riminder

DEEP LEARNING IN COMPUTER VISION

Slim FRIKHA
Lead Computer Vision AI Researcher - RIMINDER

DEEP LEARNING PRACTICAL COURSE
ECOLE POLYTECHNIQUE, 12/04/2018

Program & Course Logistics

- **Course 1 :** (05-04-18)
  - Introduction to Deep Learning - Mouhidine SEIV (Riminder)
- **Course 2 : (12-04-18)**
  - **Deep Learning in Computer Vision - Slim FRIKHA (Riminder)**
- **Course 3 :** (19-04-18)
  - Deep Learning in NLP - Paul COURSAUX (Riminder)
- **Course 4 :** (26-04-18)
  - Efficient Methods and Compression for Deep Learning - INVITED GUEST
- **Course 5:** (03-05-18)
  - Introduction to Deep Learning Frameworks - INVITED GUEST
- **Course 6:** (10-05-18)
  - Deployment in Production and Parallel Computing - INVITED GUEST

Talk outline

I. *Computer vision overview*

II. *Deep learning for image classification*

III. *Deep learning for object detection and semantic segmentation*

IV. *Human versus Machine*

Why computer vision is important

Google images search

Microsoft Kinect

Google Street View

Credit Card scanner

Self-driving cars

OCR in ATM check deposits

Smartphone face unlock

Number plate recognition

Vision Biometrics

3-D Printing

Tasks overview

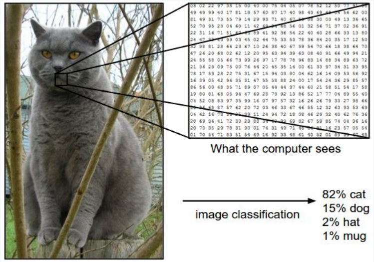
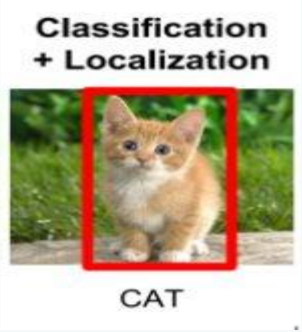
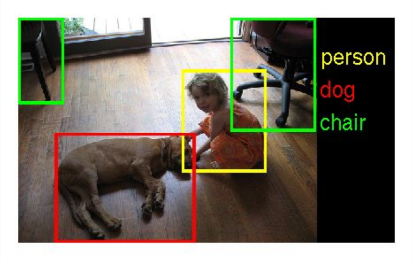


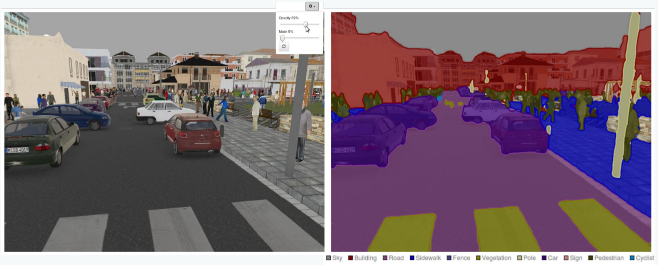
Image classification



Object localization



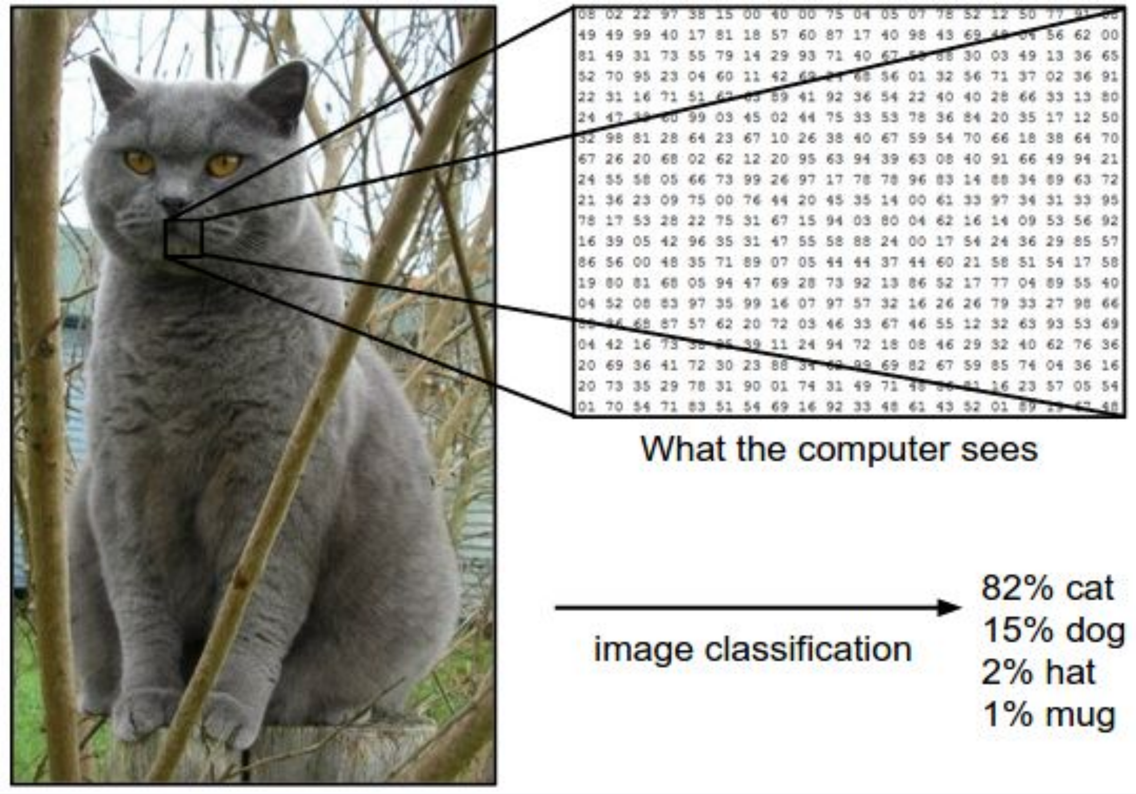
Object detection



Semantic segmentation

Image captioning

Image classification problem



Why convolutions?

