

BDMI—课程编号—01510243



SaturnLab



清华大学
Tsinghua
iCenter

卷积网络/循环网络

Convolution Networks/Recurrent Networks

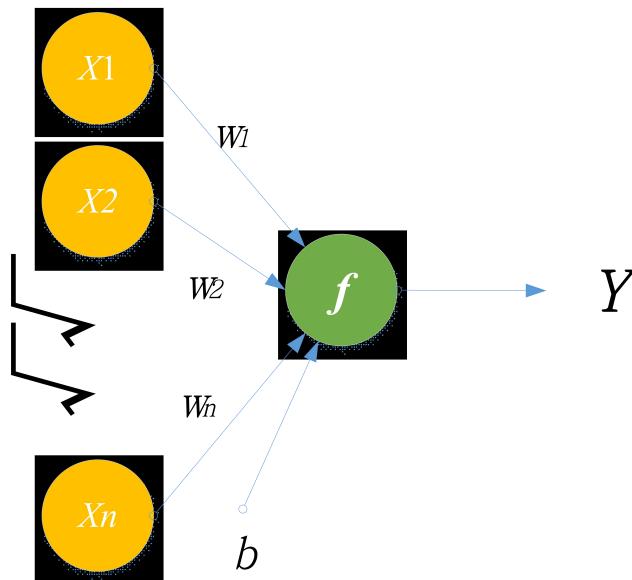
Shiyu Huang

huangsy1314@163.com

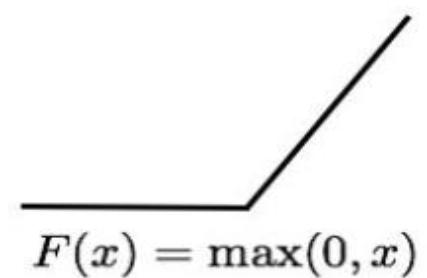
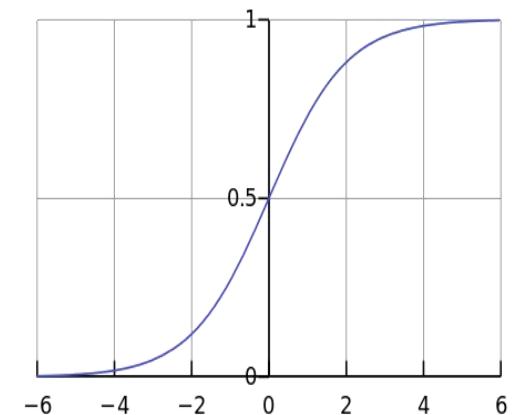
<https://huangshiyu13.github.io>

人工神经元

- 单个人工神经元 (Artificial Neuron)：
 - 一组输入的线性加权叠加
 - 经过一个非线性变换进行输出

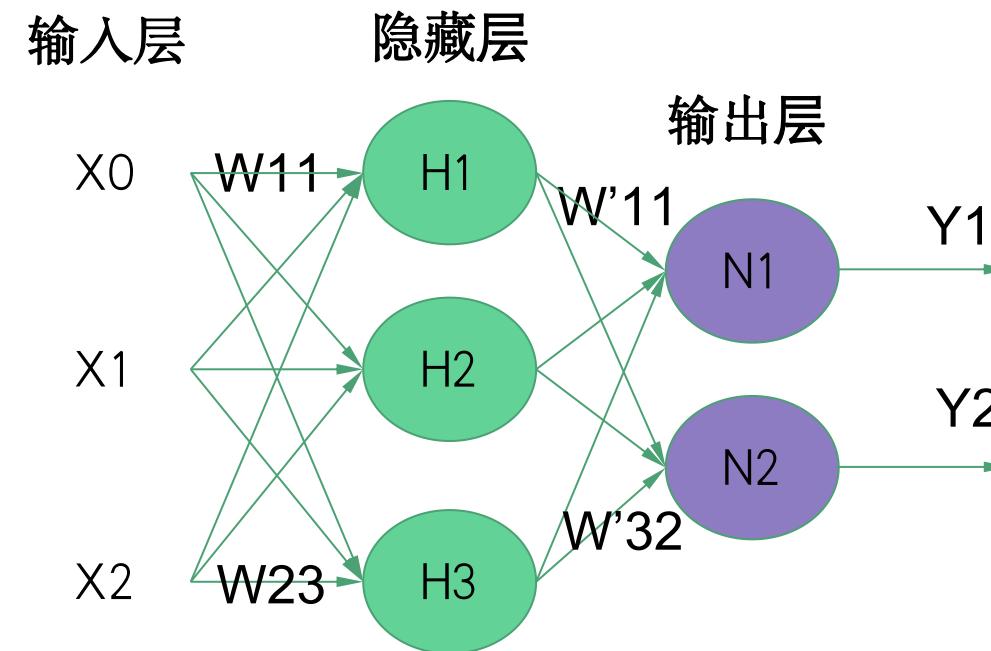


$$y = f \left(\sum_{i=1}^N w_i X_i + b \right)$$



多层人工神经网络 (Multilayer neural networks)

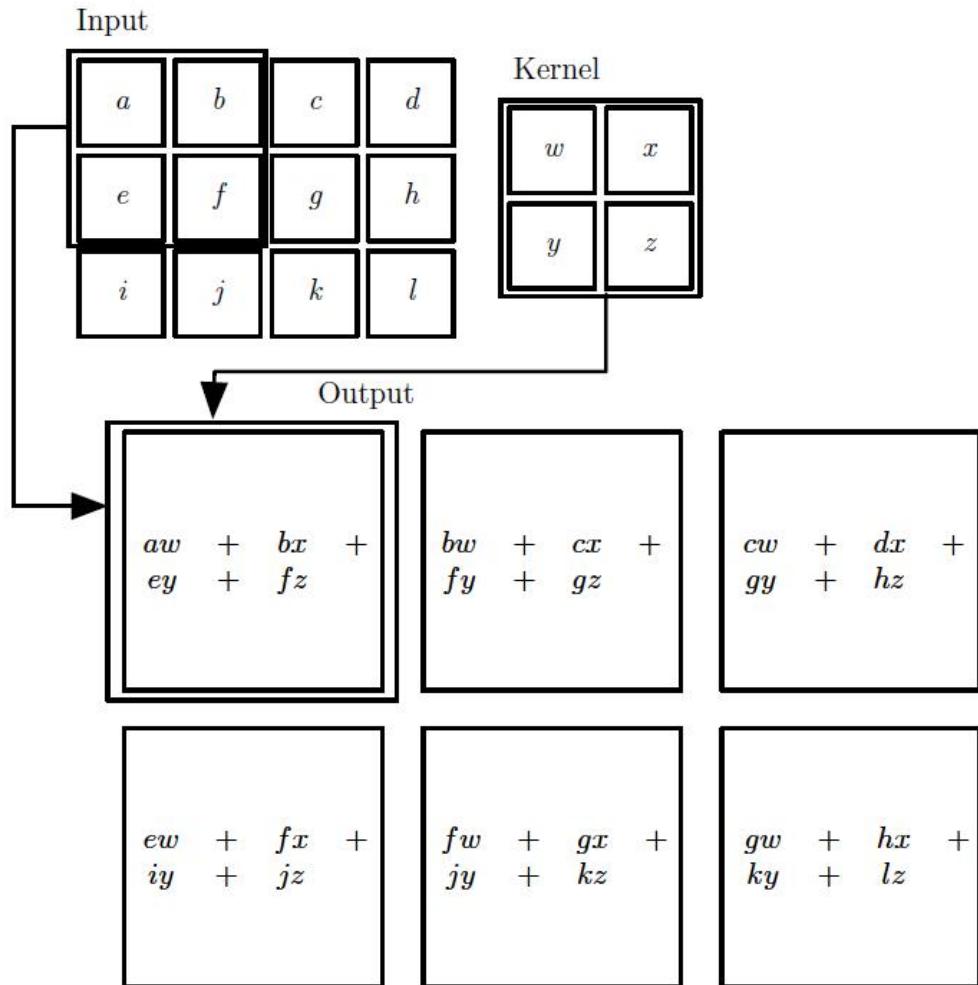
- 单个神经元可以进行线性分类，而多个神经元的组合就可以完成复杂的分类工作。
- 多层神经网络，又称为深度神经网络，在实际中表现出更好的性能。



