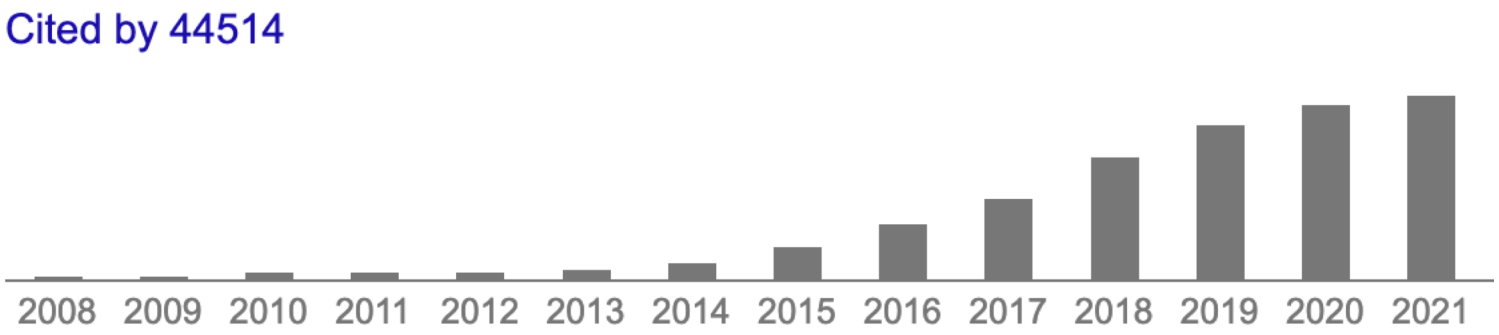
Машинное обучение в науках о Земле

Историческая справка

- 1998: LeNet-5
  - Yann LeCun, Léon Bottou, Yoshua Bengio, Patrick Haffner, “Gradient-based learning applied to document recognition”

Fig. 4. Size-normalized examples from the MNIST database.



Историческая справка

- 1998: LeNet-5
  - Yann LeCun, Léon Bottou, Yoshua Bengio, Patrick Haffner, “Gradient-based learning applied to document recognition”

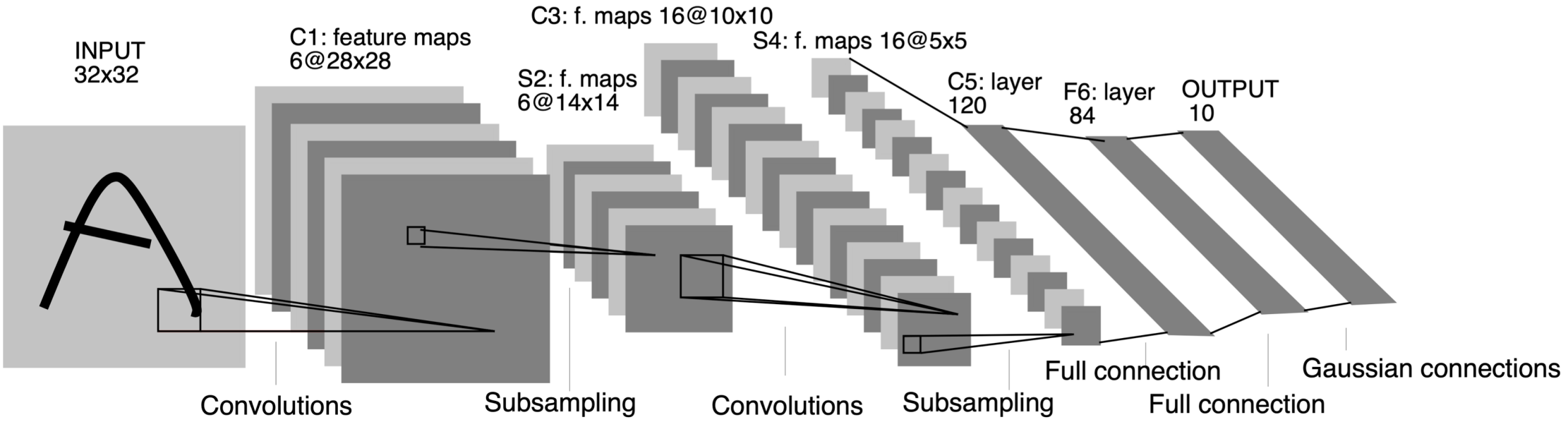


Fig. 2. Architecture of LeNet-5, a Convolutional Neural Network, here for digits recognition. Each plane is a feature map, i.e. a set of units whose weights are constrained to be identical.

