干货 | 图解LSTM神经网络架构及其11种变体（附论文）

2016-10-02 机器之心

选自FastML

作者：Zygmunt Z.

机器之心编译

参与：老红、李亚洲

就像雨季后非洲大草原许多野生溪流分化成的湖泊和水洼，深度学习已经分化成了各种不同的专门架构。

并且，每个架构都会有一个图解，这里将详细介绍它们。

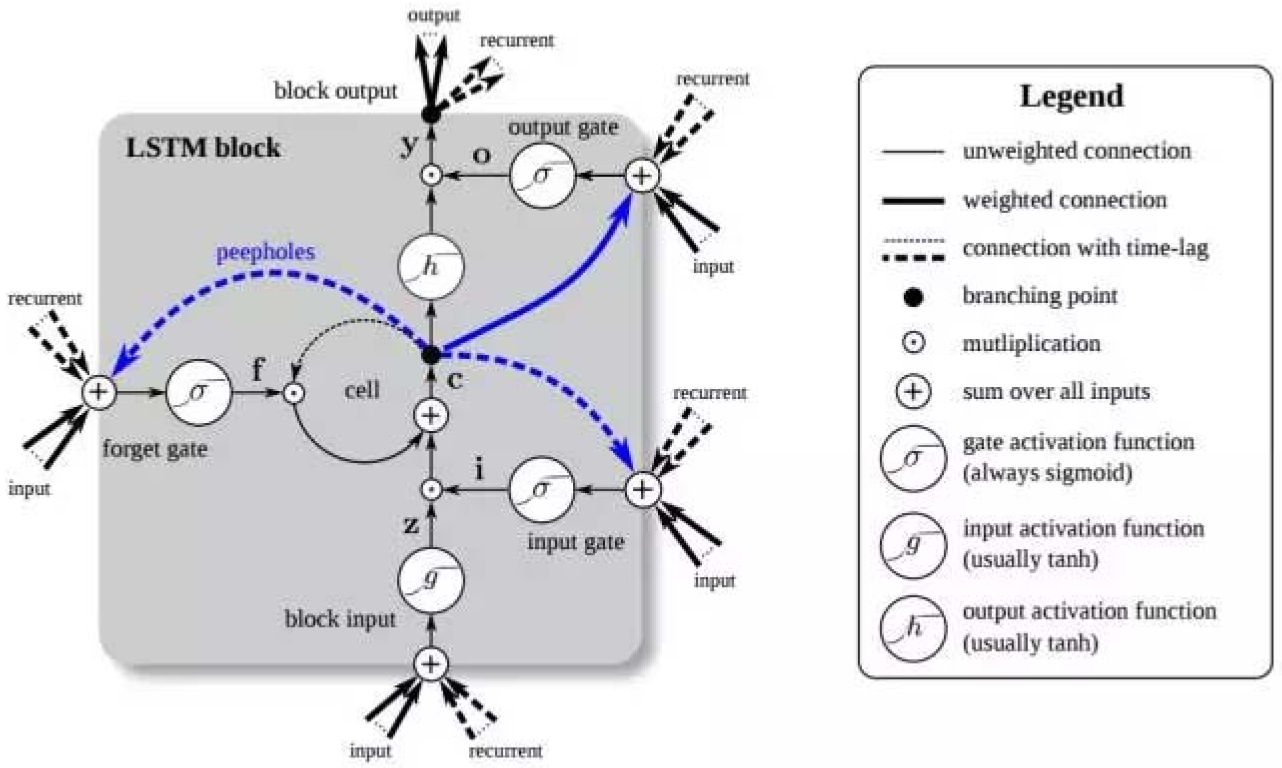
神经网络在概念上很简单，并且它们十分动人。在层级上，有着一堆同质化的元素和统一的单位，并且它们之间还存在在一系列的加权连接。这就是神经网络的所有，至少从理论上来说是这样。然而，时间证明的结果却有所不同。并非工程的特性，我们现在拥有的是建筑工程，而非工程的特性，正如 Stephen Merrity 描述的那样：

> *深度学习的浪漫主义描述通常预示着手工制作工程特性的日子一去不复返了，这个模型的本身是足以先进到能够解决问题的。正如大多数广告一样，它同时具备真实性和误导性。*
>
> *虽然深度学习在很多情况下简化了工程特性，但它肯定还没有彻底地摆脱它。随着工程特性的减少，机器学习模型本身的结构变得越来越复杂。大多数时候，这些模型架构会特定于一个给定的任务，就像过去的工程特性那样。*
>
> *需要澄清一下的是，这仍然是很重要的一步。结构工程要比工程特性更具一般性，并且提供了许多新的机会。正如我们提到的，我们不能无视这样一个事实：我们离我们想要达到的还很远。*