Classical machine learning: input format, features engineering

Dense layers: too many parameters

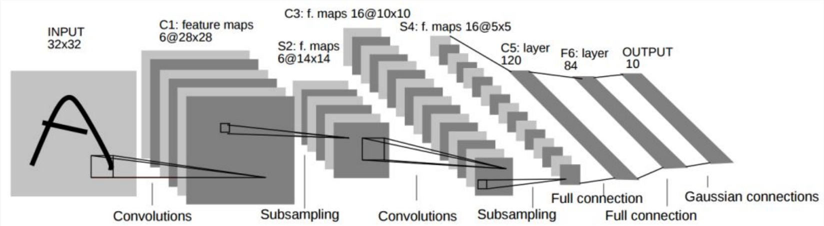
Recurrent neural networks: 1D sequences, loss of spatial information

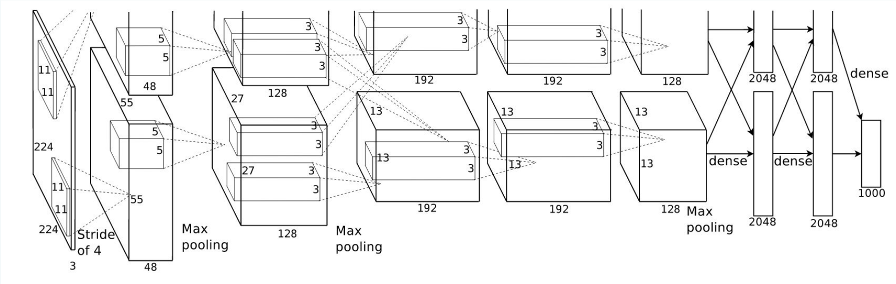
Convolutional Neural Networks

Neurocognitron [Fukushima 1980]

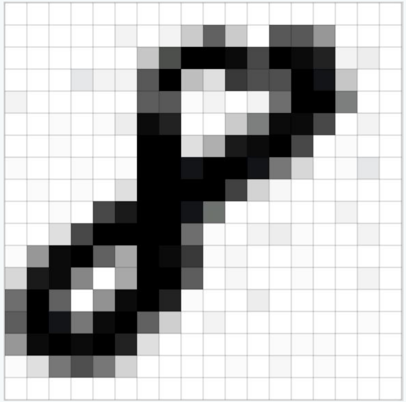
LeNet-5 [Lecun 1998]

Alexnet [Krizhevsky, Sutskever, Hinton 2012]

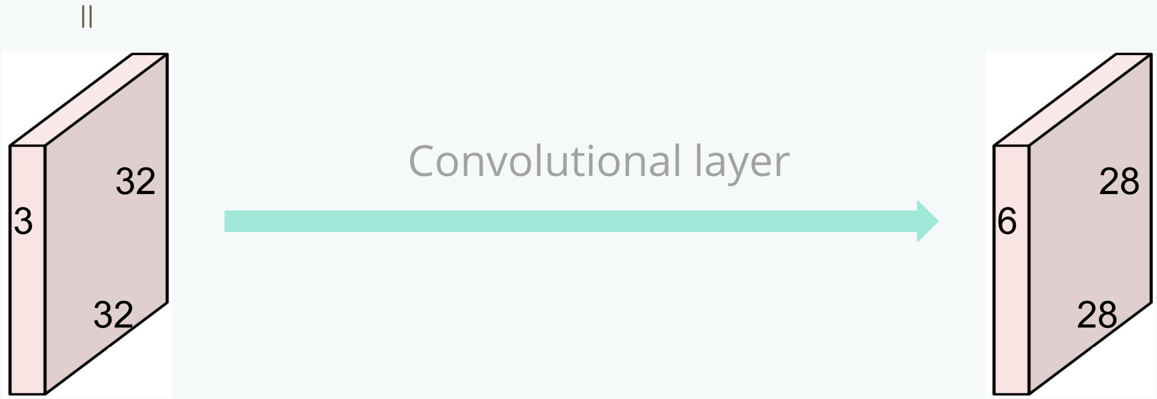




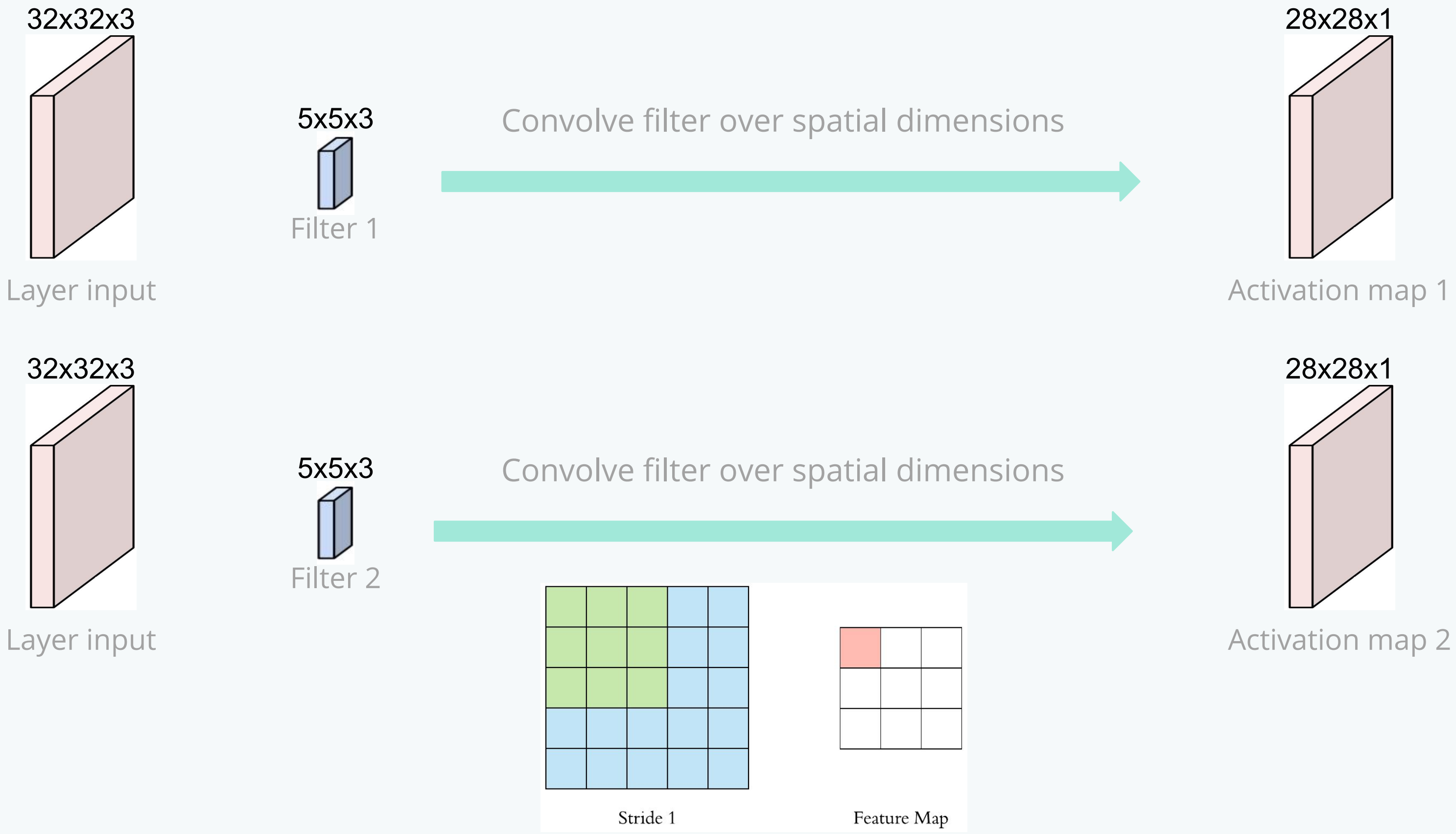
What is a convolution layer?



