Machine Learning for Modelling: *Supervised Learning*

Simone Bianco

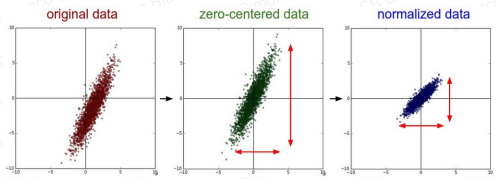
1

Convolutional Neural Networks (CNNs)

2

Data pre-processing

- **Local mean subtraction:** subtract the mean from original data
  - $X = X - \mu$
- **Normalization:** scale original data to a specific range
  - $X = \frac{X-\mu}{\sigma}$
- ...



53

Preventing overfitting

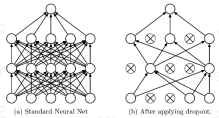
- Overfitting is a problem since CNNs have many parameters
- Forms of regularization approaches
  - Weight decay (Penalizes weight magnitudes)
  - Dropout
  - Data augmentation

54

Dropout

x_{ijk} → Dropout → $y_{ijk} = \frac{1}{1-p}\begin{cases} x_{ijk} & \text{with prob. } p \\ 0 & \text{otherwise} \end{cases}$

- Randomly switches off individual neurons

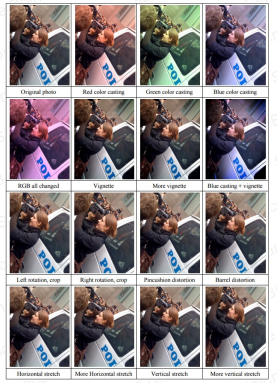


x_{ijk} → Dropout (test) → $y_{ijk} = x_{ijk}$

55

Data augmentation

- Augment the training set by "jittering" samples
- Label preserving image transformations:
  - Horizontal flip
  - Random crop
  - Color casting
  - Geometric distortion

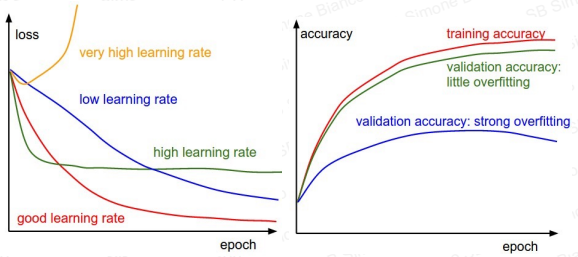


R. Wu, S. Yan, Y. Shan, Q. Dang and G. Sun, **Deep image: Scaling up image recognition**, *arXiv preprint arXiv:1501.02876, 2015.*

56

Diagnose training

- Check the effects of learning rate on the loss trend
- Track the validation/training accuracy curve trend



57

Validation

- Split the data into training, validation, and test subsets
- Use training set to optimize model parameters
- Use validation test to choose the best model
- Use test set only to evaluate performance

58

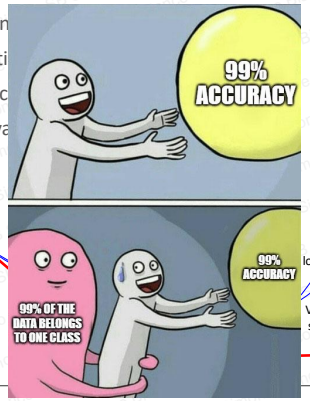
Validation

- Split the data into training
- Use training set to optimi
- Use validation test to cho
- Use test set only to evalua

59

Validation

- Split the data into train
- Use training set to opti
- Use validation test to c
- Use test set only to eva