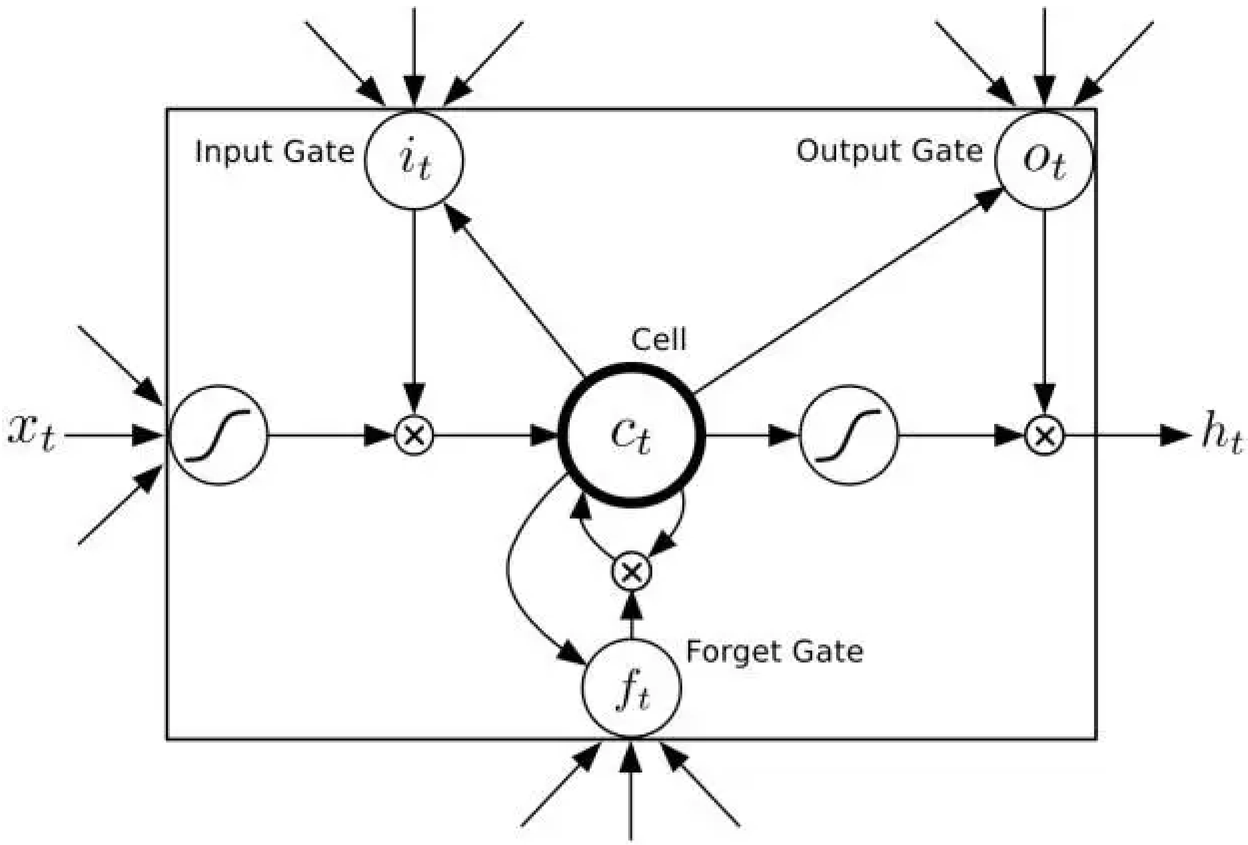
LSTM 图解

怎样解释这些架构？自然地，我们可以通过图解，图解往往可以让阐述变得更清晰。

让我们先来看看如今最流行的两种网络，CNN 和 LSTM：

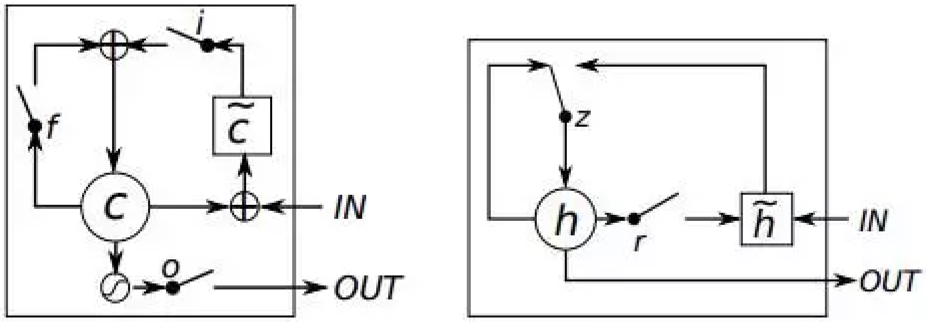


很简单吧，我们再更仔细地研究下：



正如大家所言，你可能有很多不理解的数学问题，但你会慢慢习惯它们。幸运地是，我们有很多非常好的解释。

仍觉得 LSTM 太复杂了？那让我们来试试简单的版本，GRU (Gated Recurrent Unit)，相当琐碎。



(a) Long Short-Term Memory (b) Gated Recurrent Unit

Figure 1: Illustration of (a) LSTM and (b) gated recurrent units. (a) i, f and o are the input, forget and output gates, respectively. c and $\tilde{c}$ denote the memory cell and the new memory cell content. (b) r and z are the reset and update gates, and h and $\tilde{h}$ are the activation and the candidate activation.