卷积网络

Convolution Networks

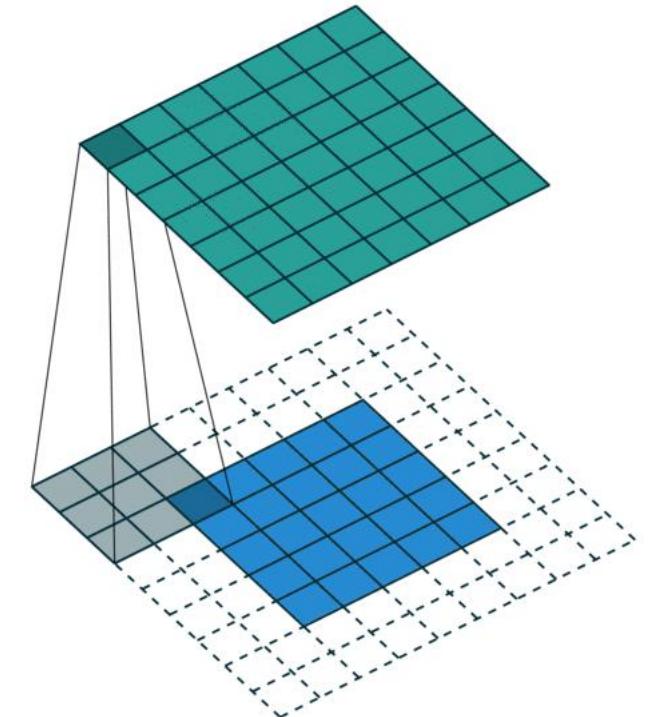
卷积运算 (Convolution)



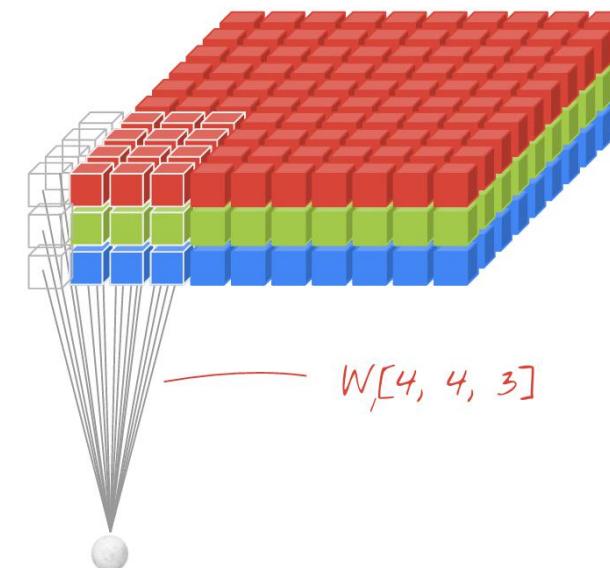
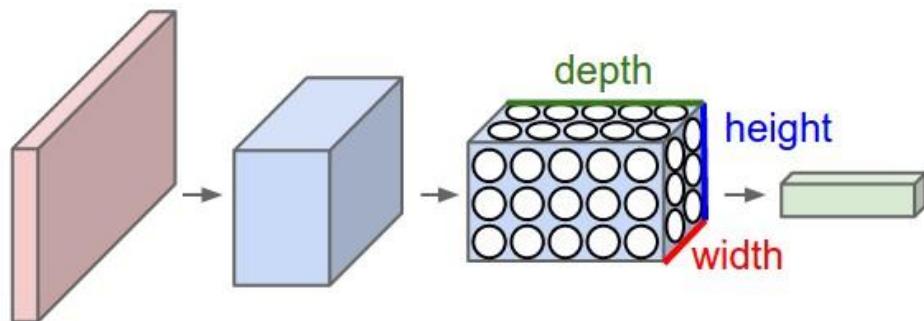
- 卷积是一种张量运算
- 输入是多维数组的数据
- 卷积核是一个多维数组，参数由学习算法得到的
- 卷积核数目一般选32、64等
- 这些多维数组都是张量 (Tensor)。

2d卷积核

- 卷积核是一个多维数组，其中参数由学习算法得到的
- 输出长度计算：
 - 输入的长度(W)，
 - 卷积核的大小(F)，
 - 卷积核移动的步长stride(S)，
 - 填充zero padding(P)
 - 输出的长度 $L = (W-F+2P)/S+1$
- 并行化：
 - 做一个和输出一样大小的Layer，
 - Layer里面所有的神经元参数都一样！



3d卷积核



- 输入是3d的
- 有多个卷积核

卷积网络

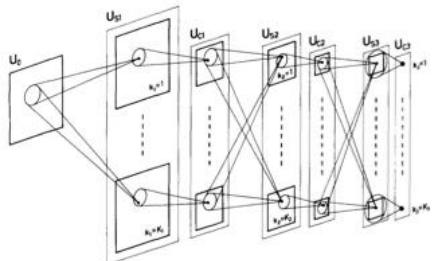
- 卷积网络 (Convolutional neural network, 简称CNN)
- 特点：局部区域的权重W共用 (weight sharing) (空间维度)
- 每一个卷积层后通常紧跟着一个下采样层subsample
 - 如max-pooling 方法或avg-pool完成下采样。



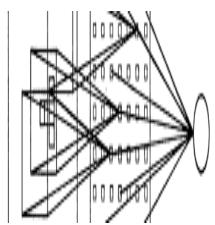
卷积网络

- 权重共享 (Parameters sharing) 和池化 (pooling) 操作
 - 利用了波形或图像信息的局部和谐性 (local coherence) , 学习其中的不变量特征 (invariant features)
- 卷积网络的基本的结构 (a series of stages of the form) :
 - Convolution/bias/non-linearity activation (ReLU or sigmoid functions)/pooling
 - 有时需要添加进去正则化层 (Normalization layers) , 如LCN(local contrast normalization)
 - LCN操作在最大池化层之后, 其目标是减去平均值, 除以标准差。LCN操作具有亮度不变性的特点, 对于图像识别用处很大。

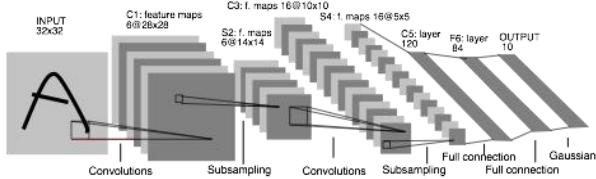
卷积网络发展



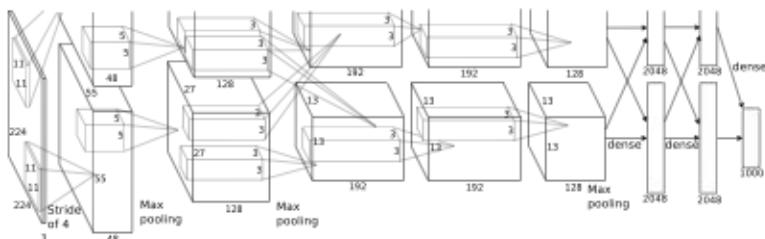
Fukushima 1980
Neocognitron



Rumelhart, Hinton, Williams 1986
“T” versus “C” problem



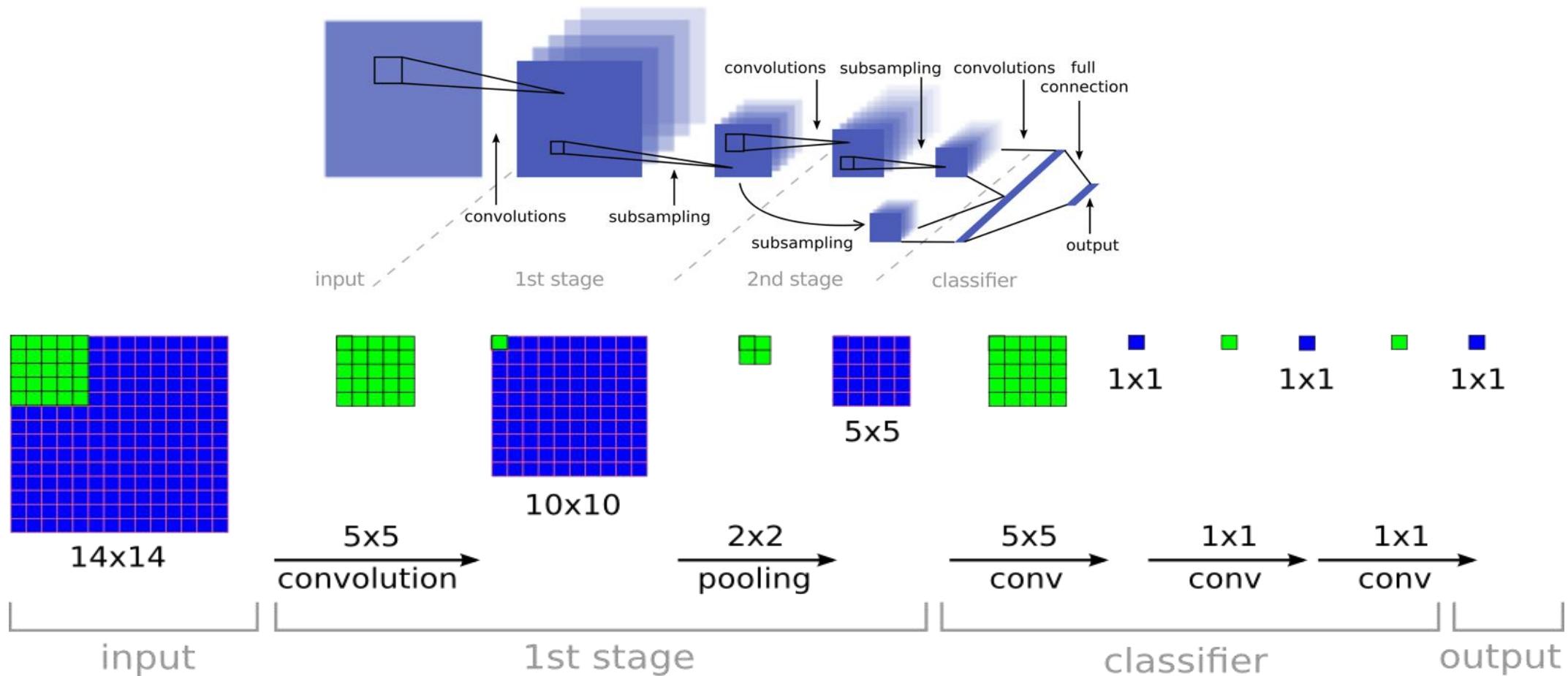
LeCun et al. 1989–1998
Hand-written digit reading