IMAGENET Construction and Analysis of a Large Scale Image Ontology

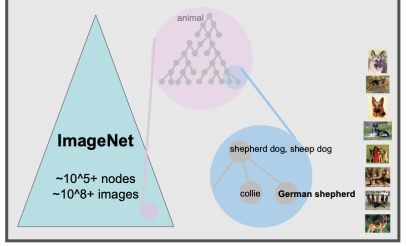
Li Fei- Fei [1,2] Jia Deng [1] Minh Do [1] Hao Su [1] Kai Li [1]

1. Computer Science Department, Princeton University, USA 2. Psychology Department, Princeton University, USA

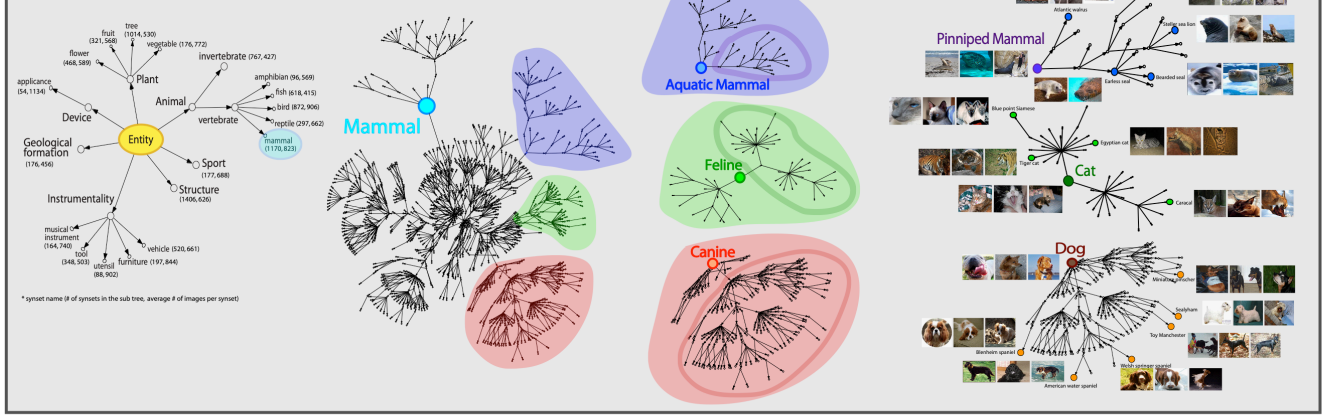
correspondence: feifeili@princeton.edu

ImageNet Overview

- An image ontology database
- Based on the WordNet backbone [Fellbaum98]
- Every node is a synonym set, or 'synset', depicting a particular concept
- ~100,000 noun synsets
- 500~2000 images per synset



ImageNet Trees

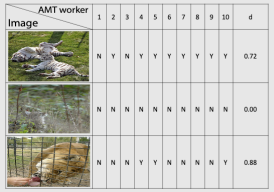


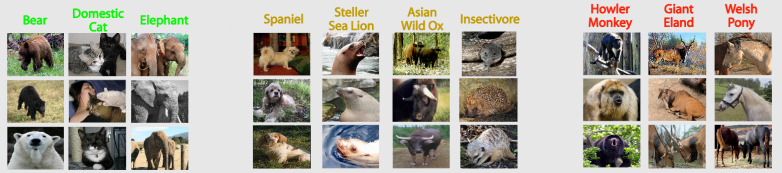
Synset Discriminability

- What do we have? Multiple AMT workers vote on whether an image belongs to a synset
- Intuition. Divergence (d) of votes reflect discriminability of the image: the higher the d, the less discriminable the image.
- How do we measure? Information theoretic analysis (entropy)

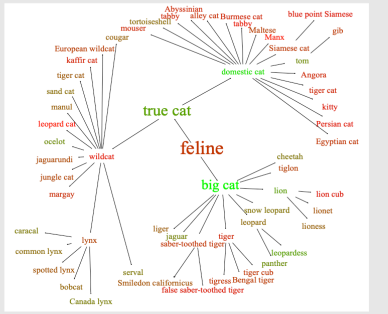
$d(\text{image}) = -(f \log(f) + (1-f)\log(1-f))$ $D(\text{synset}) = \text{average}(d)$