尤其是这一个，被称为 minimal GRU：

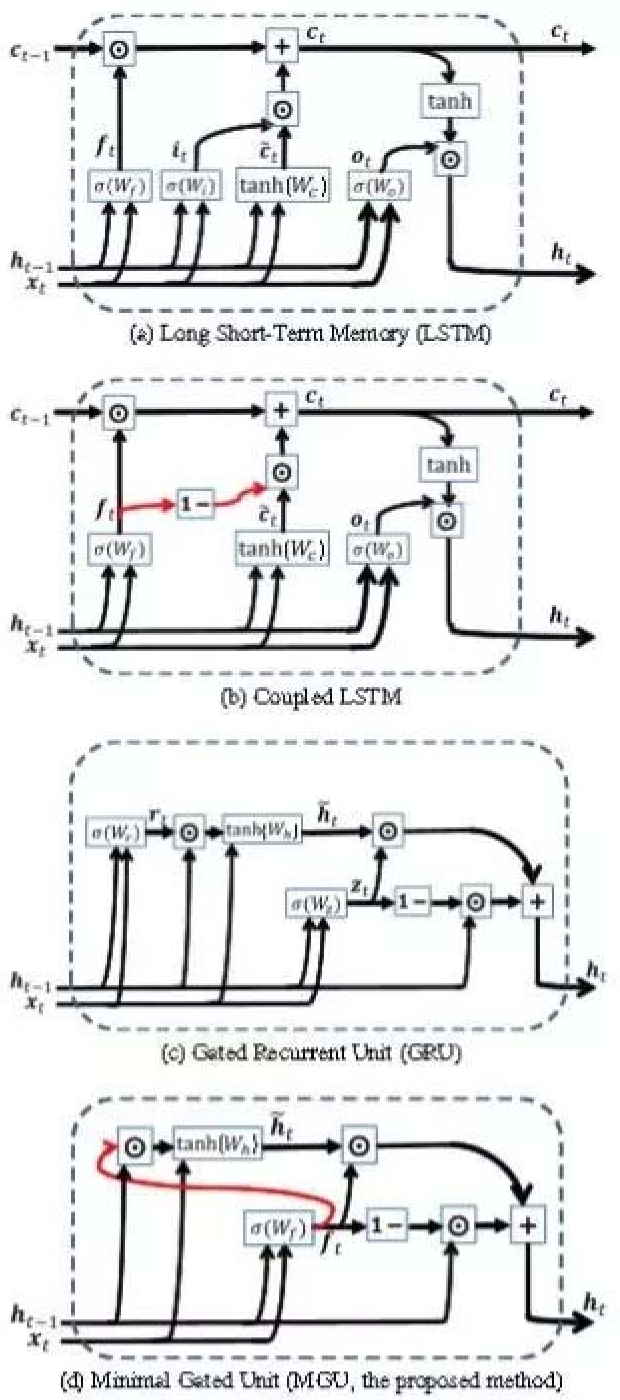


Figure 1. Data flow and operations in various gated RNN models. The direction of data flow are indicated by arrows, and operations on data are shown in rectangles. Five types of element wise operations (logistic sigmoid, tanh, plus, product and one minus) are involved. For operations with parameters (logistic sigmoid and tanh), we also include their parameters in the rectangle. These figures are different from diagrams that illustrate gates as switches, but match better to the equations in Table 1.