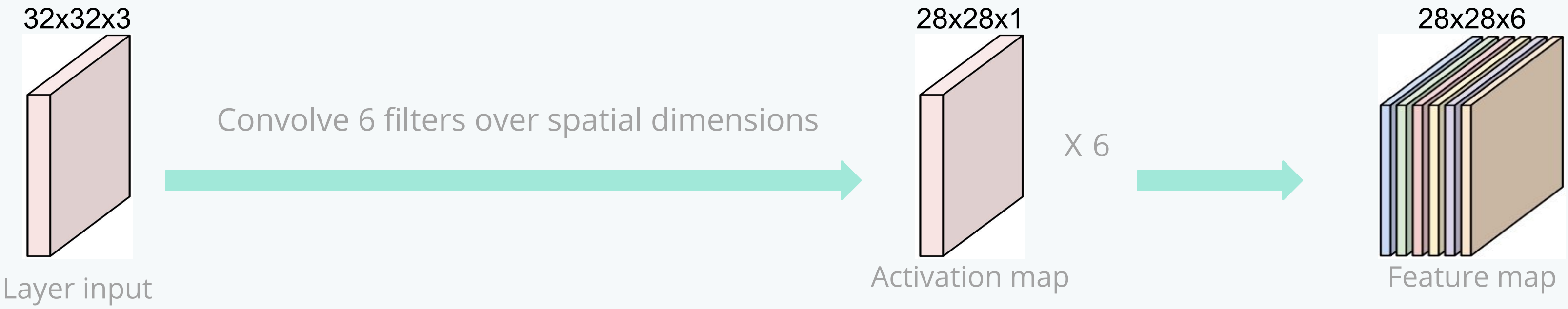
An image is just a 3D array of numbers

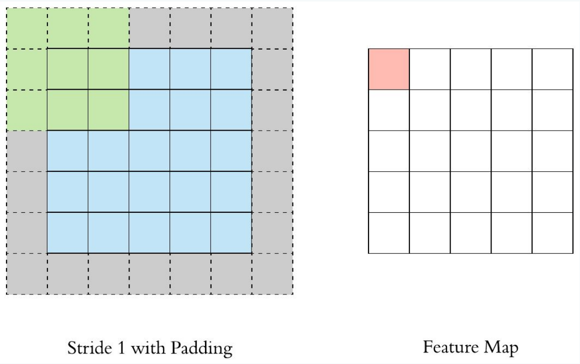


Convolution operation



Convolution operation

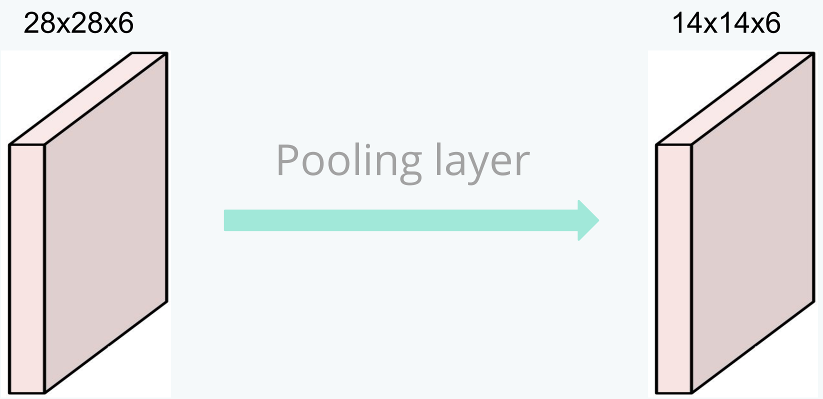


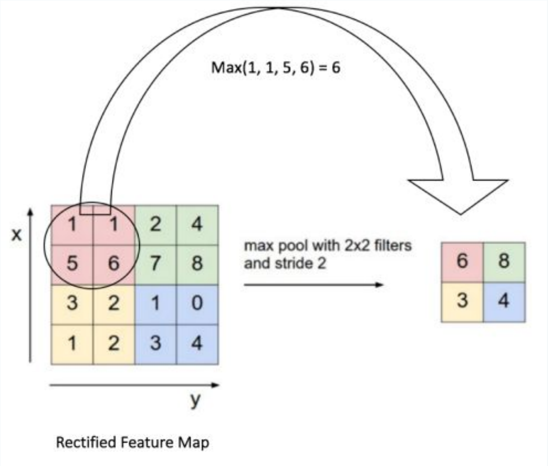


Pooling operation

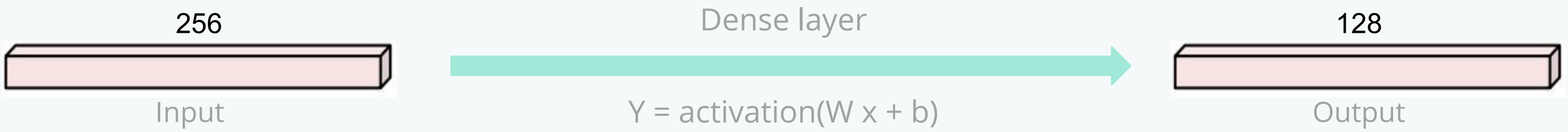
Pooling = downscaling spatial dimension

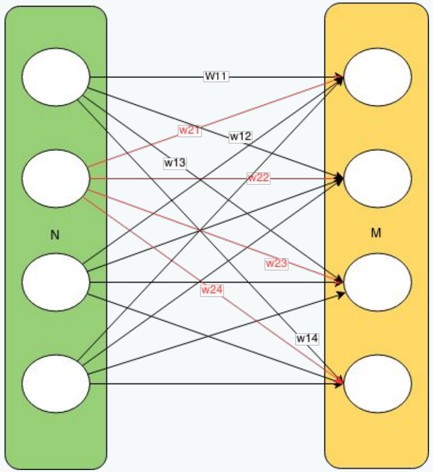
Different types: Max, Average, Sum etc.