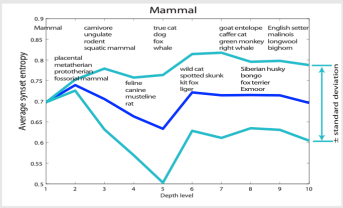
* where f is the normalized frequency of the 'yes' votes the image receives

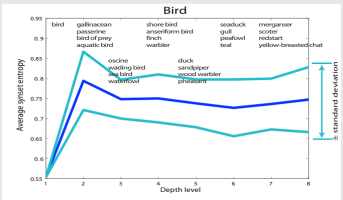


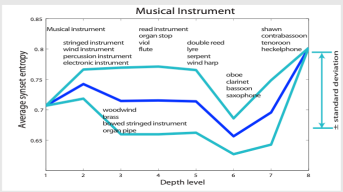


more "discriminable" Synsets — less "discriminable" Synsets





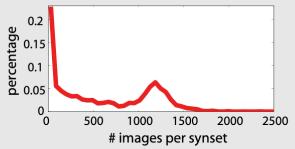




- Synsets along the WordNet semantic hierarchy tree paths display patterns of discriminability
- More discriminable synsets tend to agree with "basic level" categorization of Rosch et al. 1978

Properties of ImageNet

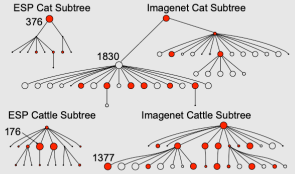
Scale



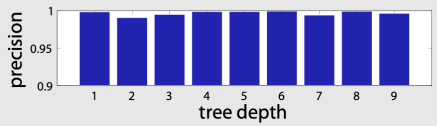
Comparison with others

	ImageNet	TinyImage	LabelMe	ESP	LHill
LabelDisam	Y	Y	N	N	Y
Clean	Y	N	Y	Y	Y
DenseHie	Y	Y	N	N	N
FullRes	Y	N	Y	Y	Y
PublicAvail	Y	Y	Y	N	N
Segmented	N	N	Y	N	Y

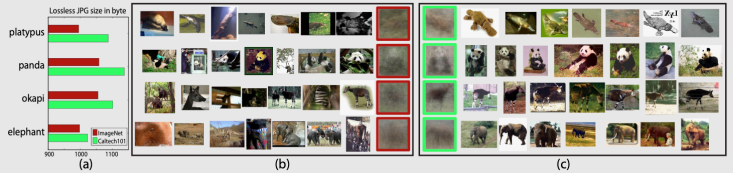
Hierarchy



Accuracy

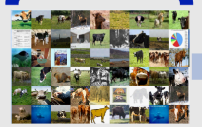


Diversity



Construction of ImageNet

Step 1: Collect Images



- Use multiple search engines (google, msn, yahoo, flickr)
- Use multiple languages (Chinese, Spanish, Dutch and Italian)
- Over 10,000 img/synset on average

Step 2: Clean the Images

Set up a task on Amazon Mechanical Turk (AMT)

- An online global workers' market
- Host online tasks for clients

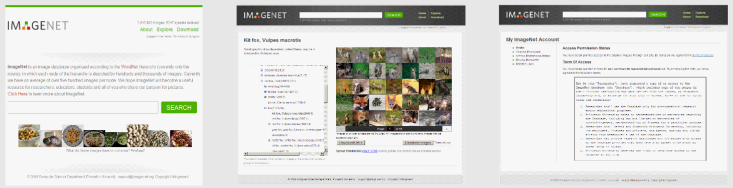
Workers do annotation on AMT

- Multiple annotations for each image

Annotation result

- An average of >97% accuracy

www.image-net.org



References

J.Deng, W.Dong, R.Socher, L.-J.Li, K.Li and L.Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. CVPR2009

C.Fellbaum, WordNet: An Electronic Lexical Database, MIT Press, 1998

L.von Ahn and L.Dabbish. Labeling images with a computer game. In CHI04, pages 319-326, 2004