Krizhevsky, Sutskever, Hinton 2012
ImageNet classification breakthrough “SuperVision” CNN



LeNet-5：手写字体识别



[x] LeCun, Y., et al. "Handwritten digit recognition with a back-propagation network." NIPS, 1989.

卷积网络的新发展

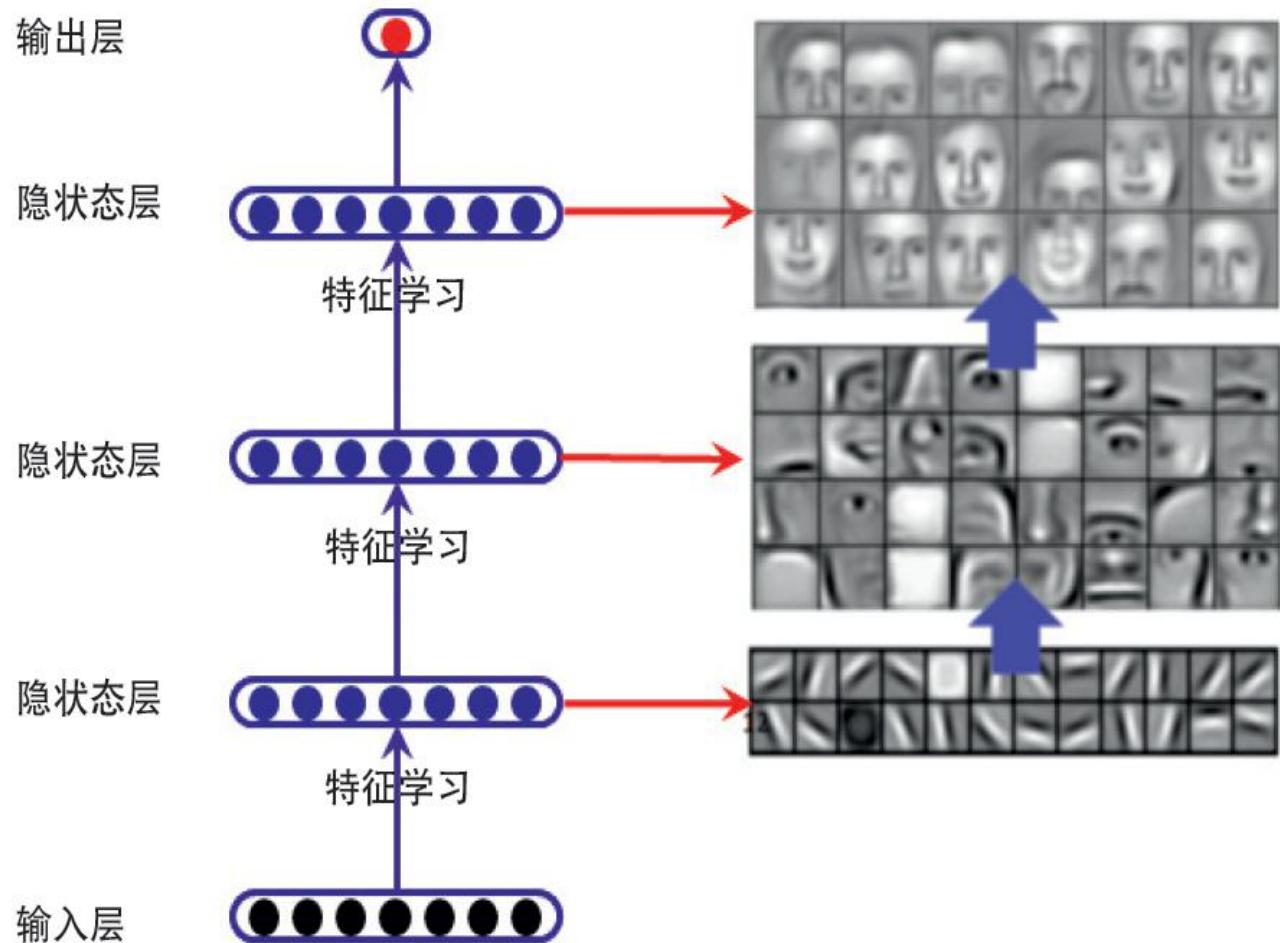
- 新的网络结构不断提出：
 - AlexNet—>NiN—>VGG—>GoogleNet—>ResNet —>DenseNet
- 网络规模与深度都在增加，准确率提升
 - NiN 引入 1×1 卷积层（Bottleneck layer）和全局池化；
 - VGG将 7×7 卷积核替换成3个 3×3 卷积核，起到了降参数的作用；
 - GoogLeNet引入了Inception模块；
 - ResNet引入了残差思想，增加了Skip Connection；
 - DenseNet的DenseBlock中将当前层的输出特征，与之后所有的层做直连。

卷积网络

- Densely Connected Convolutional Networks, CVPR 2017. <https://arxiv.org/abs/1608.06993>.
- SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size.
<https://arxiv.org/abs/1602.07360>.
- Xception: Deep Learning with Depthwise Separable Convolutions.
<https://arxiv.org/abs/1610.02357>.
- ResNet: Deep residual learning for image recognition, CVPR 2016.
<https://arxiv.org/abs/1512.03385>.
- GoogLeNet(Inception V3): Going deeper with convolutions, CVPR 2015.
<https://arxiv.org/abs/1409.4842>.
- VGG: Very Deep Convolutional Networks for Large Scale Image Recognition, 2014.
<https://arxiv.org/abs/1409.1556>.
- NiN: Network In Network, 2013. <https://arxiv.org/abs/1312.4400>.

