

### TONGJI UNIVERSITY

# 遥感图像处理与分析课程报告

**姓 名 王忠**

**学 号 1633299**

**学 院 测绘与地理信息学院**

**专 业 测绘科学与技术**

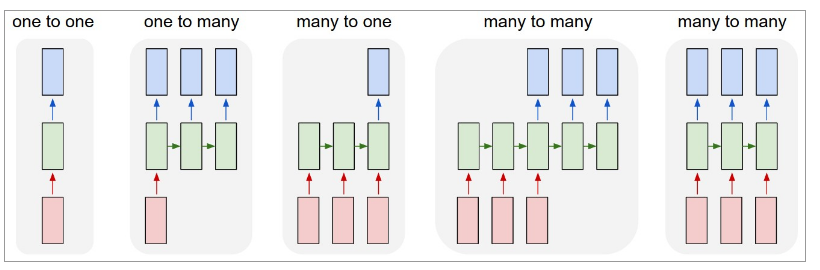
**课程教师 张 绍 明**

**联系电话 13162551729**

**2017 年 1 月 6 日**

# RNN Review : RNN网络在不同领域的应用

## （一）常用结构框架



（图中，红色代表输入，蓝色代表输出，绿色代表隐状态）

1. 一对一，DNN的模式，定长输入产生定长的输出，典型应用如图像分类等任务；
2. 一对多，变长输出，DNN不能处理，可应用于图像标注等任务；
3. 多对一，变长输入，定长输出，可应用于文本情感分析等；
4. 多对多，RNN里面的encoder-decoder架构，在图像标注，语言翻译等领域应用广泛；
5. 多对多，可用于视频逐帧标注等；

以上结构中，在时间方向，长度是任意可变的，使之能够有效地处理序列数据。

## （二）RNN在自然语言处理（NLP）中的应用

### Sequence to Sequence Learning with Neural Networks

#### 作者机构

Ilya Sutskever，Oriol Vinyals，Quoc V. Le. Google. 2014

#### 针对的问题

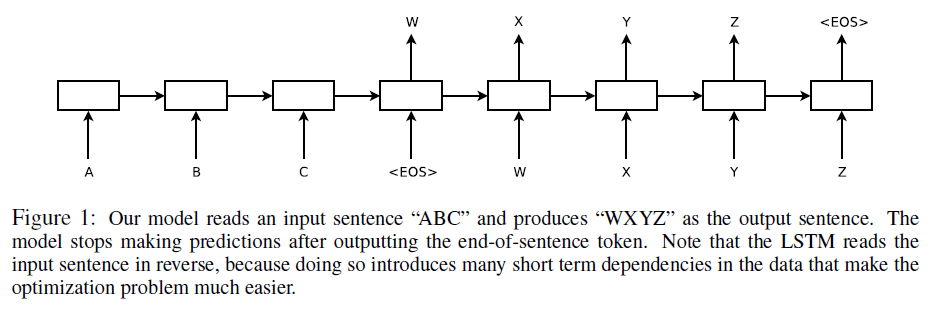
一般前馈深度网络在处理序列数据时存在很大的局限性，无法有效地建模前后数据间的关系。利用RNN对序列数据处理的独特优势来建模机器翻译。

#### 创新点

提出了RNN里的encoder-decoder的架构。为了跨越不同语言之间的鸿沟，首先利用编码器将源语言编码，然后利用解码器从编码后的数据里翻译出目标语言。

#### 方法

当输入序列和输出序列存在不同的长度，而且他们之间不存在直接性的关系时，利用RNN来建模是一件很困难的事情。一个简单的策略就是先利用一个RNN将输入进行编码，然后再利用另一个RNN将编码后的数据进行解码。



1. 解码和编码使用不同的LSTM结构，参数不共享；
2. 试验中，采用深层LSTM（4层）比单层LSTM效果好；
3. 训练时，发现将源语言的顺序翻转，加入到训练集中，能够更快地收敛。（个人理解，类似于图像处理中镜像，缩放等data augmentation方法）