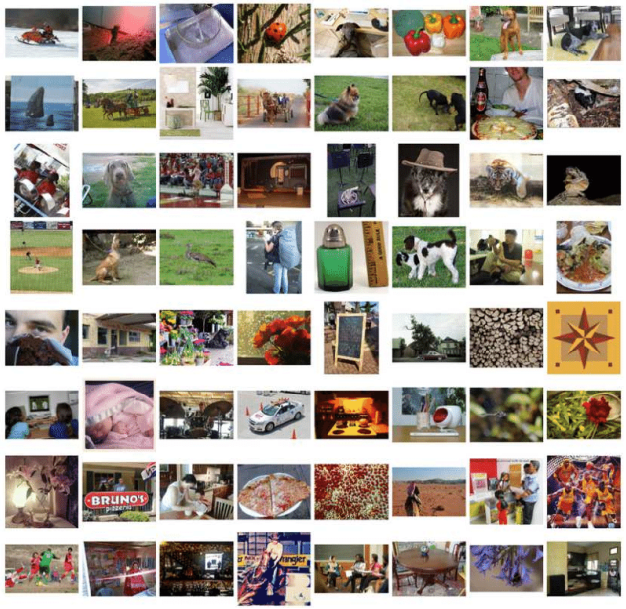
B.Russell, A.Torralba, K.Murphy, and W.Freeman. Labelme: A database and web-based tool for image annotation. IJCV, 77(1-3):157-173, May 2008

B.Yao, X.Yang, and S.Zhu. Introduction to a large-scale general purpose ground truth database: Methodology, annotation tool and benchmarks. In EMMCVPR07, pages 169-183, 2007

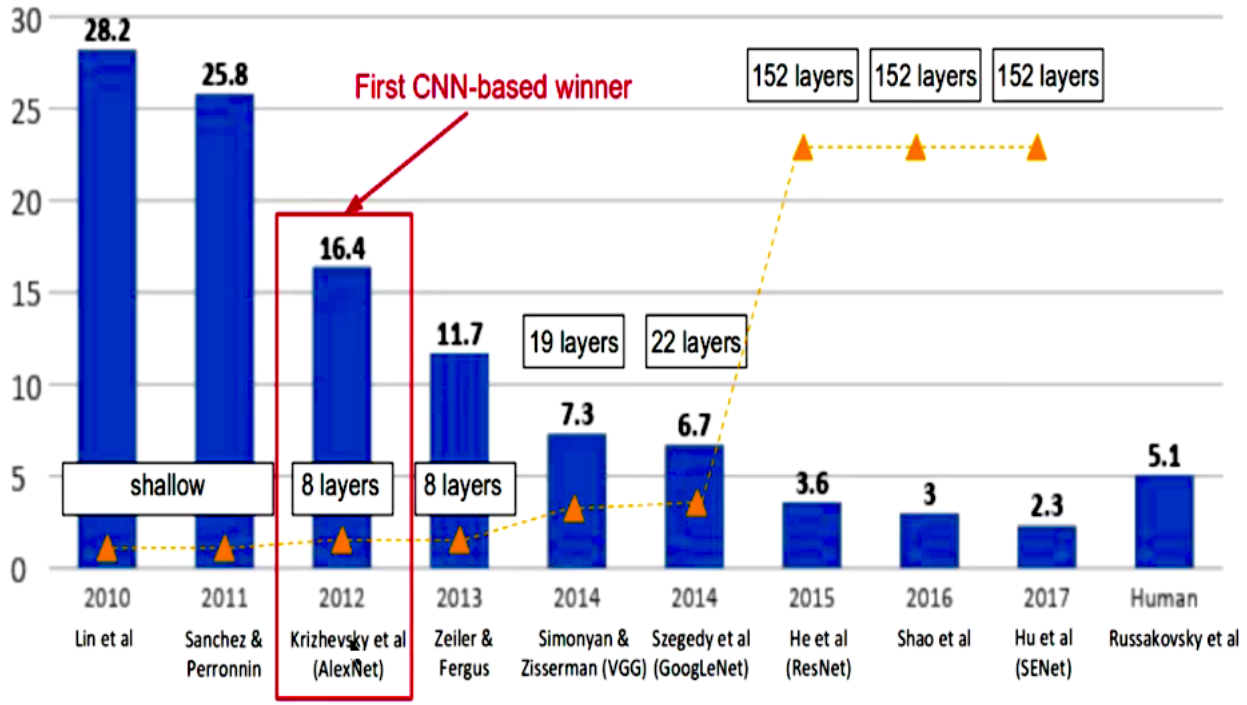
A.Torralba, R.Fergus, and W.Freeman. 80 million tiny images: A large data set for nonparametric object and scene recognition. PAMI, 30(11):1958-1970, November 2008

Историческая справка

- 2009: ImageNet (не сеть!)
  - J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li and L. Fei-Fei, **ImageNet**: A Large-Scale Hierarchical Image Database. IEEE Computer Vision and Pattern Recognition (CVPR), 2009
  - https://image-net.org/
  - 14’197’122 изображений
  - 1000 классов
  - ave. size: 469x387 px.