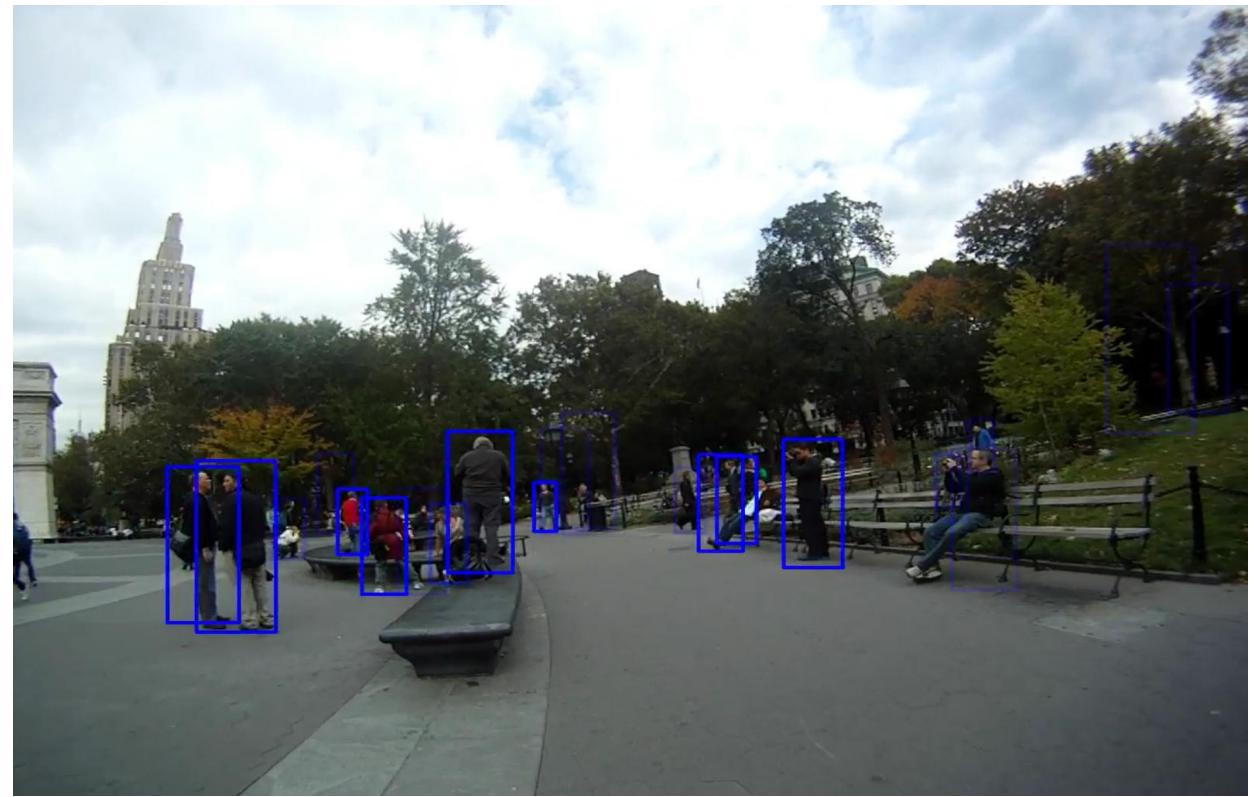
深度卷积网络—人脸识别

- 特征学习 (feature learning)
- 特征工程 (feature engineering)
- 特征学习取代了特征工程
 - 学习得到的特征
 - 图像化显示



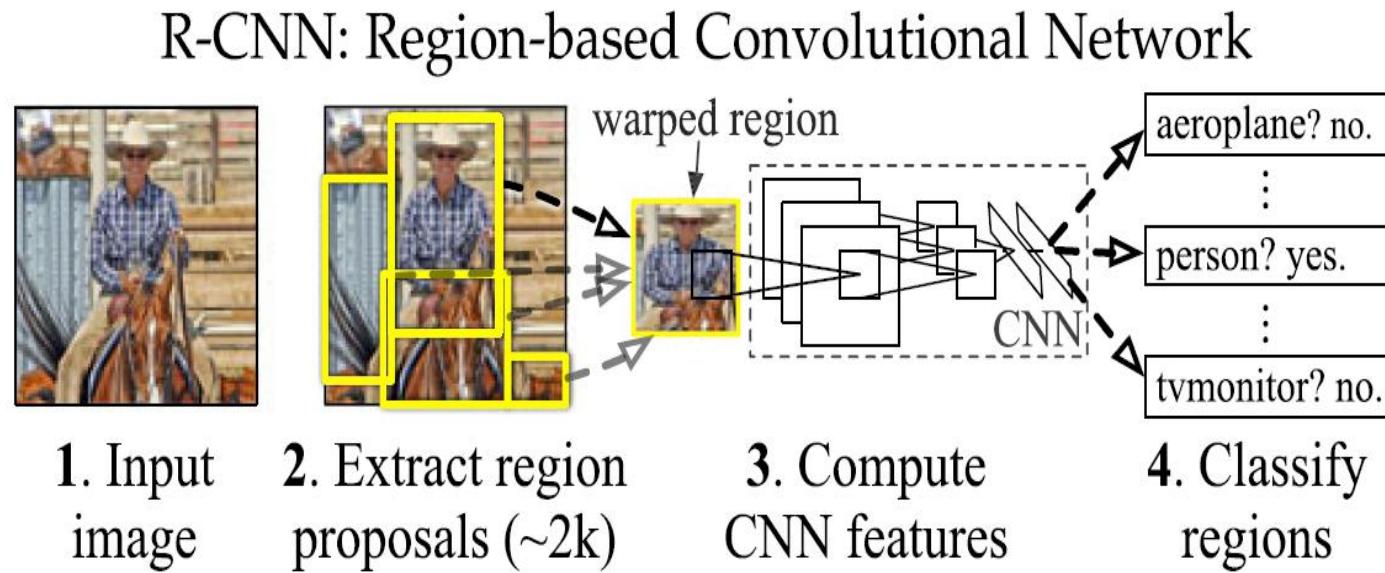
行人检测(Pedestrian detection)

- 自动驾驶技术
 - 交通标志，如信号灯检测
 - 路人检测
 - 车道线



卷积网络_计算机视觉_对象检测

- R-CNN (Region-based Convolutional Network method)
 - Region based convolutional networks for accurate object detection and segmentation, TPAMI, 2015.
 - Rich feature hierarchies for accurate object detection and semantic segmentation, CVPR 2014.



卷积网络_计算机视觉CV_对象检测OD

- [x] Rich feature hierarchies for accurate object detection and semantic segmentation, CVPR 2014.
- [x] Region based convolutional networks for accurate object detection and segmentation, TPAMI, 2015.
- [x] Fast R-CNN, ICCV 2015.
- [x] Faster R-CNN Towards real-time object detection with region proposal networks, NIPS, 2015.
- [x] Show and Tell: Lessons learned from the 2015 MSCOCO Image Captioning Challenge, IEEE PAMI 2016.
- [x] SSD: Single Shot MultiBox Detector, ECCV 2016.
- [x] You Only Look Once: Unified, Real-Time Object Detection, CVPR 2016.