更多图解

LSTM 个多各样的变体如今很常见。下面就是一个，我们称之为深度双向 LSTM：

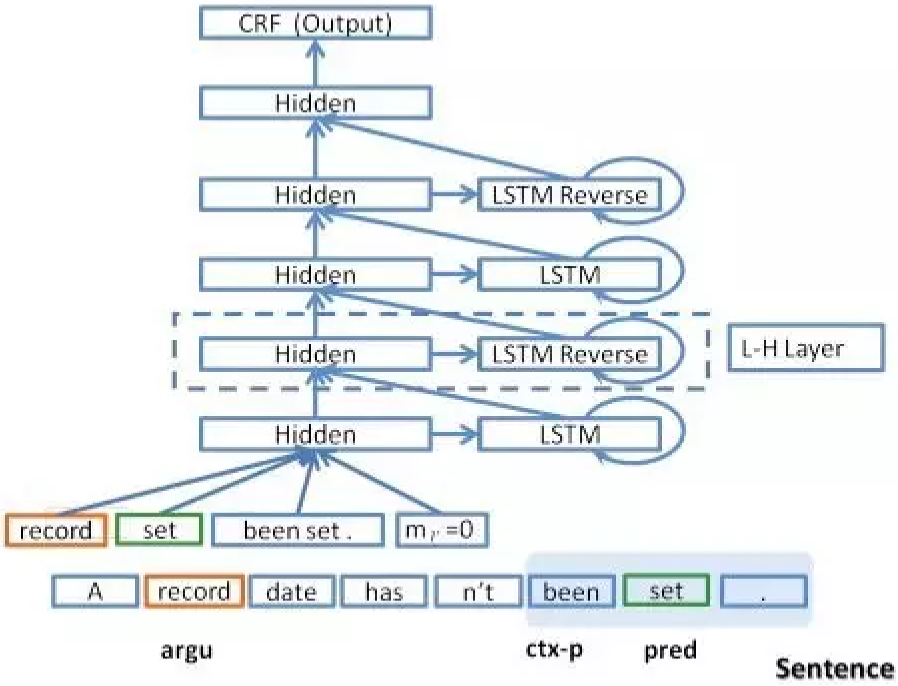


Figure 2: DB-LSTM network.Shadow part denote the predicate context within length 1.

DB-LSTM（参见论文：End-to-end Learning of Semantic Role Labeling Using Recurrent Neural Networks ）

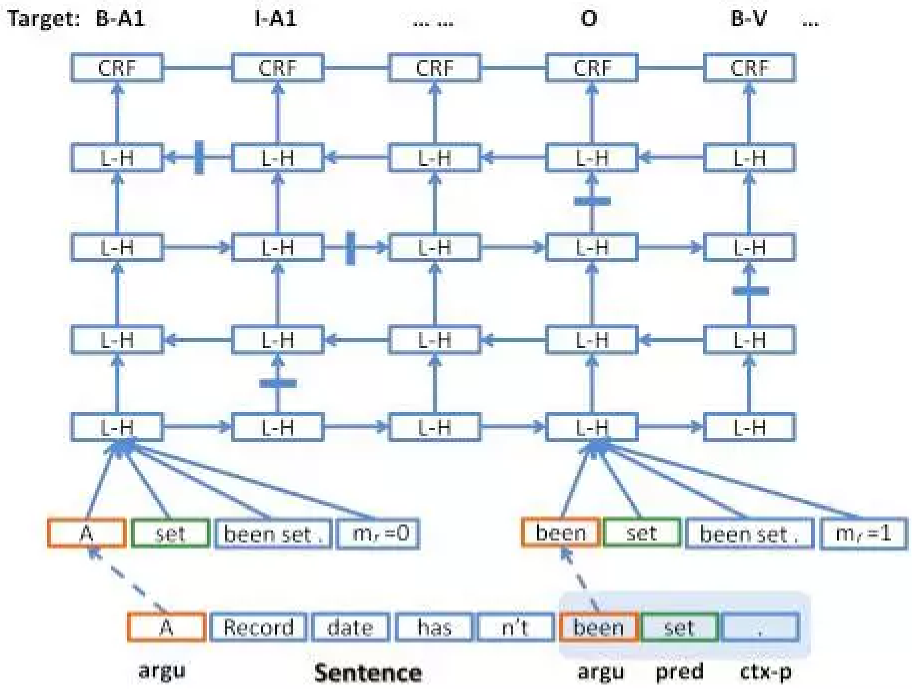


Figure 3: Temporal expanded DB-LSTM network. Bars denote that the connections are blocked by the closed gates. Shadow part denotes the predicate context.