Историческая справка



AlexNet (2012)

- Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. "**Imagenet classification with deep convolutional neural networks.**" Advances in neural information processing systems (NIPS) 25 (2012).

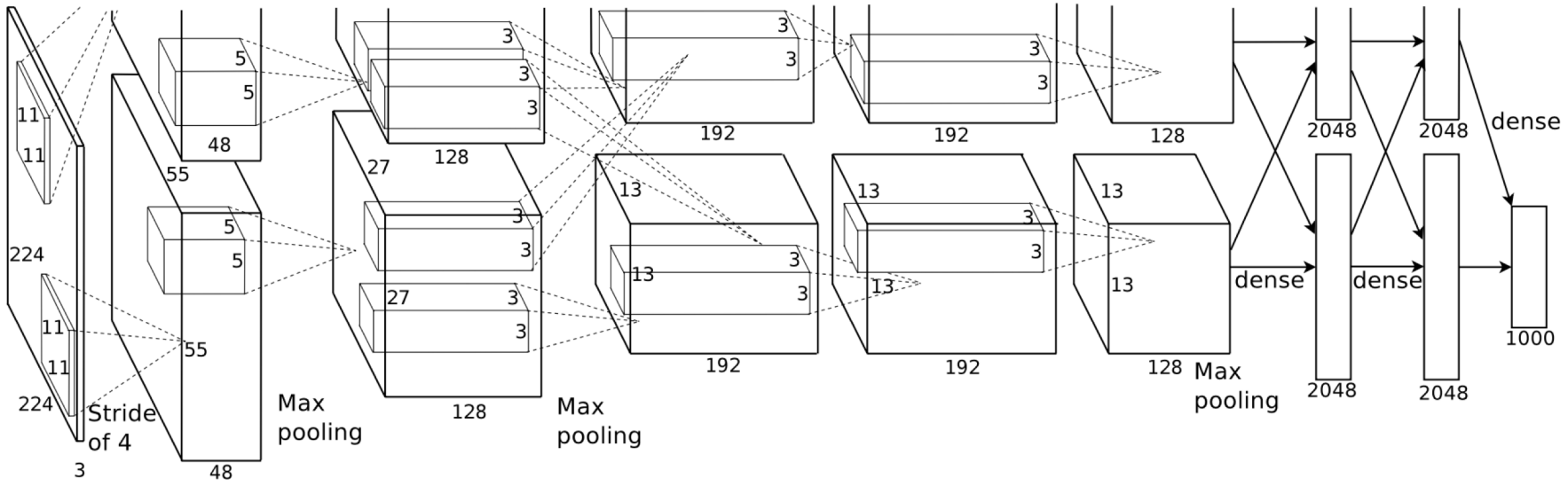
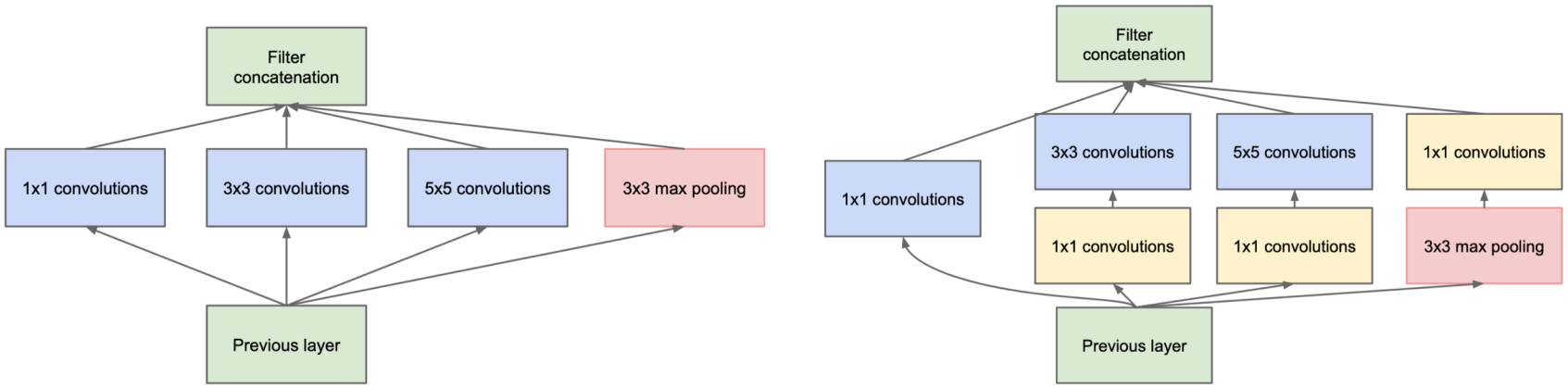


Figure 2: An illustration of the architecture of our CNN, explicitly showing the delineation of responsibilities between the two GPUs. One GPU runs the layer-parts at the top of the figure while the other runs the layer-parts at the bottom. The GPUs communicate only at certain layers. The network's input 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.