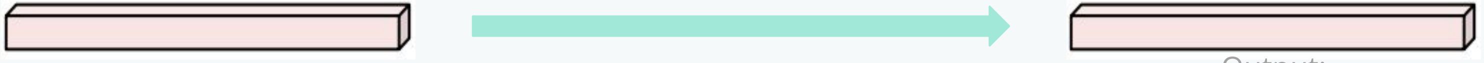


Fully connected layer





Softmax activation



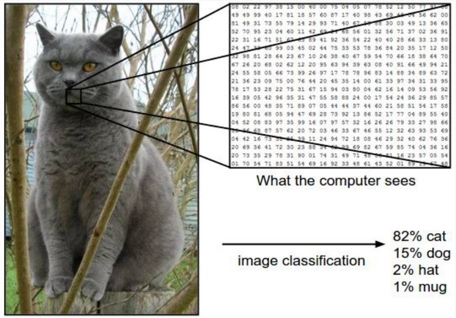
$$\sigma : \mathbb{R}^K \to (0,1)^K$$

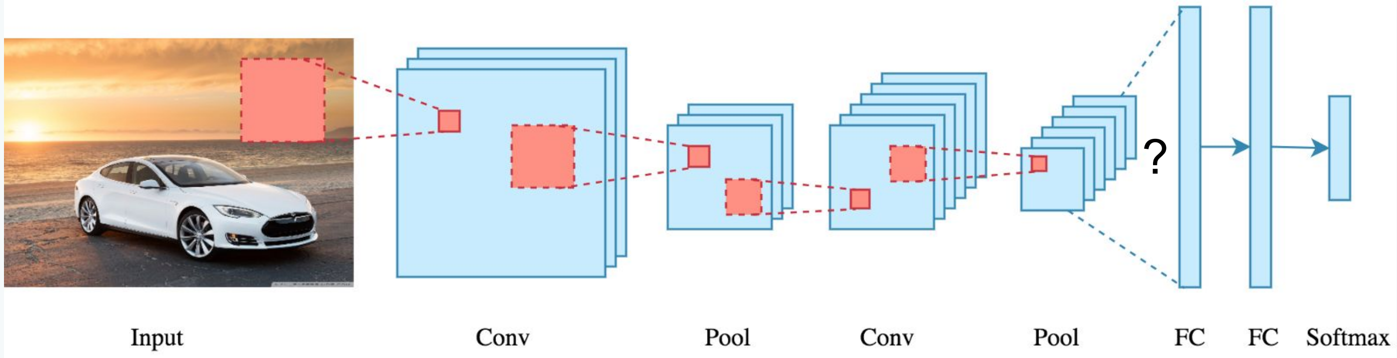
$$\sigma(\mathbf{z})_j = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}} \quad \text{for } j = 1, \ldots, K.$$

	Scoring Function	Unnormalized Probabilities	Normalized Probabilities
Dog	-3.44	0.0321	0.0006
Cat	1.16	3.1899	0.0596
Boat	-0.81	0.4449	0.0083
Airplane	3.91	49.8990	0.9315

CNN architecture

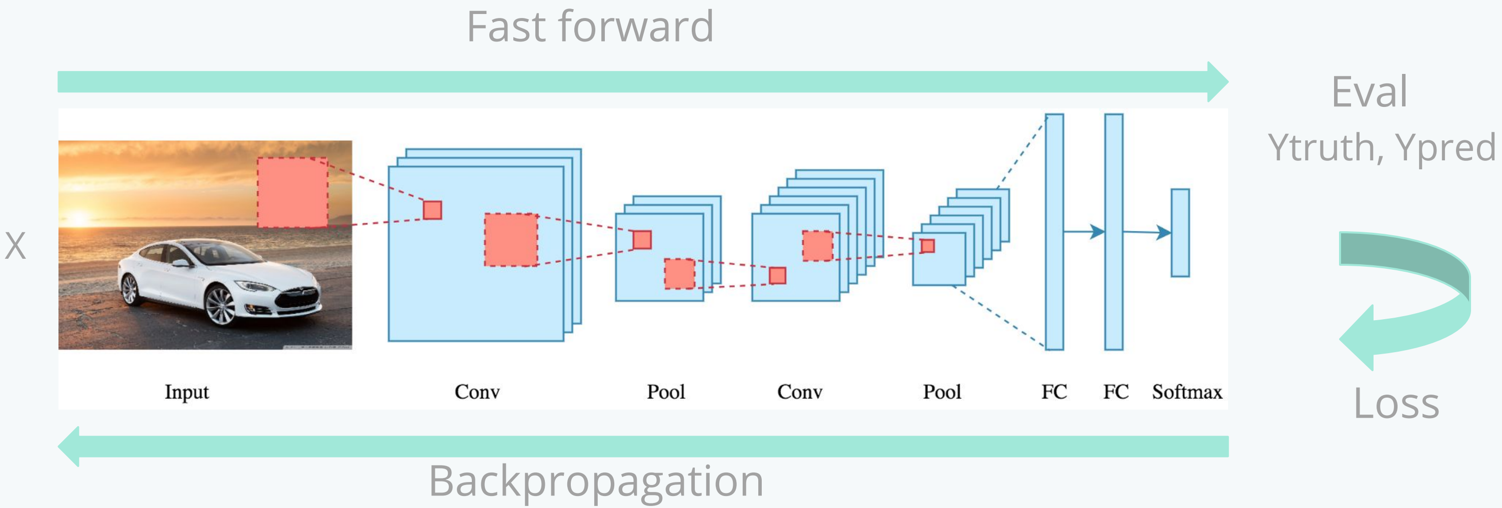
CNN
=
Convolutions + pooling + fully connected





32x32x3 | 28x28x6 | 14x14x6 | 10x10x16 | 5x5x16 | 64 | 4 | 4

Training CNNs



Regularization: data augmentation

Horizontal / vertical flip

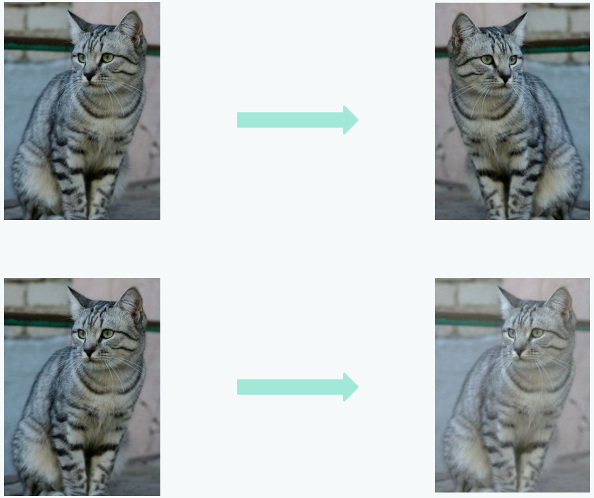
Color jitter

Random crops and scales

Translation

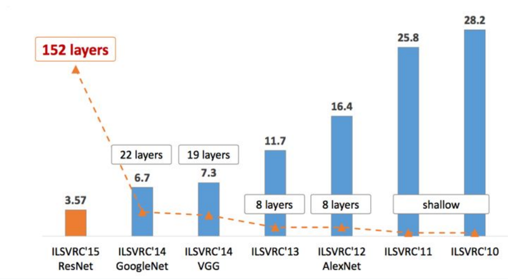
Rotation

Stretching ...