60

AA 2022/2023

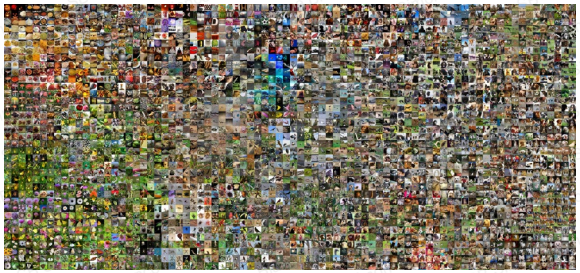
CNN architectures

61

Applications

- 14,197,122 images
- 21,841 synsets

IMAGENET



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.*

62

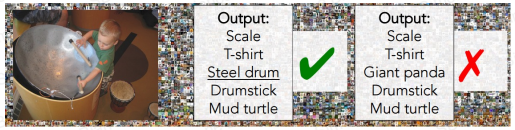
Applications

IMAGENET

- Image classification

Large Scale Visual Recognition Challenge

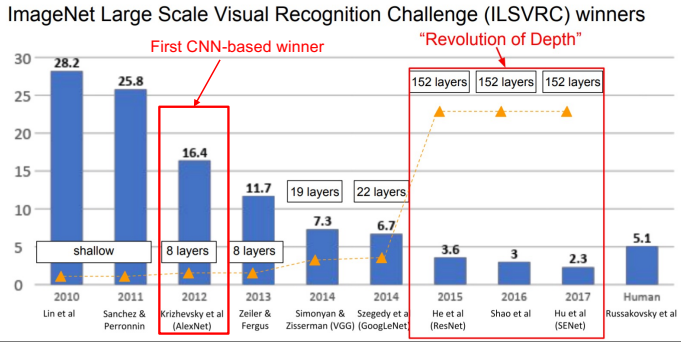
The Image Classification Challenge:
1,000 object classes
1,431,167 images



63

CNN Architectures

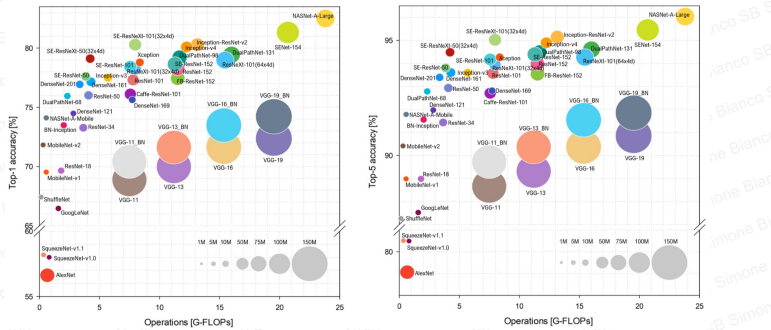
Imagenet classification error



64

CNN Architectures

Imagenet classification error

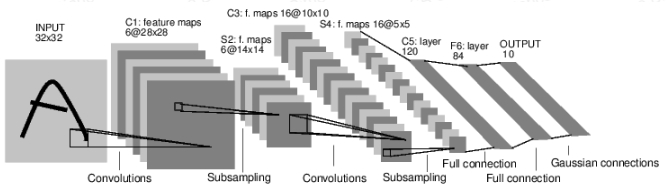


Ball chart reporting the Top-1 and Top-5 accuracy vs. computational complexity. Top-1 and Top-5 accuracy using only the center crop versus floating-point operations (FLOPs) required for a single forward pass are reported. The size of each ball corresponds to the model complexity. (a) Top-1; (b) Top-5.

65

CNN Architectures

LeNet (LeNet-5)

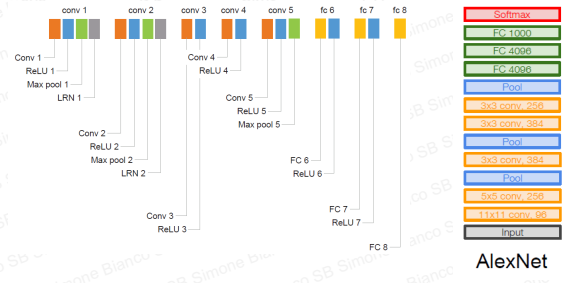


LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278-2324.

66

CNN Architectures

AlexNet

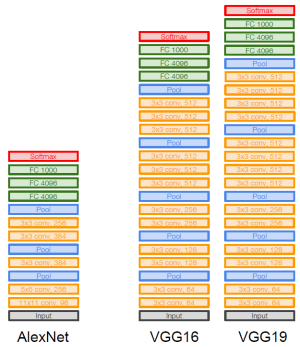


Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097-1105).

67

CNN Architectures

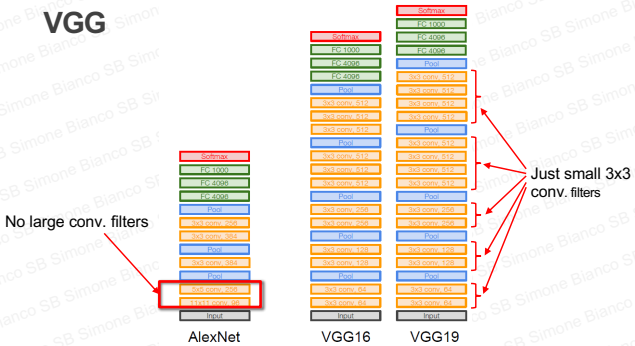
VGG



68

CNN Architectures

VGG



69

CNN Architectures

VGG

Why use smaller filters?