GoogLeNet (2014)

- Christian Szegedy, et al. "Going deeper with convolutions." Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). 2015.

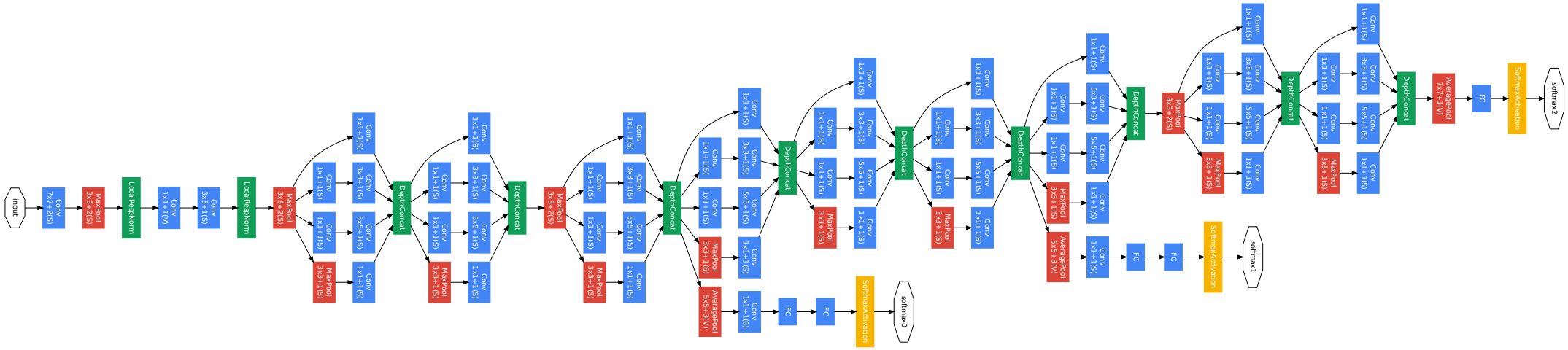


(a) Inception module, naïve version (b) Inception module with dimension reductions

Figure 2: Inception module

GoogLeNet (2014)

- Christian Szegedy, et al. "Going deeper with convolutions." Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). 2015.



GoogLeNet (2014)

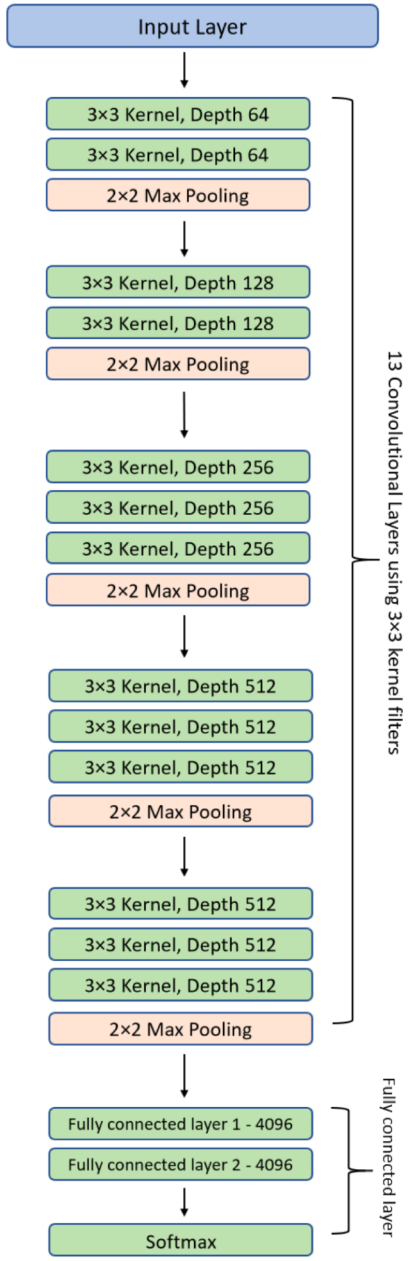
- Christian Szegedy, et al. "Going deeper with convolutions." Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). 2015.