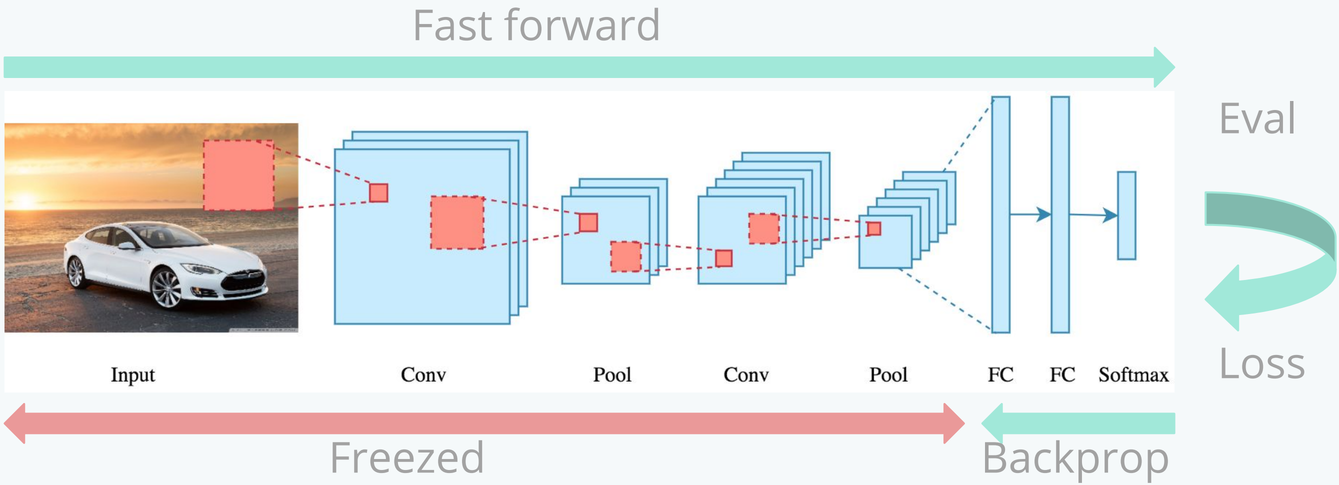


Generalization: transfer learning

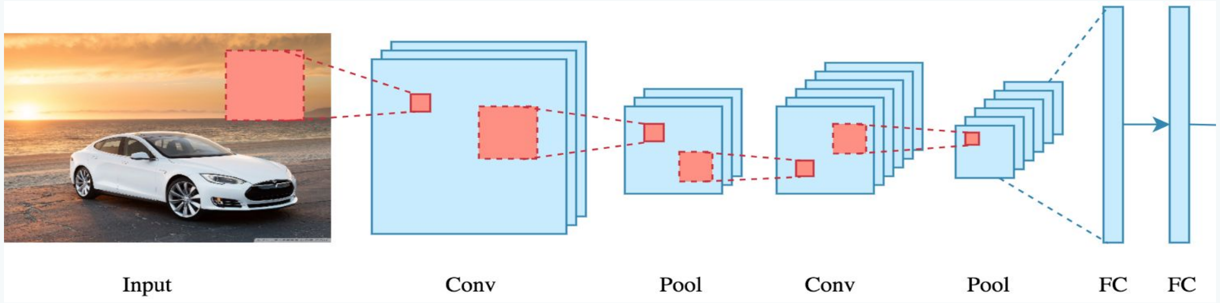
AlexNet (2012)
ZF Net (2013)
VGG Net (2014)
GoogLeNet (2015)
Microsoft ResNet (2015)...