循环网络

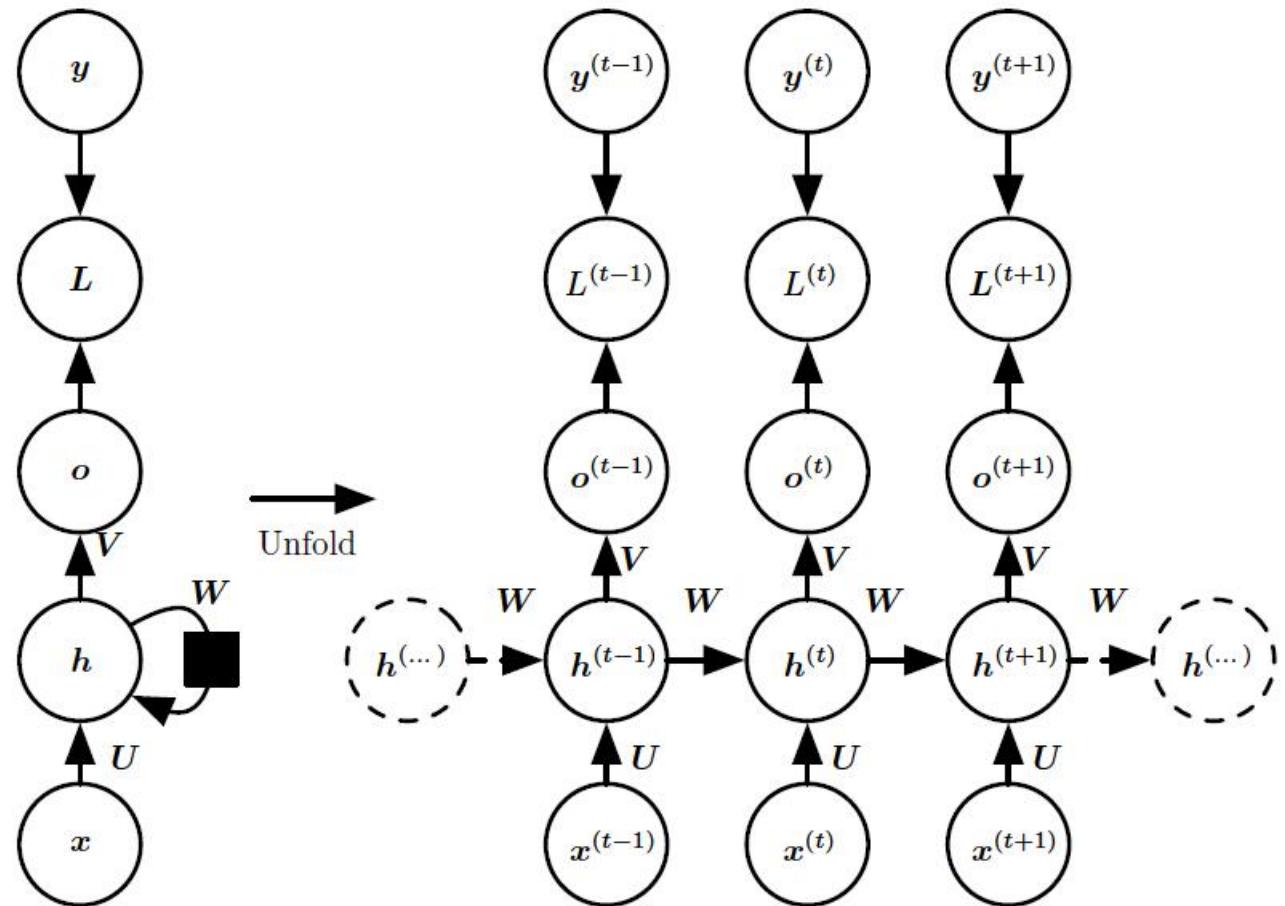
Recurrent network

循环网络—RNN

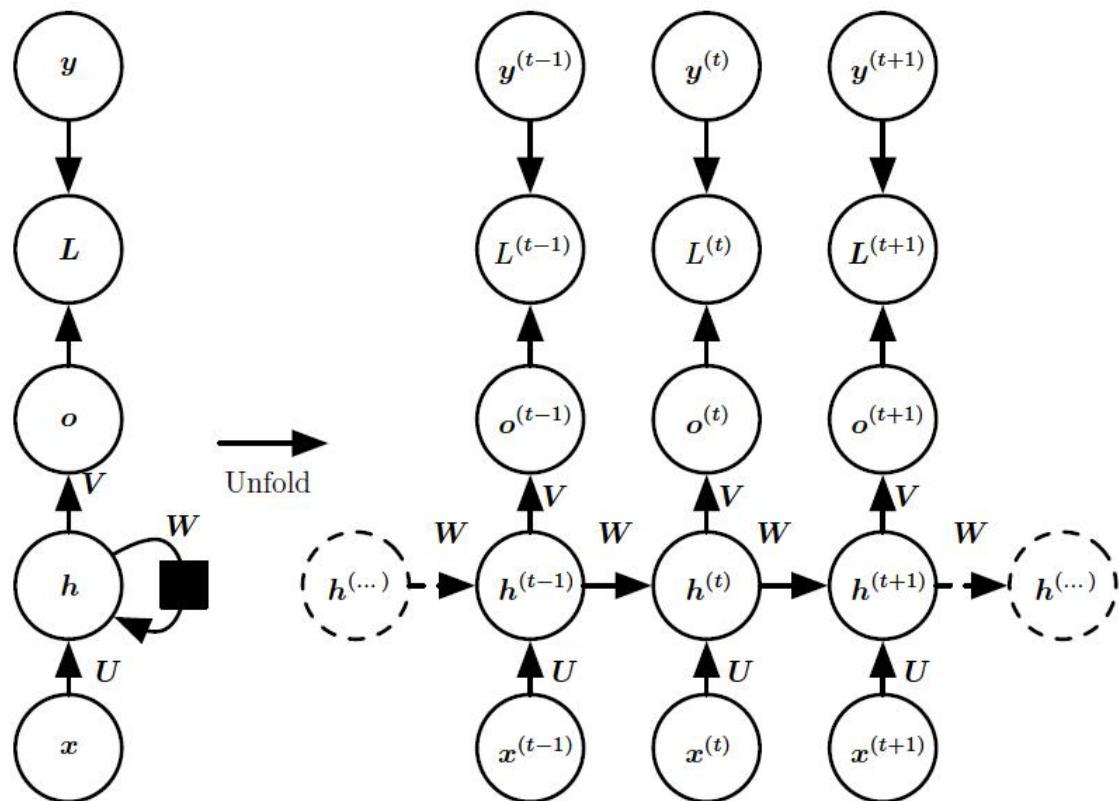
- 循环网络 (Recurrent neural network, 简称RNN)
- 在时间维度上，每一个时间步处理时，采用共享的权重
- 用于序列建模预测问题：
 - 手写识别 (handwriting recognition)、语音识别 (speech recognition)
 - 诗歌填词 (poem compose)、代码生成 (code writing)
 - 股价预测 (stock price)、天气预测 (weather forecast)
 - 机器翻译 (machine translation)、图片注释 (image caption)
 - 对联撰写 (Couplet writing)
- The Unreasonable Effectiveness of Recurrent Neural Networks,
<http://karpathy.github.io/2015/05/21/rnn-effectiveness/>.
- 对联自动生成：https://github.com/huangshiyu13/couplet_generation

RNN

- 循环网络结构
 - y 是训练目标
 - L 是损失函数
 - o 是网络输出
 - h 是状态（隐藏单元）
 - x 是网络输入
- 计算图的时间步上展开
- 举例：天气预测



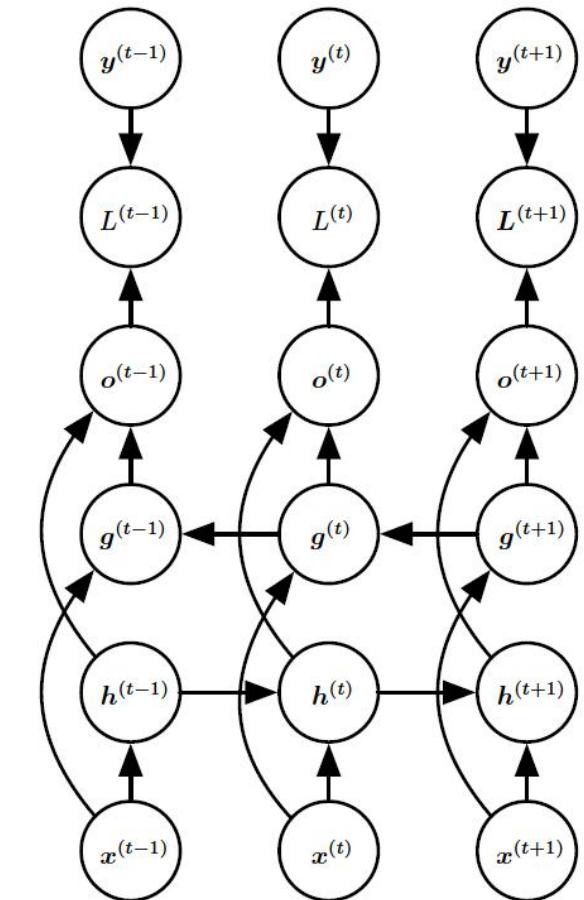
权重共享



- 循环神经网络在不同的时间步上采用相同的 U 、 V 、 W 参数
- 输入到隐藏的连接由权重矩阵 U 参数化
- 隐藏到输出的连接由权重矩阵 V 参数化
- 隐藏到隐藏的循环连接由权重矩阵 W 参数化

双向RNN (Bidirectional RNN)

- 一个时间上从序列起点开始移动的RNN 和另一个时间上从序列末尾开始移动的RNN。
- $h(t)$ 代表通过时间向前移动的子RNN 的状态, $g(t)$ 代表通过时间向后移动的子RNN 的状态。
- 输出单元 $o(t)$, 能够计算同时依赖于过去和未来且对时刻 t 的输入值最敏感的表示
- 用途: 手写识别和语音识别

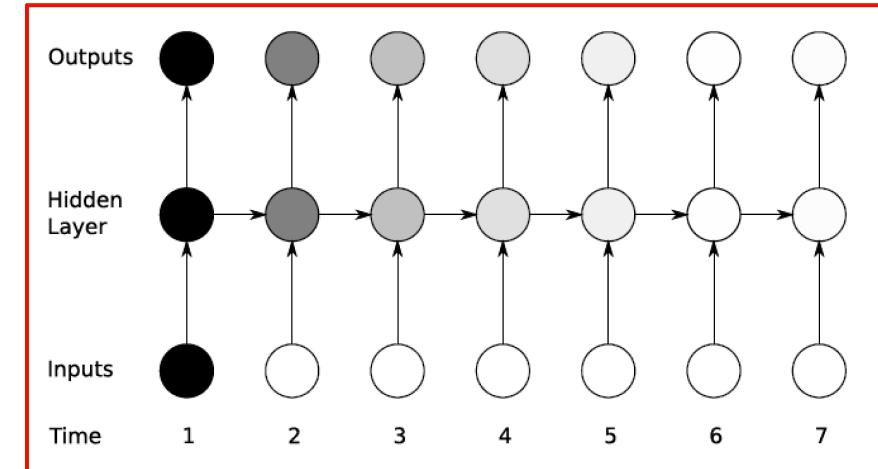


通过时间反向传播 (back-propagation through time, BPTT)

- 通过时间反向传播算法 (back-propagation through time, BPTT) 应用于展开图的反向传播参数调节。
- 梯度计算涉及执行一次前向传播 (从左到右的传播)，一次由右到左的反向传播。
- 运行时间是 $O(T)$ ，并且不能通过并行化来降低，因为前向传播图是固有顺序的。
- 每个时间步只能一前一后地计算。前向传播中的各个状态必须保存，直到它们反向传播中被再次使用，因此内存代价也是 $O(T)$ 。
- 损失函数的梯度关于各个参数的计算是计算成本很高的操作。
- 由此可见，循环网络非常强大但训练代价也很大。