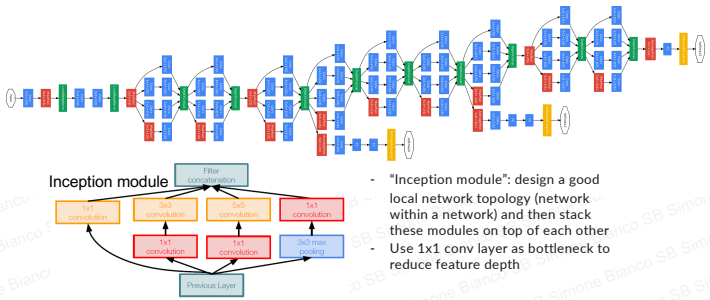
- Stack of 3x3 conv filters has the same receptive field as one 7x7
- But makes the CNN deeper, with more non-linearities
- And fewer parameters

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

70

CNN Architectures

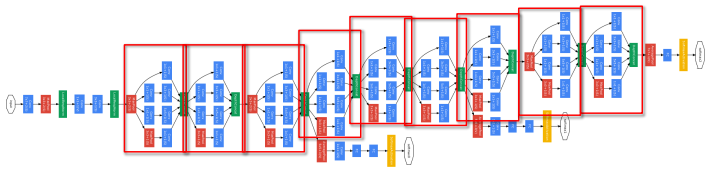
GoogLeNet (Inception)



- "Inception module": design a good local network topology (network within a network) and then stack these modules on top of each other
- Use 1x1 conv layer as bottleneck to reduce feature depth

71

CNN Architectures

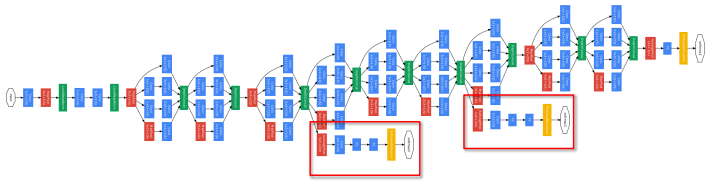


- 9 inception modules stacked one on top of each other

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015, June). Going deeper with convolutions. CVPR.

72

CNN Architectures

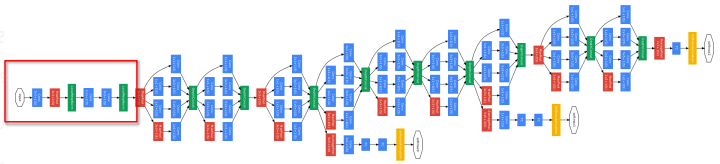


- 9 inception modules stacked one on top of each other
- 2 auxiliary classifiers to inject additional gradients at lower layers

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015, June). Going deeper with convolutions. CVPR.

73

CNN Architectures

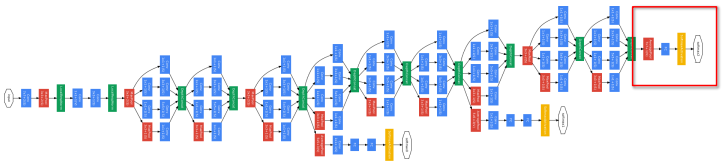


- 9 inception modules stacked one on top of each other
- 2 auxiliary classifiers to inject additional gradients at lower layers
- 1 stem network

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015, June). Going deeper with convolutions. CVPR.

74

CNN Architectures



- 9 inception modules stacked one on top of each other
- 2 auxiliary classifiers to inject additional gradients at lower layers
- 1 stem network
- 1 classifier output with expensive FC layers removed

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015, June). Going deeper with convolutions. CVPR.

75

CNN Architectures

ResNet

- So far, the deeper the better
- Can we continue on adding layer and go deeper?