top 5 test error rate



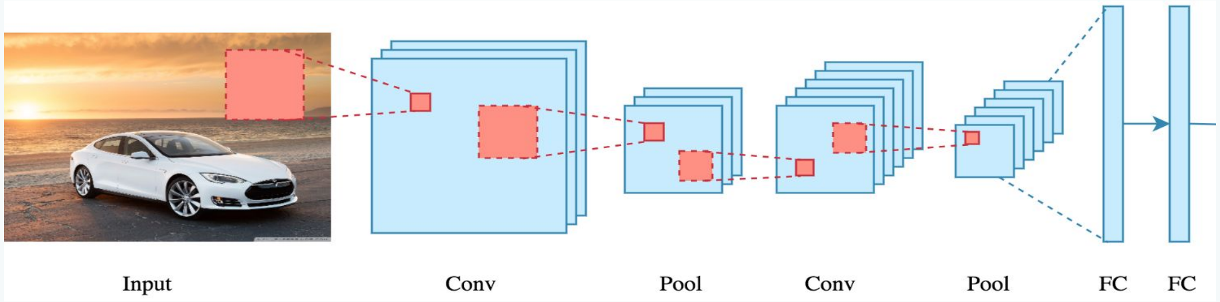
Object localization



Class

Bounding box

Objects detection

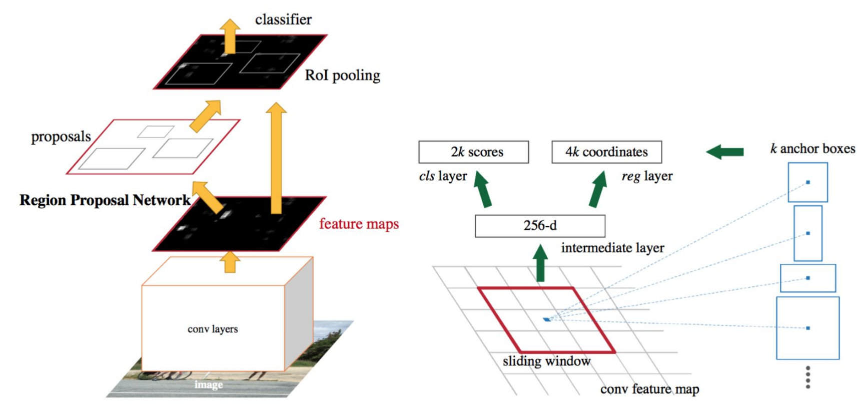


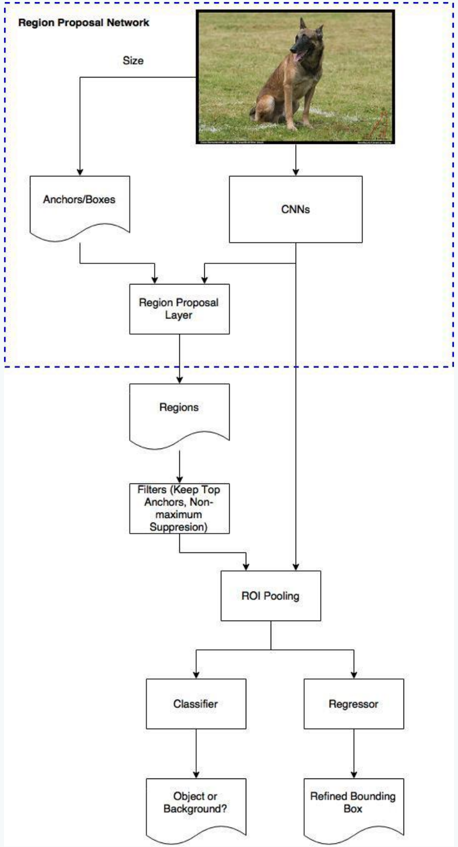
Class

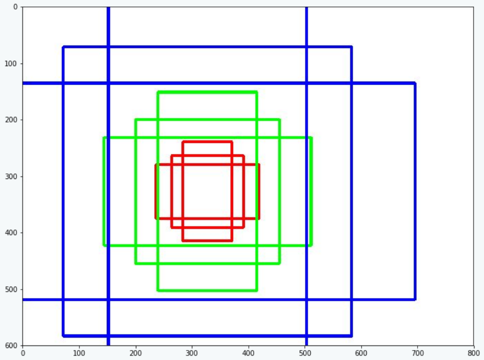
x N

Bounding box

Objects detection: Faster R-CNN





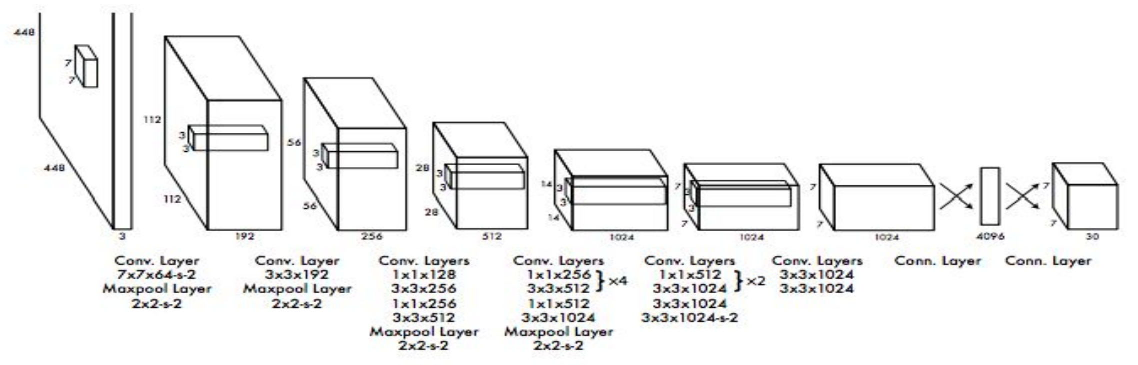


Anchor boxes example for k=9

Objects detection: YOLO

Input

448 x 448 x 3



Output

$S \times S \times (B * 5 + C)$

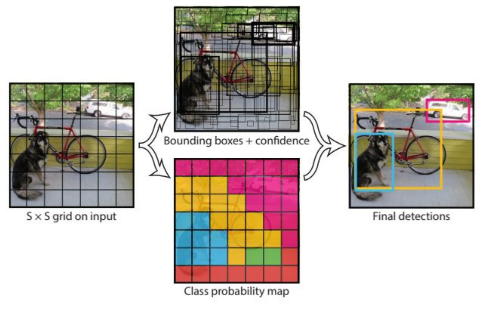
YOLO architecture example: S=7, B=2, C=30

$S \times S$ grid

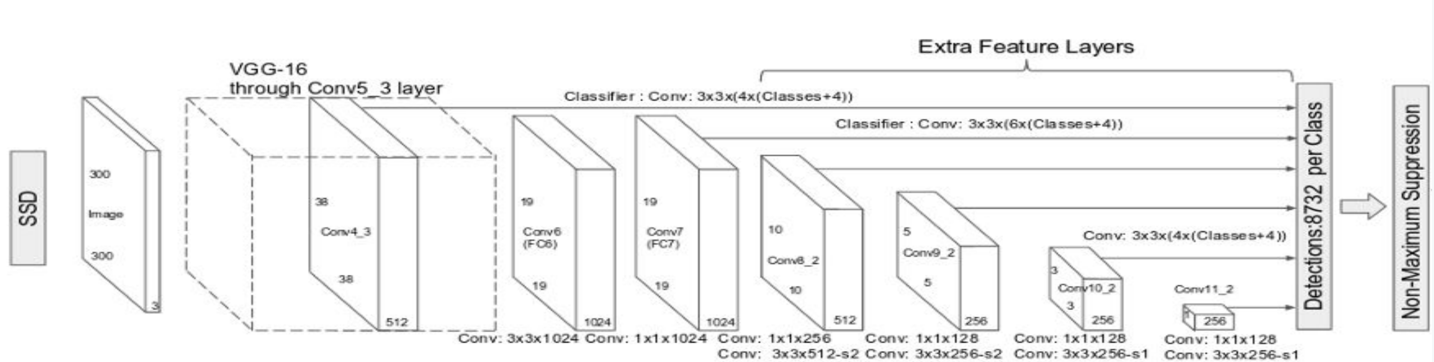
Each grid cell predicts B bounding boxes and confidence scores for those boxes

Each grid cell predicts **one** set C conditional class probabilities

Faster than Faster R-CNN but not as accurate as



Objects detection: SSD



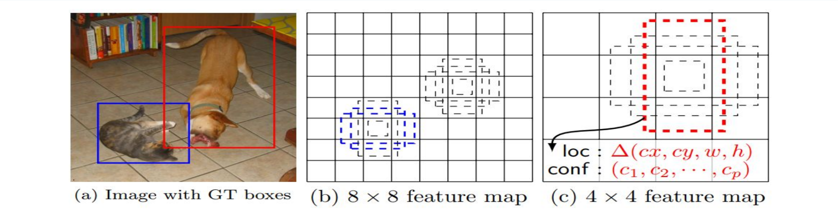
SSD architecture example

Combination of multiple ideas:

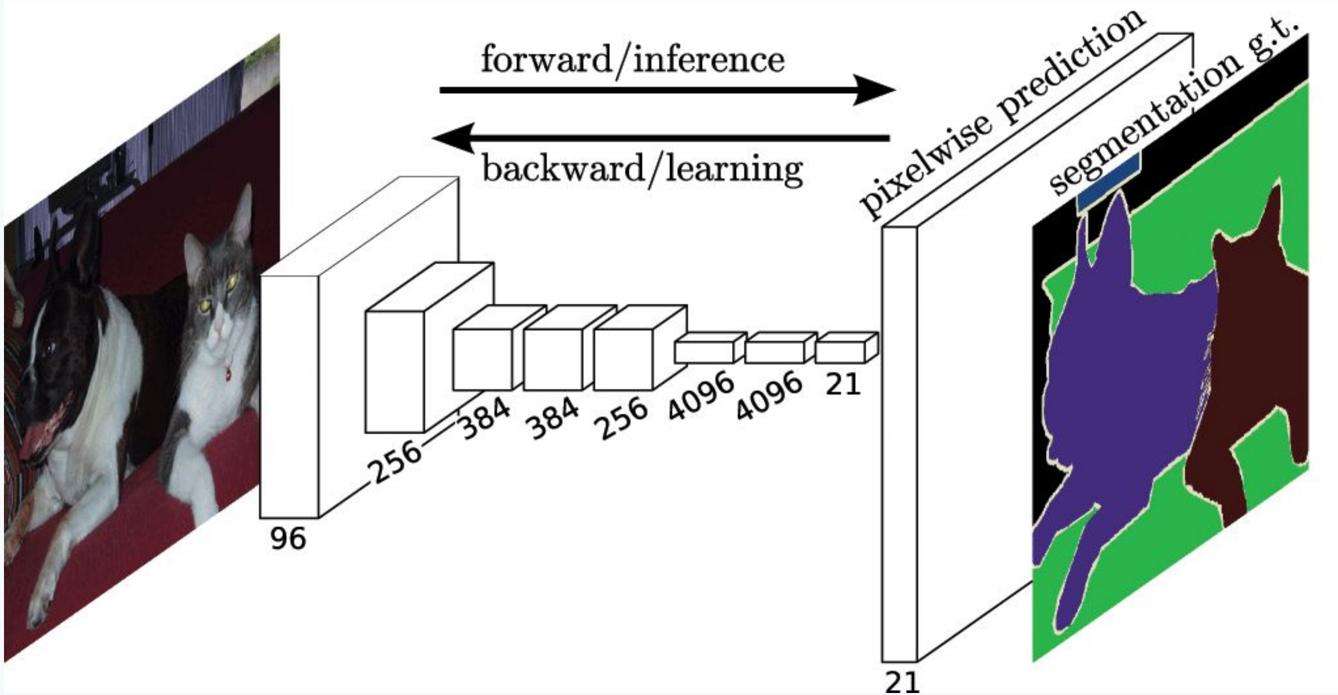
- anchor boxes
- single NN like YOLO
- multi scale support

Faster than Faster R-CNN but not as accurate

Slower than YOLO but more accurate



Semantic segmentation

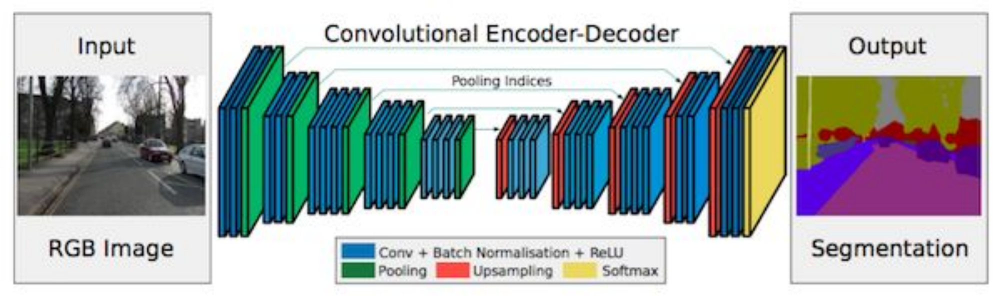


Semantic segmentation

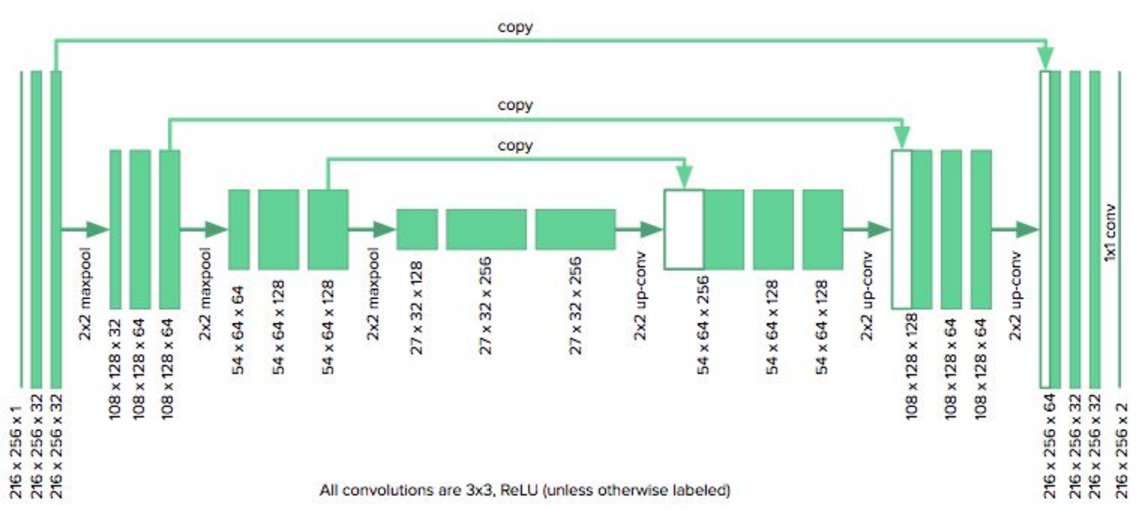
Multiple architecture and ideas too:

- FCN
- SegNet
- U-Net
- Dilated Convolutions
- DeepLab (v1 & v2)
- PSPNet
- DeepLab v3...