RNN训练的问题

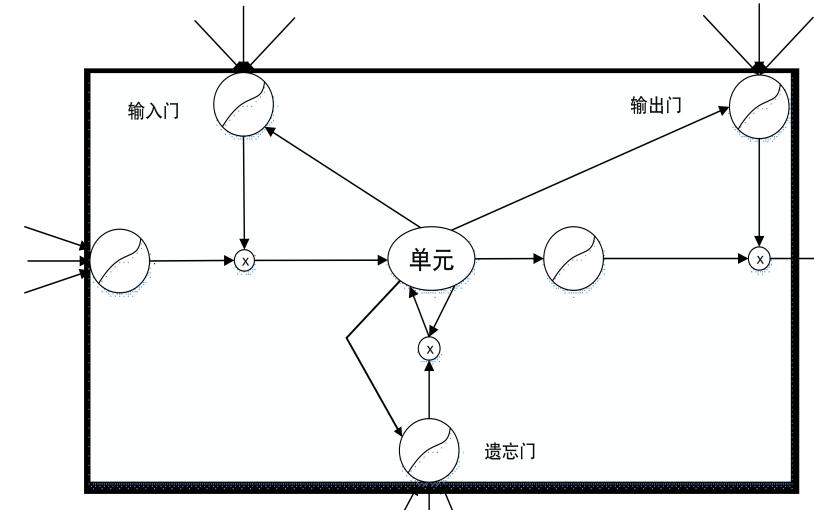
- 梯度截断 (Gradient Clipping)
 - 训练循环网络时，经常出现梯度要么太大，要么太小，为了加速训练，需要把梯度设置为一些固定数值。
 - 比如说，梯度的任何维度的数值应该小于1，如果某个维度的数值大于1，则固定设置为1.
- BPTT训练RNN时的梯度消失和梯度爆炸问题
 - 最初和最后的时间步的梯度的幅度过大或者过小
 - ReLU函数的问题
 - Sigmoid函数的饱和问题
- LSTM解决以上问题



[1] S. Hochreiter, Y. Bengio, P. Frasconi, and J. Schmidhuber. Gradient flow in recurrent nets: the difficulty of learning long-term dependencies. *A Field Guide to Dynamical Recurrent Neural Networks*, 2001.

长短时记忆网络 (Long—short term memory)

- LSTM是RNN的一个改进，LSTM增加了一个主输入单元和其他三个辅助的门限输入单元：
 - 记忆单元 (memory cell) 、输入门 (input gate) 、遗忘门 (forget gate) 及输出门 (output gate) 。
- 三个辅助输入分别控制网络是否输入，是否存储输入以及是否输出。
 - 输入门 (Input gate) 控制是否输入，遗忘门 (Forget gate) 控制是否存储，输出门 (Output gate) 控制是否输出。
- 辅助单元可以寄存时间序列的输入，在训练过程中会利用后向传播的方式进行。
- 记忆单元和这些门单元的组合，大大提升了RNN处理远距离依赖问题的能力，解决RNN网络收敛慢的问题。



<https://distill.pub/2019/memorization-in-rnns/>

LSTM

- 前向方程

$$h_t = \mathcal{H}(W_{xh}x_t + W_{hh}h_{t-1} + b_h) \quad (1)$$

$$y_t = W_{hy}h_t + b_y \quad (2)$$

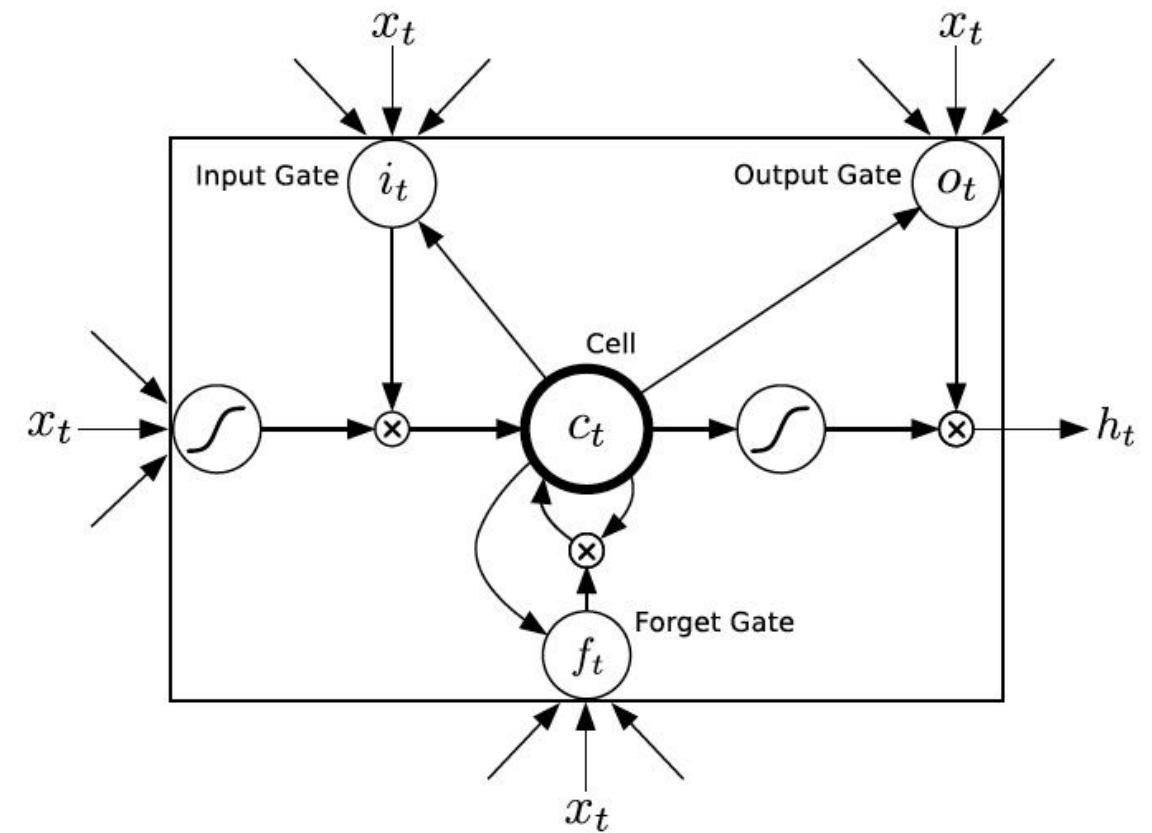
$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \quad (3)$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \quad (4)$$

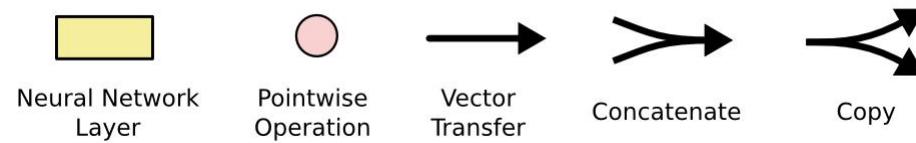
$$c_t = f_t c_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad (5)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o) \quad (6)$$