Team	Year	Place	Error (top-5)	Uses external data
SuperVision	2012	1st	16.4%	no
SuperVision	2012	1st	15.3%	Imagenet 22k
Clarifai	2013	1st	11.7%	no
Clarifai	2013	1st	11.2%	Imagenet 22k
MSRA	2014	3rd	7.35%	no
VGG	2014	2nd	7.32%	no
GoogLeNet	2014	1st	6.67%	no

Table 2: Classification performance

VGG (2014)

- Karen Simonyan and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).

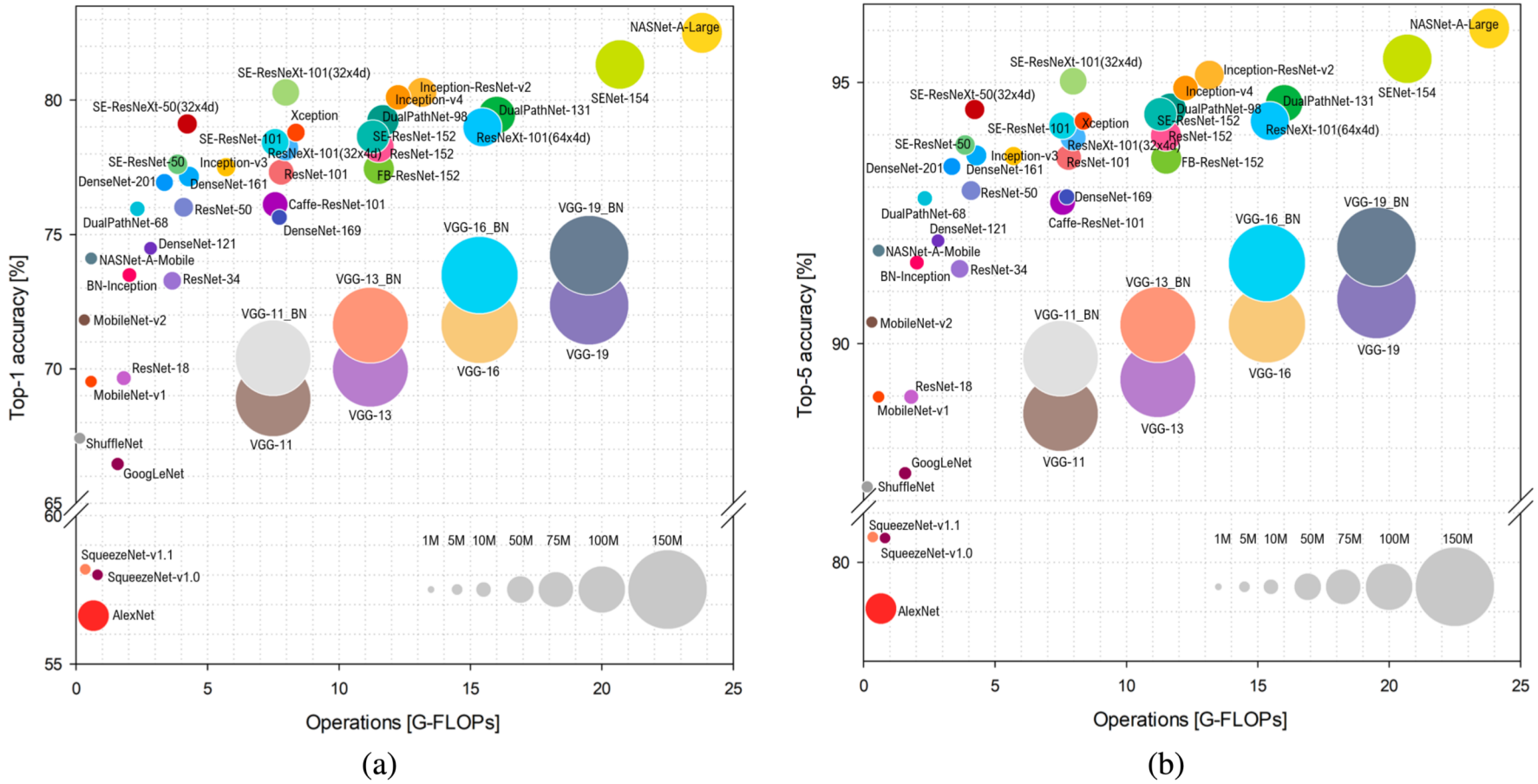


ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	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