Semantic segmentation: encoder decoder

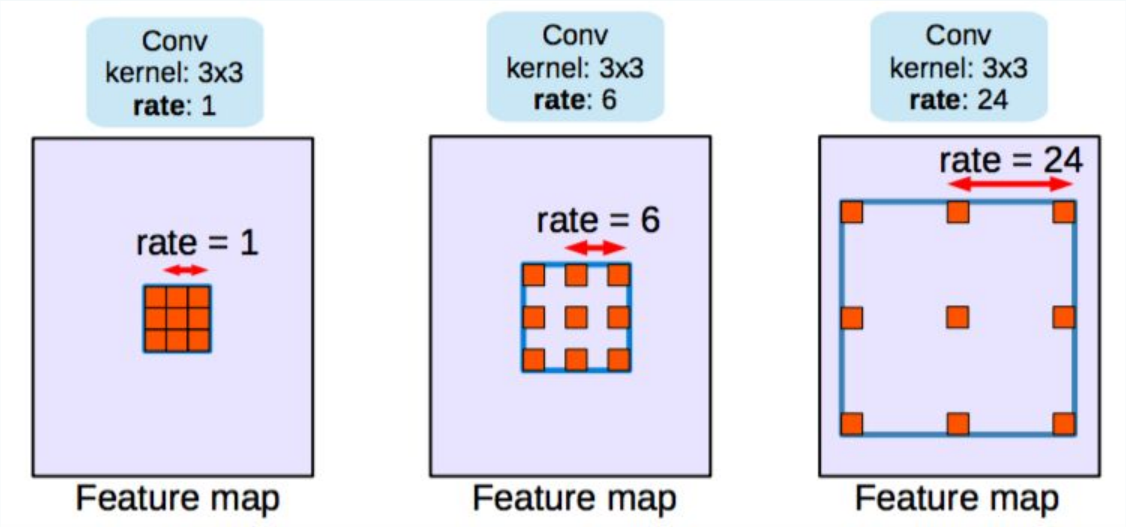


SegNet architecture

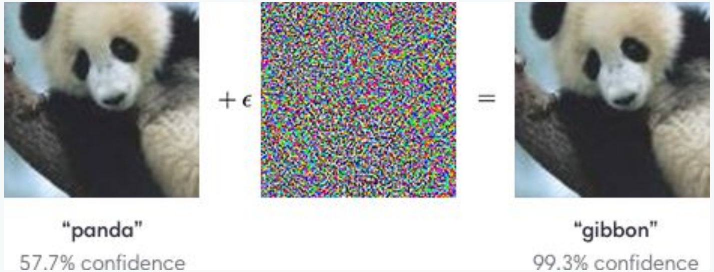


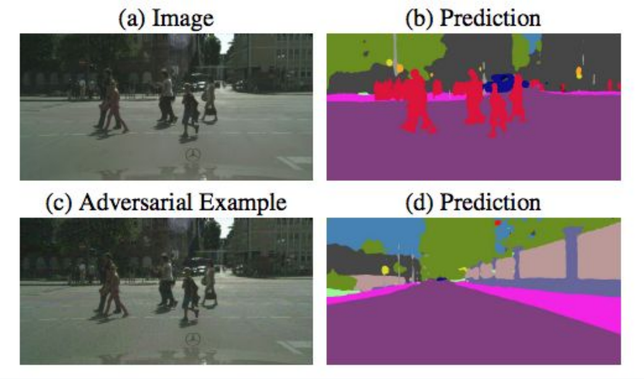
U-Net architecture

Semantic segmentation: atrous convolutions



Adversarial Attacks

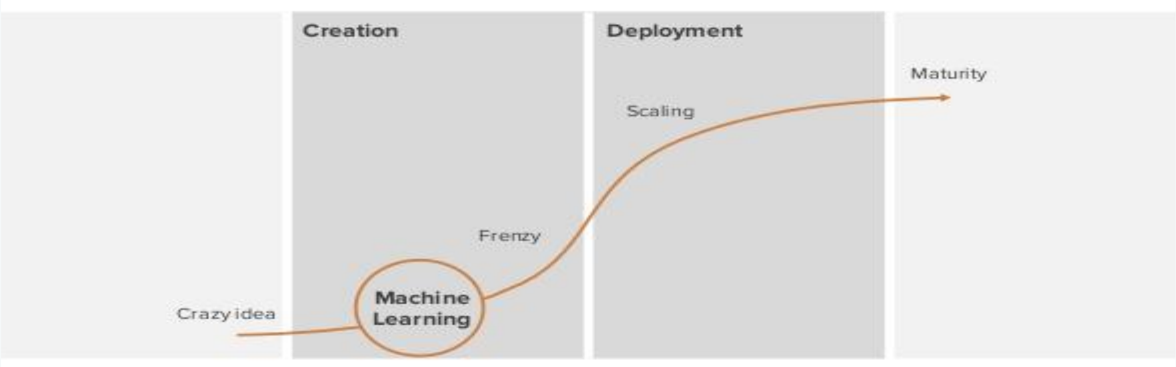




Conclusion

NNs can be better than human for specific simple tasks (classification)

Machine learning is only “at the beginning of the S-Curve”



Machine learning S-Curve

One cool thing

Semantic segmentation to improve FIFA 18 graphics

POINT OF CONTACT

Slim Frikha
Lead Computer Vision AI Researcher

slim.frikha@riminder.net

Sources

- https://medium.com/@alonbonder/ces-2018-computer-vision-takes-center-stage-9abca8a2546d
- http://cs231n.github.io/classification/
- http://cv-tricks.com/object-detection/faster-r-cnn-yolo-ssd/
- https://blog.goodaudience.com/using-convolutional-neural-networks-for-image-segmentation-a-quick-intro-75bd68779225
- https://towardsdatascience.com/image-captioning-in-deep-learning-9cd23fb4d8d2
- http://cs231n.stanford.edu/syllabus.html
- https://medium.com/@ageitgey/machine-learning-is-fun-part-3-deep-learning-and-convolutional-neural-networks-f40359318721
- https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/
- https://towardsdatascience.com/applied-deep-learning-part-4-convolutional-neural-networks-584bc134c1e2
- https://medium.com/@smallfishbigsea/faster-r-cnn-explained-864d4fb7e3f8
- https://lilianweng.github.io/lil-log/2017/12/31/object-recognition-for-dummies-part-3.html#faster-r-cnn
- https://medium.com/diaryofawannapreneur/yolo-you-only-look-once-for-object-detection-explained-6f80ea7aaa1e
- https://medium.com/@ManishChablani/ssd-single-shot-multibox-detector-explained-38533c27f75f
- https://towardsdatascience.com/understanding-ssd-multibox-real-time-object-detection-in-deep-learning-495ef744fab
- https://blog.openai.com/adversarial-example-research/
- https://www.semanticscholar.org/paper/Adversarial-Examples-that-Fool-Detectors-Lu-Sibai/dfa14959ae31c6c95ae508dd847dc7d67f04fad9
- https://futureoflife.org/2017/05/01/machine-learning-security-iclr-2017/
- http://blog.enabled.com.au/artificial-general-intelligence/
- https://towardsdatascience.com/using-deep-learning-to-improve-fifa-18-graphics-529ec44ea37e