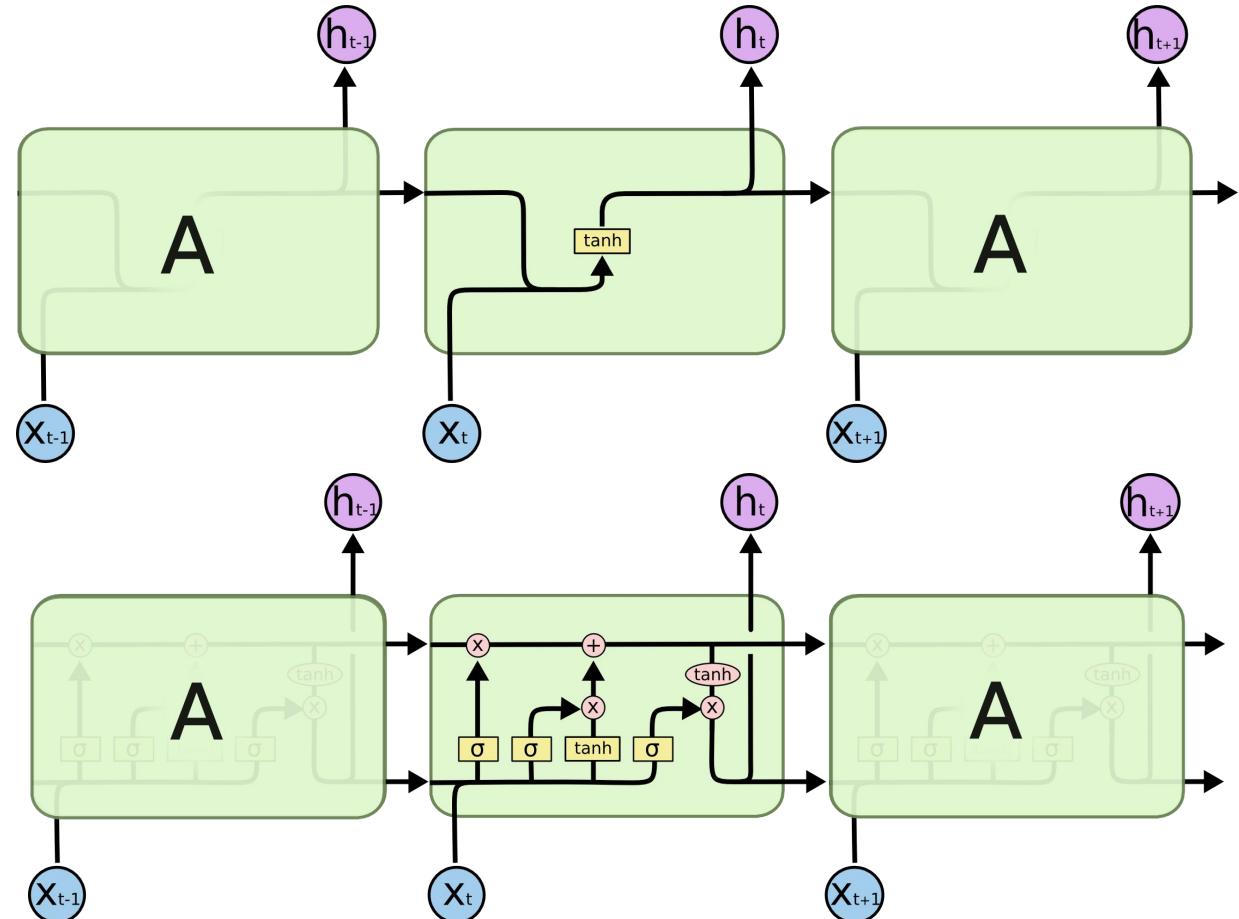
$$h_t = o_t \tanh(c_t) \quad (7)$$



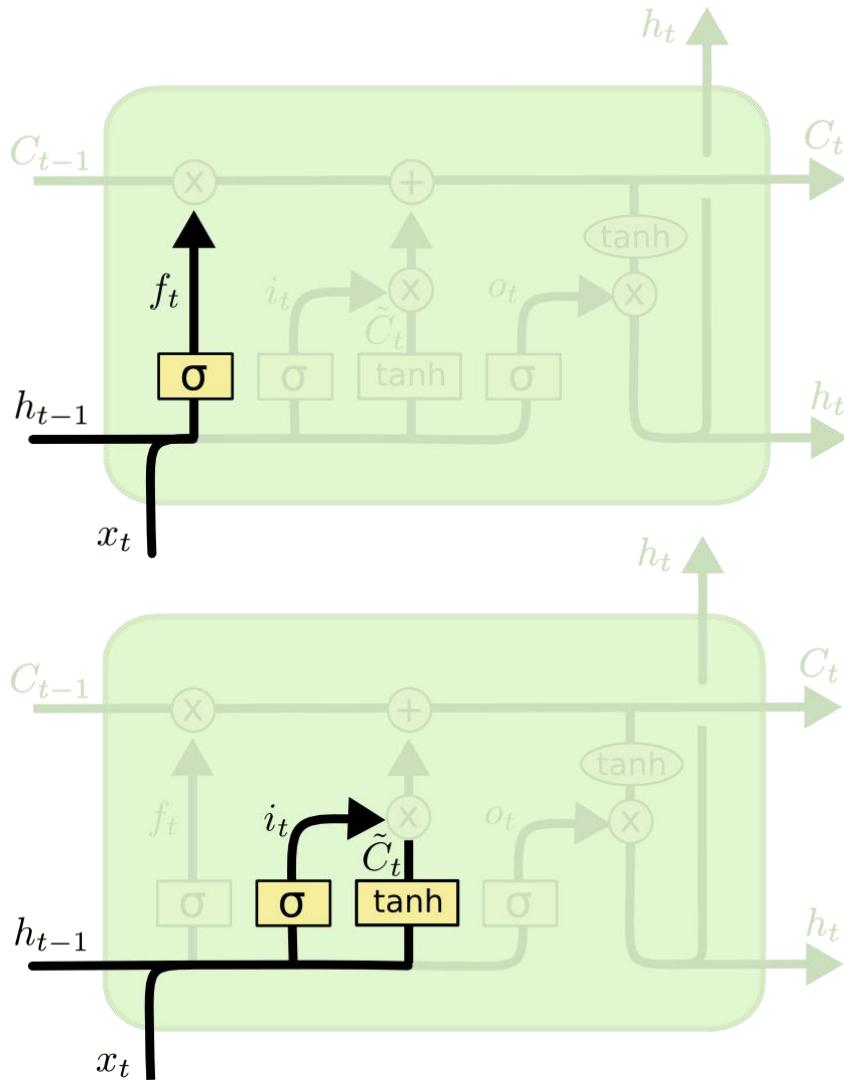
RNN与LSTM



- 记忆单元C
- 遗忘门f
- 输入门i
- 输出门o



LSTM-1

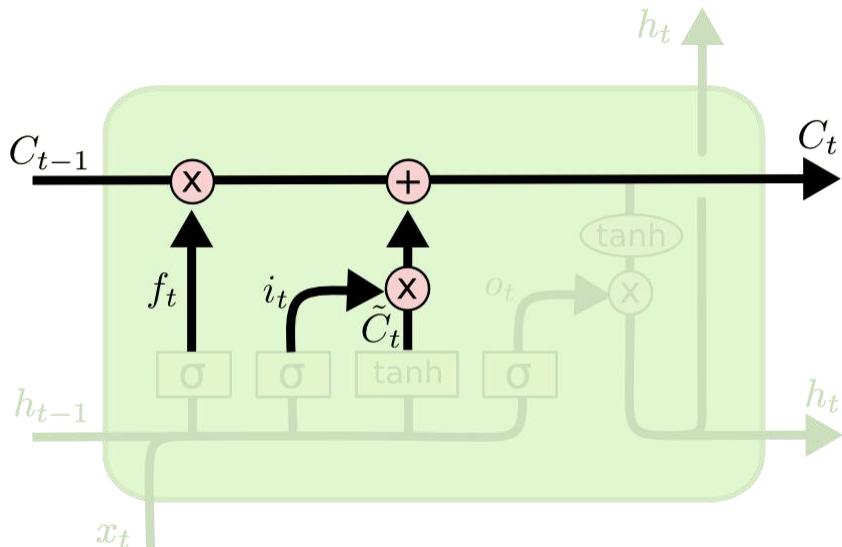


$$f_t = \sigma (W_f \cdot [h_{t-1}, x_t] + b_f)$$

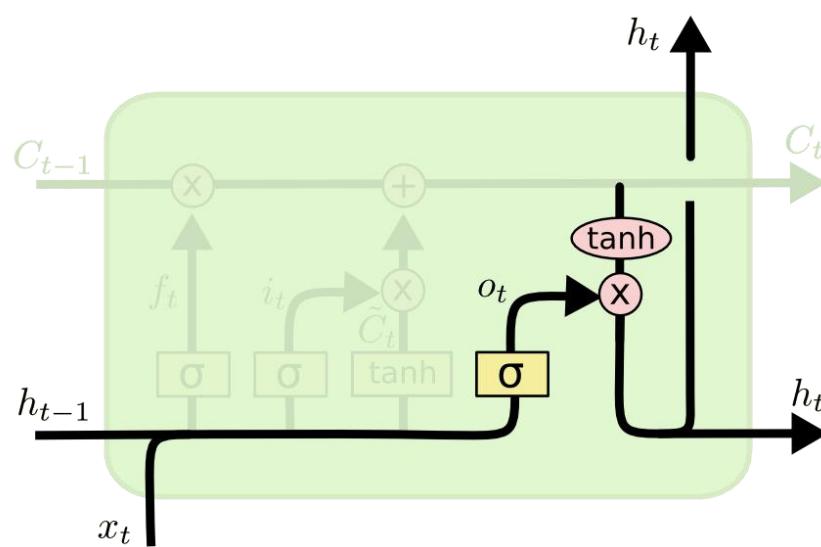
$$i_t = \sigma (W_i \cdot [h_{t-1}, x_t] + b_i)$$