剩下的也不需要加以过多说明。让我们从 CNN 和 LSTM 的结合开始说起：

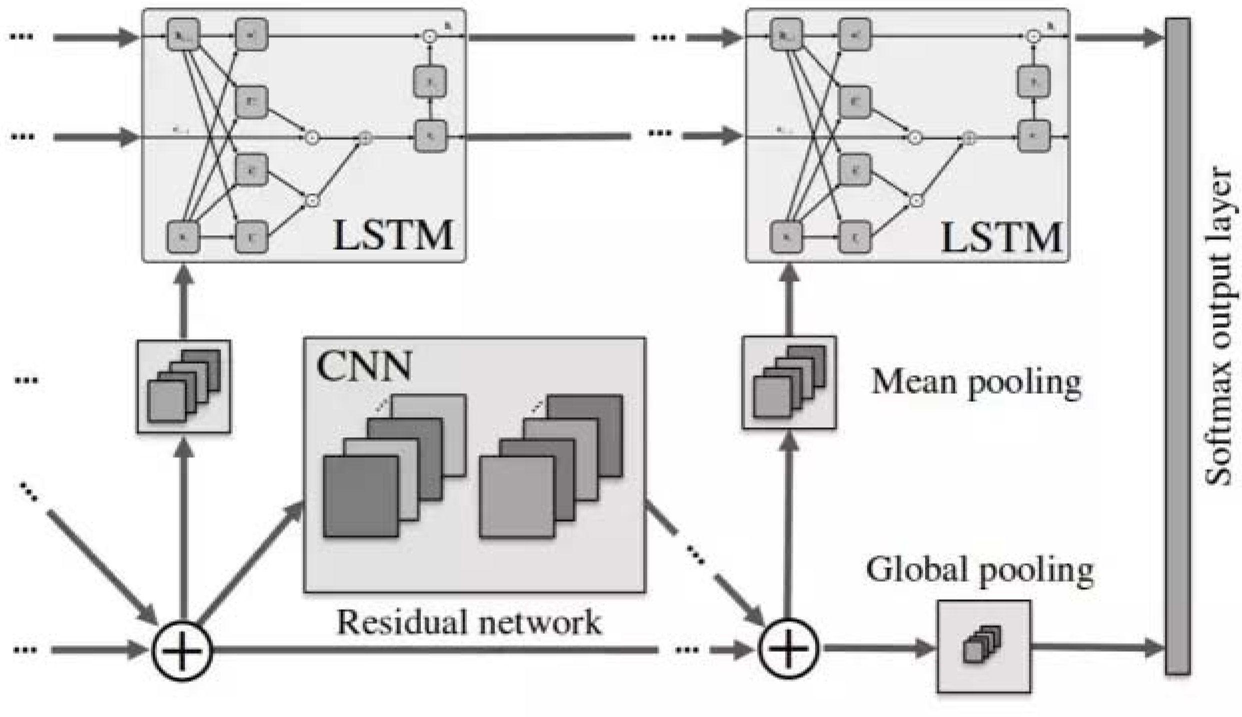
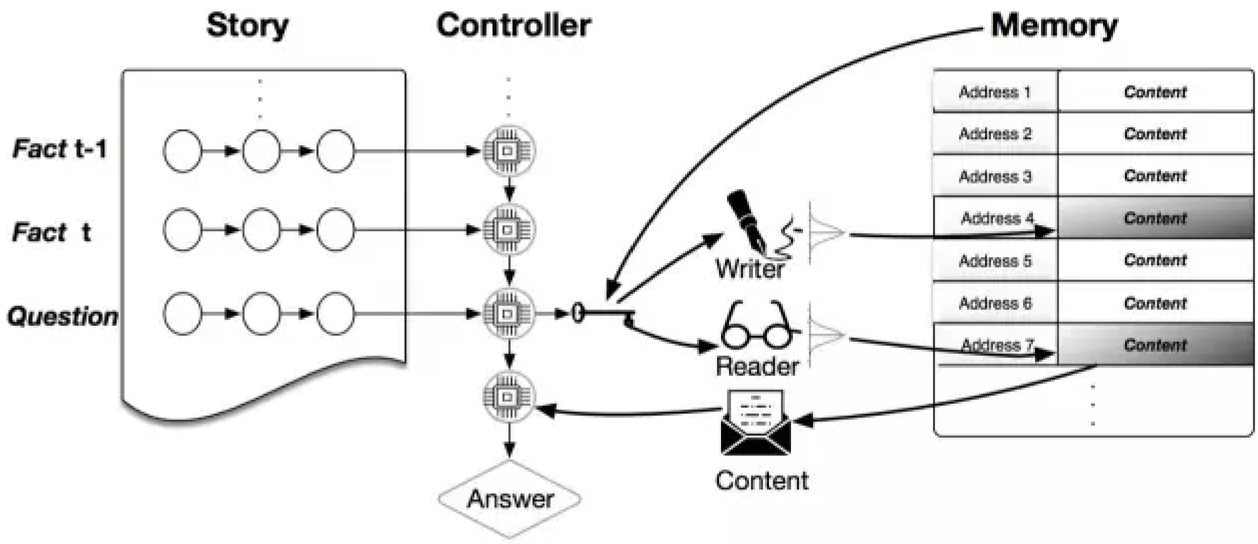


Figure 2: The repeating computational block and the final computational units of a convolutional residual memory network (CRMN).