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

76

CNN Architectures

ResNet

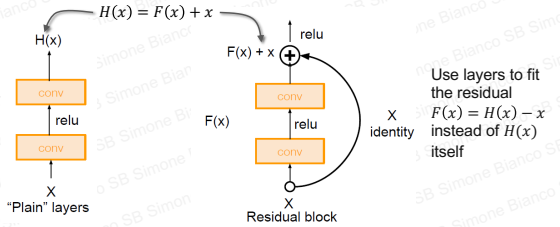
- So far, the deeper the better
- Can we continue on adding layer and go deeper?
- Unfortunately no
- Hypothesis: the problem is an optimization problem, deeper models are harder to optimize (e.g. vanishing gradients)

77

CNN Architectures

ResNet

Solution: Use network layers to fit a residual mapping instead of directly trying to fit a desired underlying mapping

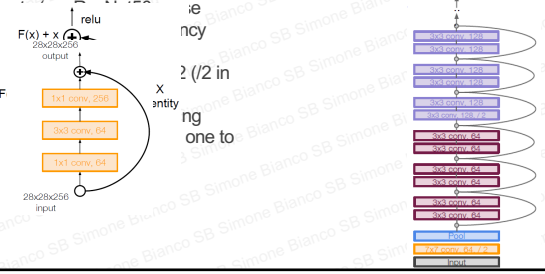


78

CNN Architectures

ResNet

- Stack residual blocks
- Every residual block has two 3x3 conv layers
- For deepe… …ie bottleneck …ncy
- Periodicall… …2 (/2 in downsamp… each dime… F … X …entity …ng
- Additional …
- No FC lay… …one to map to the …



79

How deep is enough?

GoogLeNet (2014)
VGG-VD-16 (2014)
VGG-M (2013)
AlexNet (2012)

ResNet 50 (2015)
ResNet 152 (2015)

16 convolutional layers
50 convolutional layers
152 convolutional layers

Krizhevsky, I. Sutskever, and G. E. Hinton. *ImageNet classification with deep convolutional neural networks*. In Proc. NIPS, 2012.

C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. *Going deeper with convolutions*. In Proc. CVPR, 2015.

K. Simonyan and A. Zisserman. *Very deep convolutional networks for large-scale image recognition*. In Proc. ICLR, 2015.

K. He, X. Zhang, S. Ren, and J. Sun. *Deep residual learning for image recognition*. In Proc. CVPR, 2016.

80

How deep is enough?

GoogLeNet (2014)
VGG-VD-16 (2014)
VGG-M (2013)
AlexNet (2012)

16 convolutional layers
50 convolutional layers
152 convolutional layers

81

Transfer learning

82

Transfer learning

- All the best performing architectures have millions of parameters to learn
- In order to tune them, very large databases are needed (e.g., rule of 10: 10x samples wrt number of parameters to train)

- What if we do not have such amount of data? Can we still use deep learning?

83

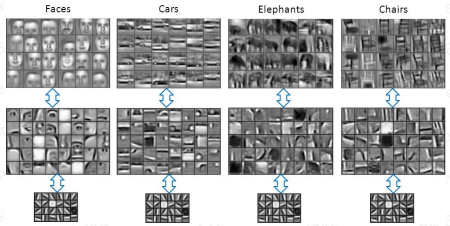
Transfer learning

- In practice, very few people train an entire CNN from scratch, because it is relatively rare to have a dataset of sufficient size.
- It is common to pretrain a CNN on a very large dataset (e.g., ImageNet which contains 1.2 million images with 1000 categories)
- and then use the CNN either for fine-tuning or as a fixed feature extractor for the task of interest.

84

Transfer learning

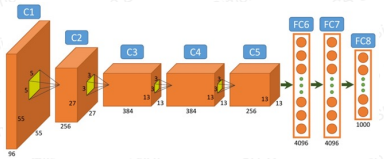
- Why does it work?
- Motivated by the observation that the earlier features of a CNN contain more generic features (e.g., edge detectors or color blob detectors) that should be useful to many tasks
- Later layers of the CNN becomes progressively more specific to the details of the classes contained in the original dataset.



