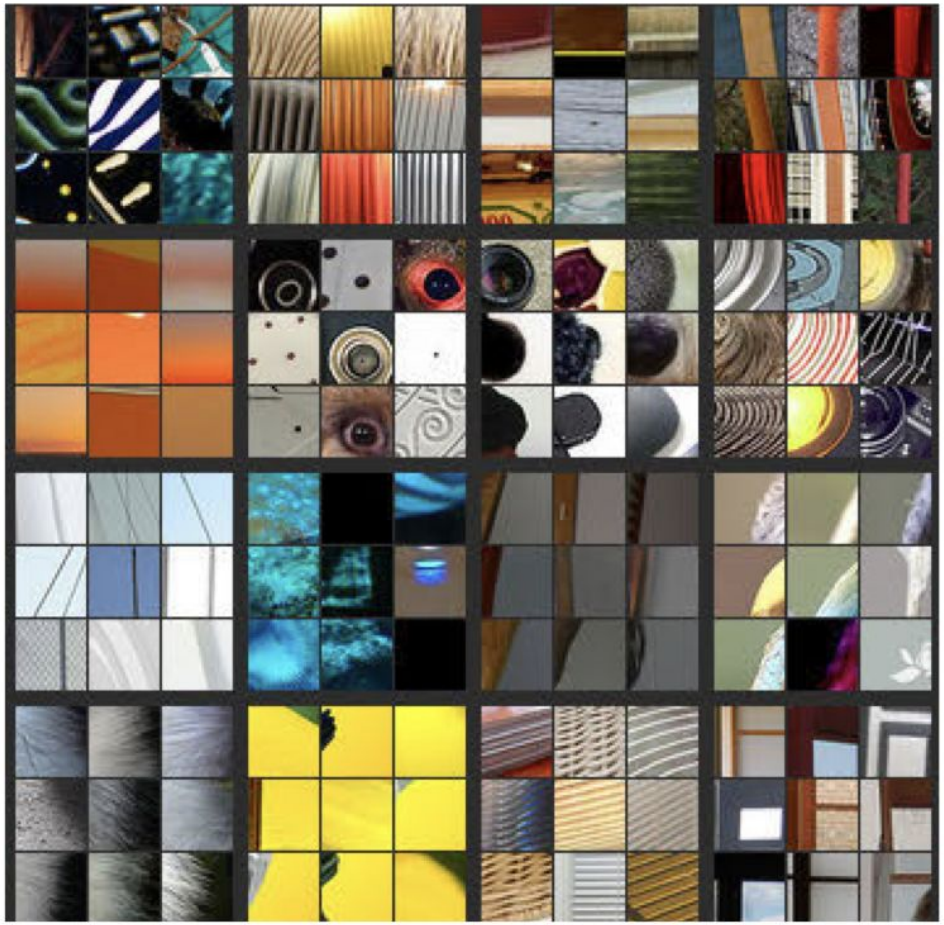
https://arxiv.org/pdf/1311.2901



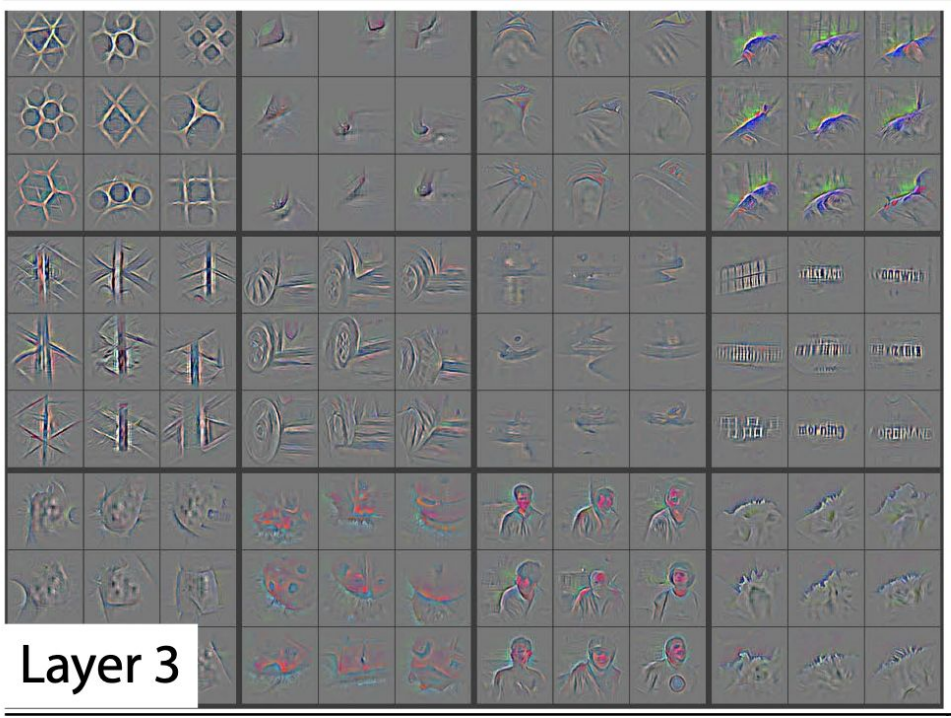
Layer 1

https://arxiv.org/pdf/1311.2901





https://arxiv.org/pdf/1311.2901



https://arxiv.org/pdf/1311.2901