卷积残差记忆网络（参见论文：Convolutional Residual Memory Networks）



动态 NTM（参见论文：Dynamic Neural Turing Machine with Soft and Hard Addressing Schemes）

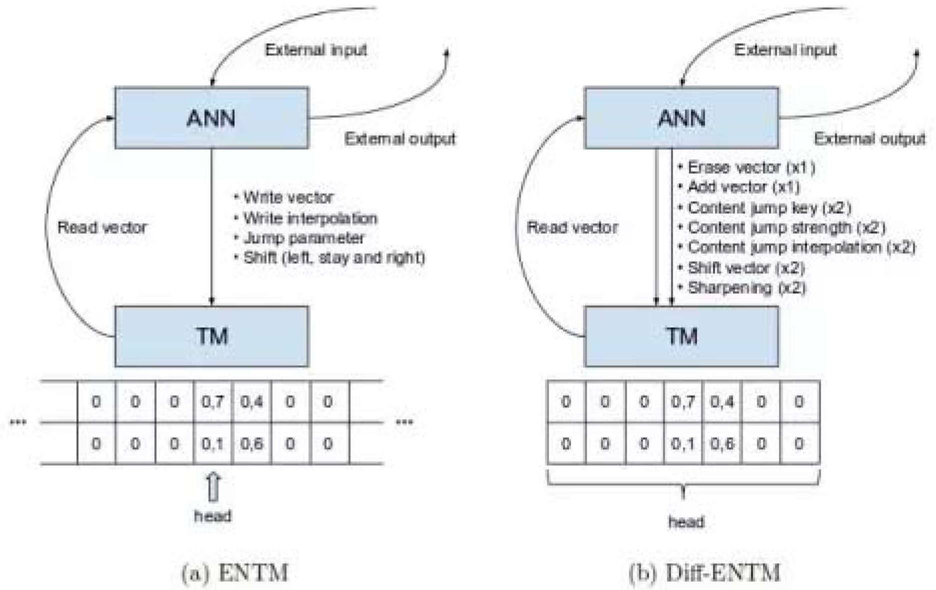
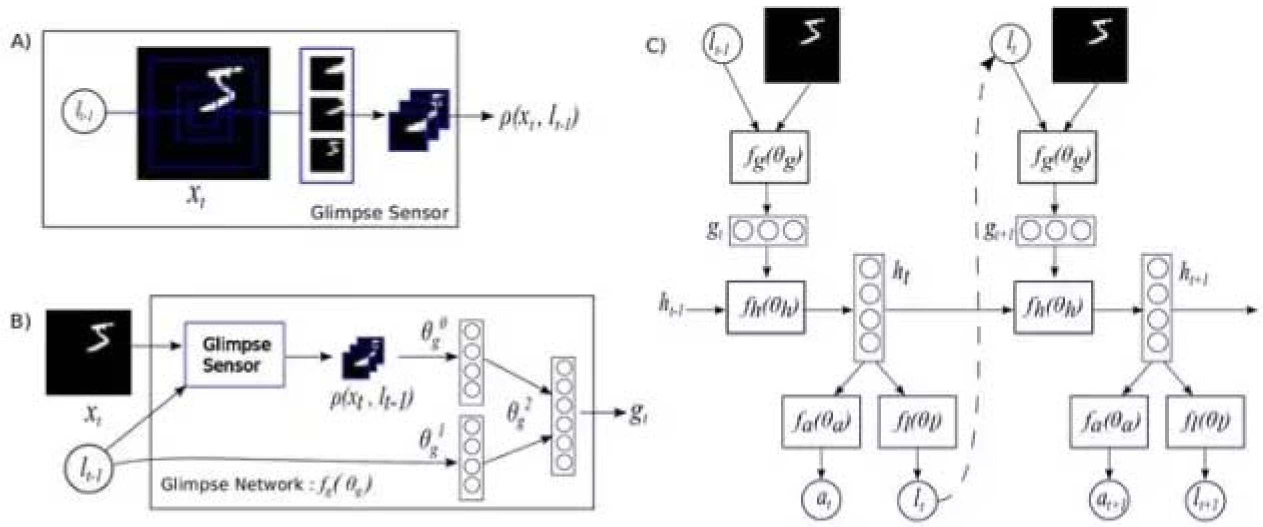
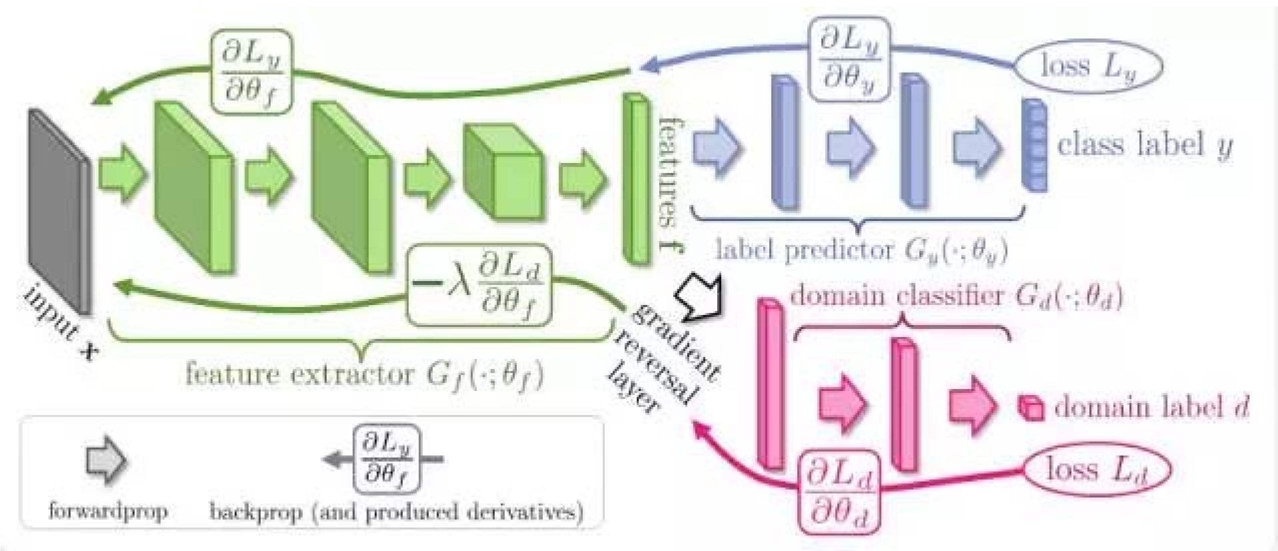