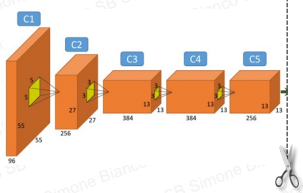
85

CNN fine-tuning

- Task 1 (e.g. ImageNet, 1000 categories)



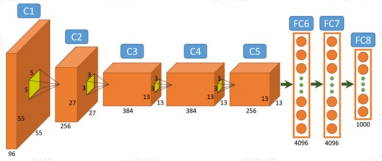
- Task 2 (10 categories)



86

CNN fine-tuning

- Task 1 (e.g. ImageNet, 1000 categories)

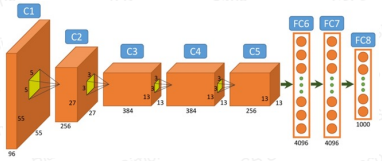


- Task 2 (10 categories)

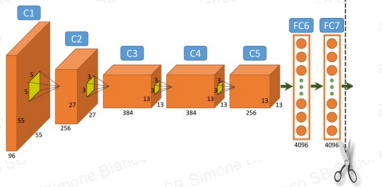
87

CNN as feature extractor

- Task 1 (e.g. ImageNet, 1000 categories)



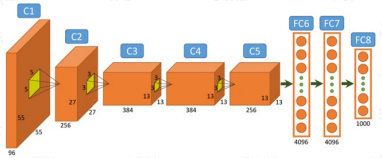
- Task 2 (10 categories)



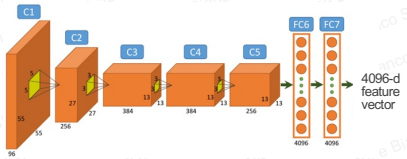
88

CNN as feature extractor

- Task 1 (e.g. ImageNet, 1000 categories)



- Task 2 (10 categories)



89

How to decide what type of TL you should perform?

- *New dataset is **small** and **similar** to original dataset:*

→ Train a linear classifier on CNN features from higher layers

- *New dataset is **large** and **similar** to original dataset*

→ Fine-tune the CNN

- *New dataset is **small** but **very different** from original dataset*

→ Train a linear classifier on CNN features from lower layers

- *New dataset is **large** and **very different** from original dataset*

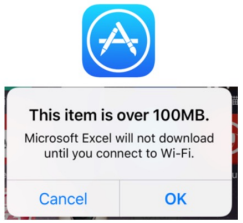
→ Train CNN from scratch or fine-tune it

90

Model Compression

The problem: If running DNN on Mobile…

APP developers suffer from the model size!



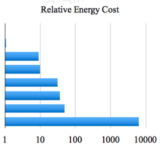
91

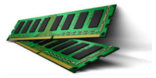
Model Compression

The problem: If running DNN on Mobile…

HW engineers suffer from the model size! (embedded systems, limited resources)

Operation	Energy [pJ]	Relative Cost
32 bit int ADD	0.1	1
32 bit float ADD	0.9	9
32 bit Register File	1	10
32 bit int MULT	3.1	31
32 bit float MULT	3.7	37
32 bit SRAM Cache	5	50
32 bit DRAM Memory	**640**	**6400**





92

Model Compression

The problem: If running DNN on the Cloud…

Network Delay

Power Budget

User Privacy

Intelligent but Inefficient

93

Model Compression

Smaller Size
Compress Mobile App Size by 35x-50x

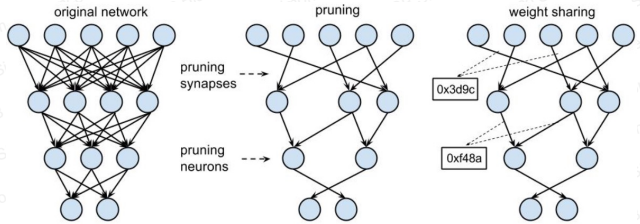
Accuracy
no loss of accuracy improved accuracy

Speedup
make inference faster

- AlexNet: 35×, 240MB => 6.9MB
- VGG16: 49× 552MB => 11.3MB
- Both with no loss of accuracy on ImageNet
- Weights fits on-chip SRAM, taking 120x less energy than DRAM

94

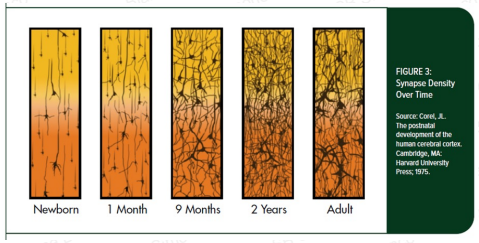
Model Compression



95

Model Compression

- **50 trillions** of synapses are generated in the human brain during the first few months of birth.
- **1 year old**, peaked at **1000 trillions**
- Pruning begins to occur.
- **10 years old**, a child has nearly **500 trillions** synapses
- This 'pruning' mechanism removes redundant connections in the brain.



96