Историческая справка



Историческая справка

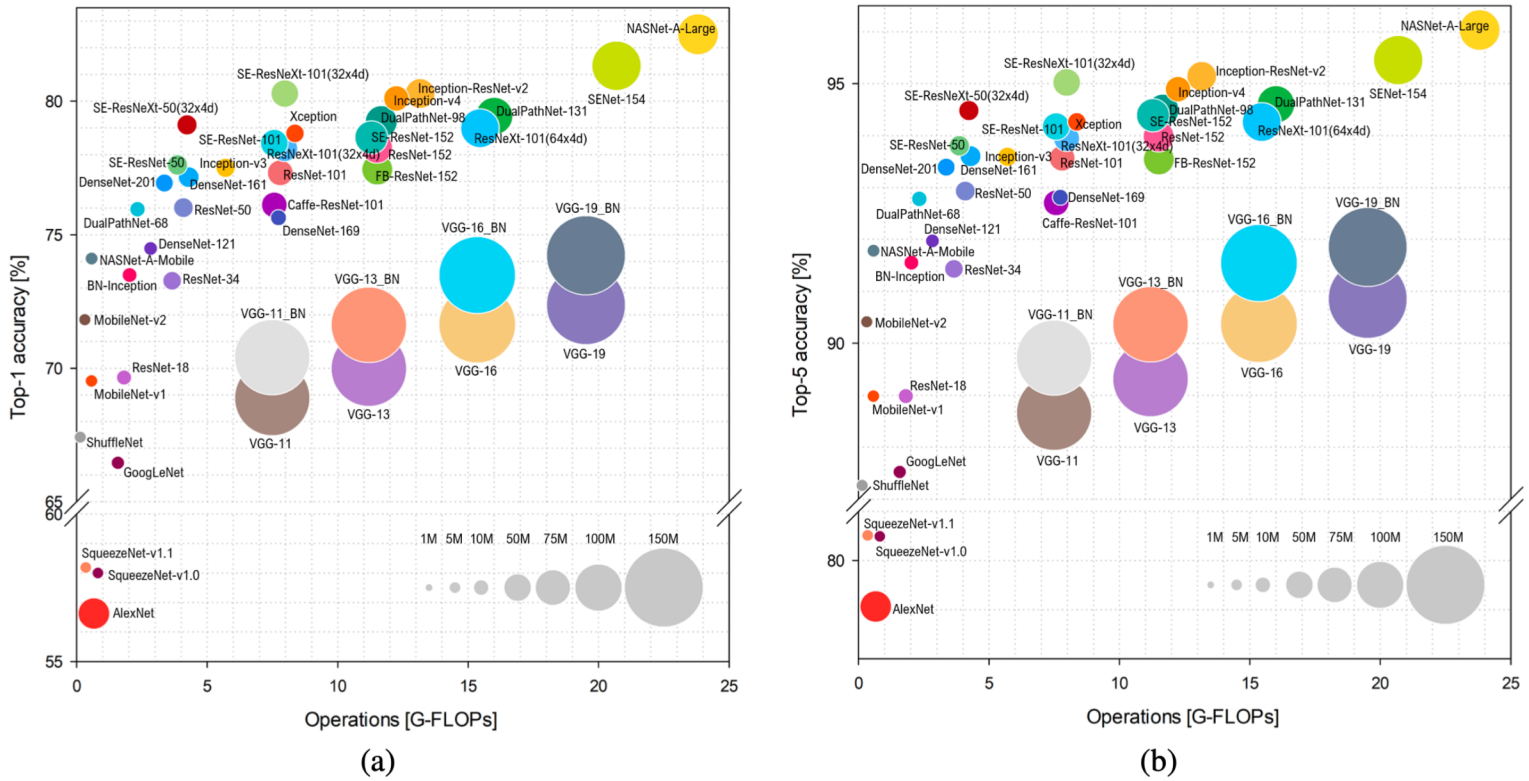


FIGURE 1: Ball chart reporting the Top-1 and Top-5 accuracy *vs.* computational complexity. Top-1 and Top-5 accuracy using only the center crop versus floating-point operations (FLOPs) required for a single forward pass are reported. The size of each ball corresponds to the model complexity. (a) Top-1; (b) Top-5.

Историческая справка

- ResNet: He, Kaiming, et al. "Deep residual learning for image recognition." CVPR, 2016.
- DenseNet: Huang, Gao, et al. "Densely connected convolutional networks." *CVPR*. 2017.