Figure 1: Evolvable Neural Turing Machines. This figure shows the activation flow between the ANN and the memory bank for the ENTM and the Diff-ENTM architecture. Extra ANN outputs determine the vector to be written to memory and the movement of the read and write heads. The ANN receives the content of the current memory location as input at the beginning of the next time-step. In addition to the NTM specific inputs and outputs, the ANN has domain dependent actuators and sensors. Notice how the Diff-ENTM head always addresses all memory locations, and has an increased number of connections from the ANN to the TM compared to the ENTM.

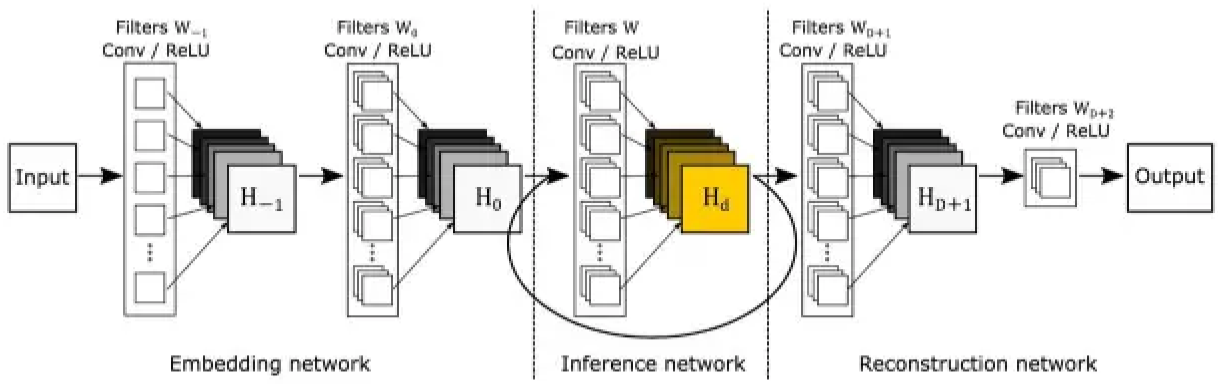
可发展神经图灵机（参见论文：Evolving Neural Turing Machines for Reward-based Learning）