$$\tilde{C}_t = \tanh (W_C \cdot [h_{t-1}, x_t] + b_C)$$

LSTM-2



$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

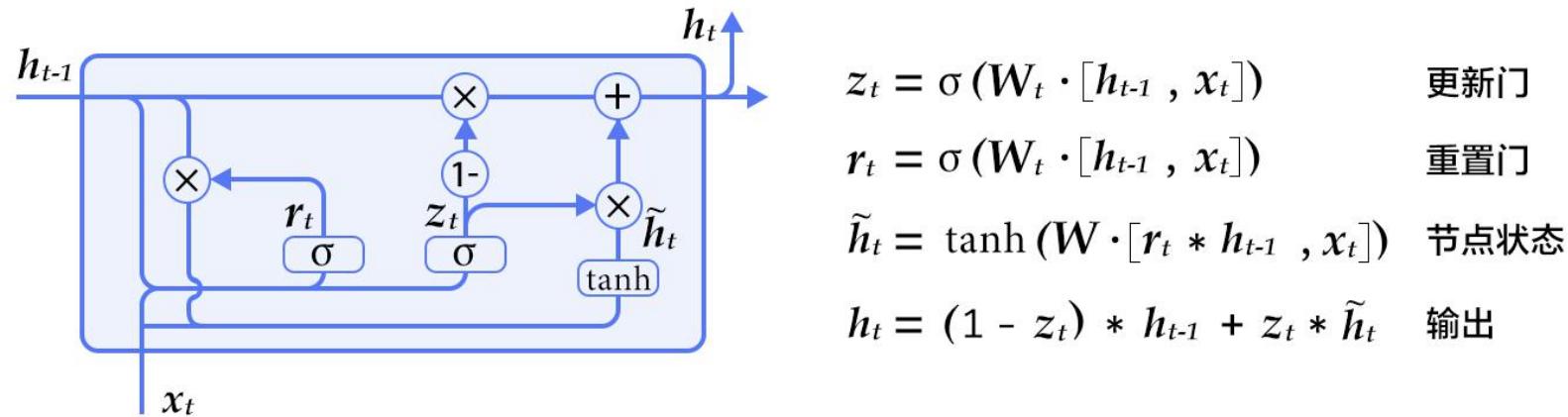


$$o_t = \sigma (W_o [h_{t-1}, x_t] + b_o)$$

$$h_t = o_t * \tanh (C_t)$$

门控循环单元 (Gated Recurrent Unit)

- GRU (Gated Recurrent Unit) 是Cho等提出的LSTM的简化版本，也是RNN的一种变种
- GRU单元只有两个门：
 - 更新门 (update gate)，LSTM的输入门 (Input Gate) 和遗忘门 (Forget Gate) 合并，用于控制历史信息对当前时刻隐层输出的影响。如果更新门接近1，会把历史信息传递下去。
 - 重置门 (reset gate)，如果重置门关闭，会忽略掉历史信息，即历史不相干的信息不会影响未来的输出。
- LSTM的记忆状态C (cell state) 和隐藏状态H (hidden state) 合并为一个状态



序列对序列模型

Sequence-to-Sequence Models

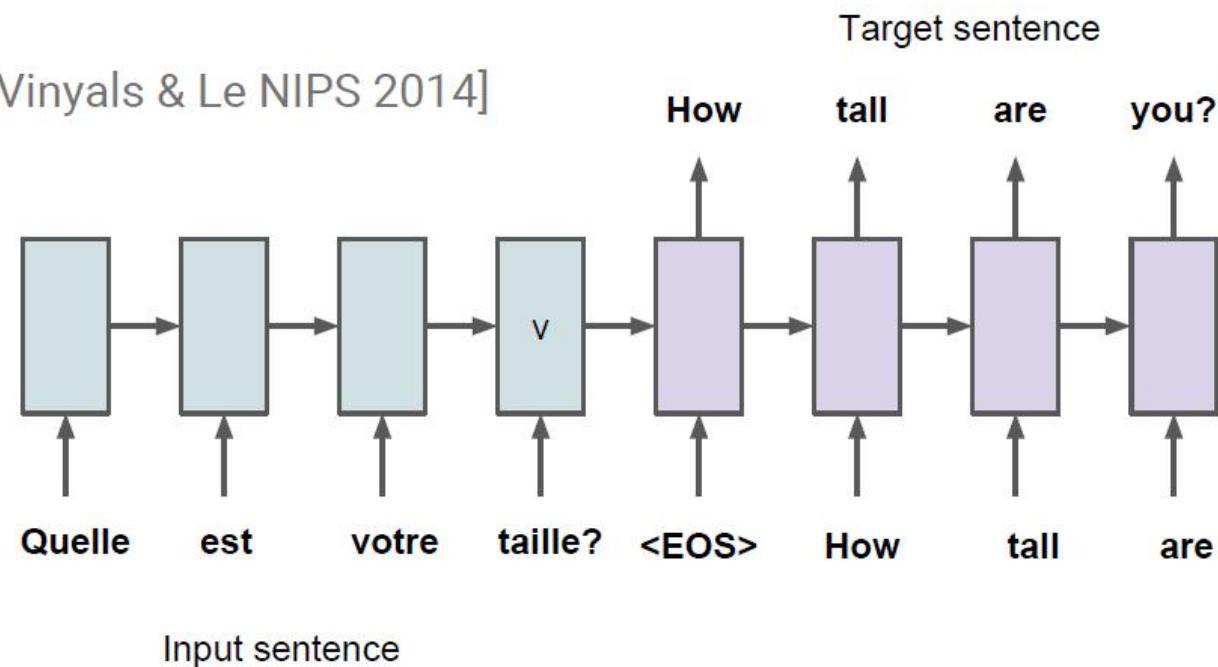
对联自动生成

input : 青山不墨千秋画	output : 绿水无弦万古琴
input : 春眠不觉晓	output : 花落自知秋
input : 无边落木萧萧下	output : 不尽流云处处游
input : 两只黄鹂鸣翠柳	output : 一弯新月钓清溪
input : 万紫千红春无限	output : 五风十雨岁有余
input : 月透柳帘窥案卷	output : 风吹竹管动窗纱

[x] https://github.com/huangshiyu13/couplet_generation

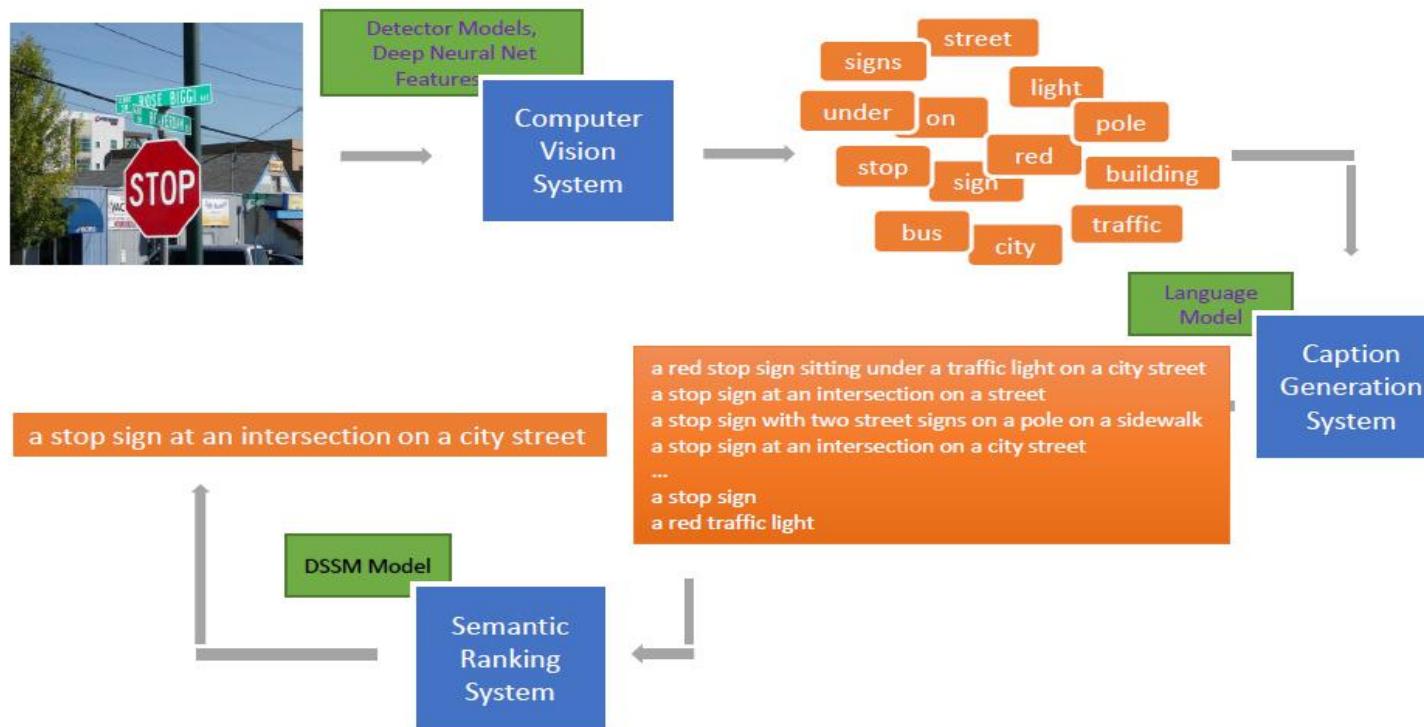
机器翻译（英语—法语）

[Sutskever & Vinyals & Le NIPS 2014]



- [x] I. Sutskever, O. Vinyals, & Q. V. Le, Sequence to sequence learning with neural networks. NIPS 2014.

深度学习的应用—图片注解



Fang, Hao, et al. "From captions to visual concepts and back." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015.

Thanks~

Questions?