视觉注意的循环模型（参见论文：Recurrent Models of Visual Attention）

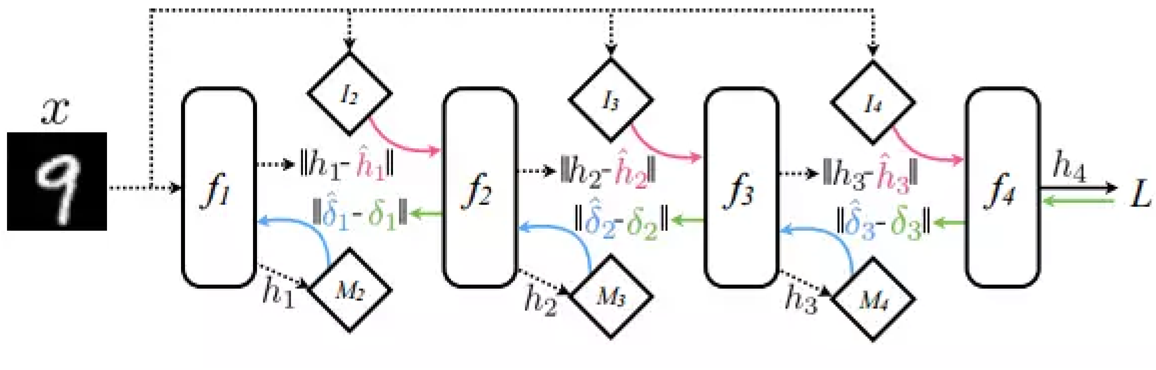


通过反向传播无监督域适应（参见论文：Unsupervised Domain Adaptation by Backpropagation）



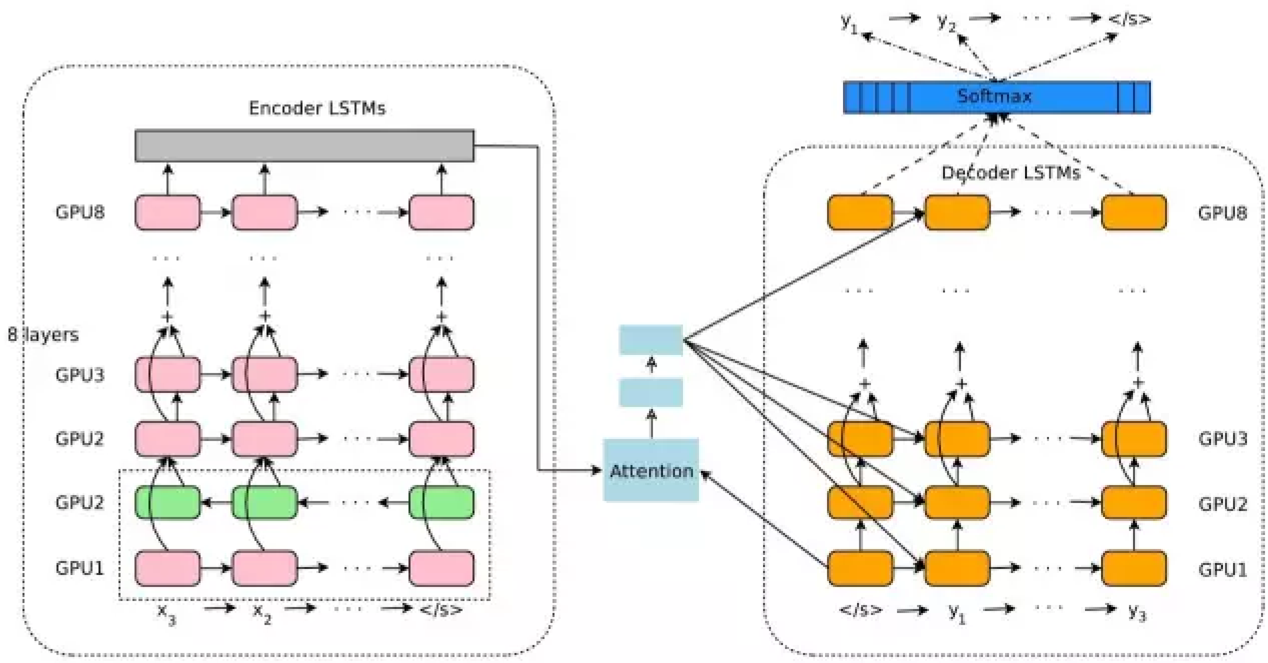
进行图像超分辨率的深度递归 CNN（参见论文：Deeply-Recursive Convolutional Network for Image Super-Resolution）

带有合成梯度的多层感知器的图解在清晰度上得分很高：



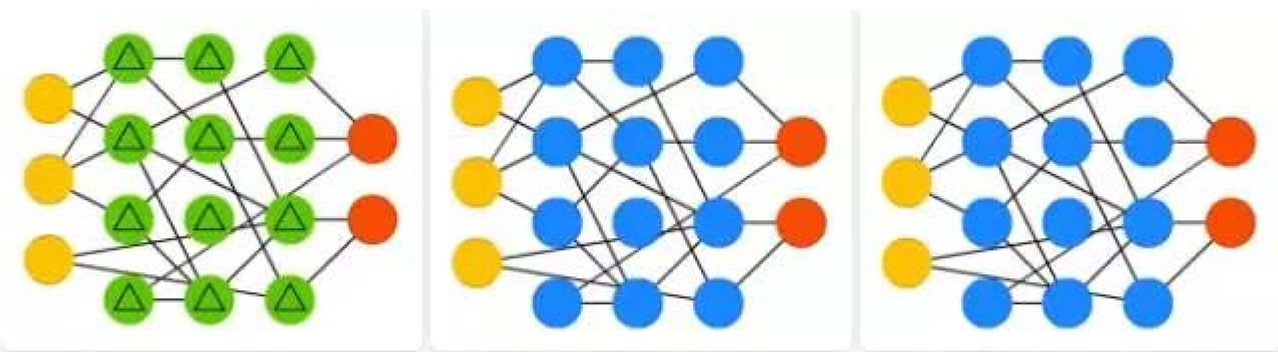
带有合成梯度的 MLP（参见论文：Decoupled Neural Interfaces using Synthetic Gradients）

每天都有新的成果出现，下面这个就是新鲜的，来自谷歌的神经机器翻译系统：



一些完全不同的东西

Neural Network ZOO（一篇描述神经网络架构的文章，机器之心同样进行了编译）的描绘非常简单，但很多都华而不实，例如：ESM, ESN 和 ELM。



它们看上去像没有完全连接的感知器，它们看上去像没有完全连接的感知器，但它们应该代表的是一种液体状态机、一个回声状态网络和一个极端学习机。

LSM 和 ESN 有何不同？很简单，LSM 有着三角状绿色的神经元。而 ESN 和 ELM 又有什么不同呢？它们都有蓝色的神经元。

讲真，虽然类似，,ESN 是一个递归网络而 ELM 则不是。而这种区别也可在架构图中见到。

阅读原文