#### 评价

学术价值在于encoder-decoder结构的提出，工业价值在于将语言翻译推进到一个新的台阶(state-of-art)。类似的结构后来被用于语音识别（Towards End-to-End Speech Recognition with Recurrent Neural Networks——Google Deepmind, Alex Graves），同样达到了很好的效果。

## （三）RNN在图像以及视频分析领域的应用

### Show, Attend and Tell: Neural Image Caption Generation with Visual Attention

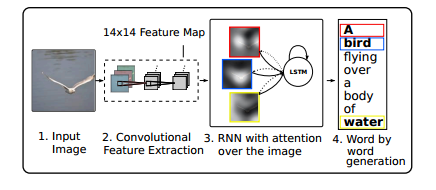
#### 作者机构

[Kelvin Xu](https://arxiv.org/find/cs/1/au:+Xu_K/0/1/0/all/0/1), [Jimmy Ba](https://arxiv.org/find/cs/1/au:+Ba_J/0/1/0/all/0/1), [Ryan Kiros](https://arxiv.org/find/cs/1/au:+Kiros_R/0/1/0/all/0/1), [Kyunghyun Cho](https://arxiv.org/find/cs/1/au:+Cho_K/0/1/0/all/0/1), [Aaron Courville](https://arxiv.org/find/cs/1/au:+Courville_A/0/1/0/all/0/1), [Ruslan Salakhutdinov](https://arxiv.org/find/cs/1/au:+Salakhutdinov_R/0/1/0/all/0/1), [Richard Zemel](https://arxiv.org/find/cs/1/au:+Zemel_R/0/1/0/all/0/1), [Yoshua Bengio](https://arxiv.org/find/cs/1/au:+Bengio_Y/0/1/0/all/0/1). 19 Apr 2016.

#### 针对问题和创新点

给定图像，生成图像的描述。引入attention mechanism（聚焦机制）用于图像描述生成，每次由图像的焦点patch生成新的描述词。

#### 方法



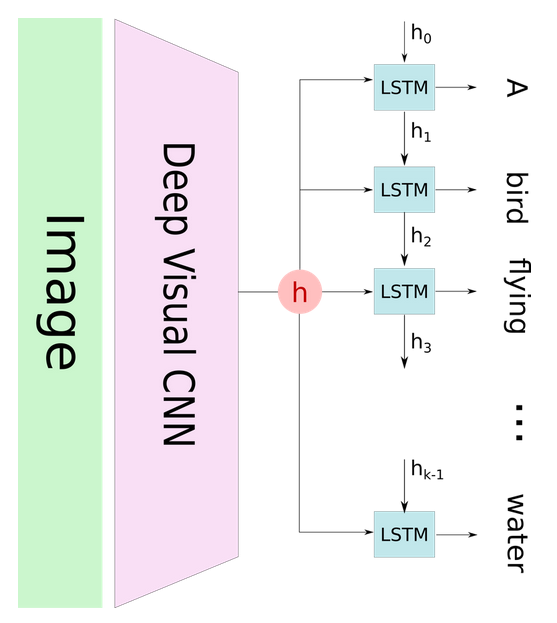
**Attention mechanism**

<https://blog.heuritech.com/2016/01/20/attention-mechanism/>

A wealth of results in the past few yearsuggest that visual structure can be better captured by a sequence of partial glimpses, or foveations, than by a single sweep through the entire image (Larochelle & Hinton,2010; Denil et al., 2012; Tang et al., 2013; Ranzato, 2014Zheng et al., 2014; Mnih et al., 2014; Ba et al., 2014; Sermanet et al., 2014).

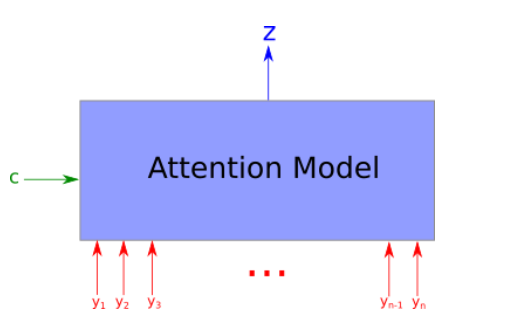
动物视觉系统观察物体时倾向于集中观察主要的部分以获取足够的观察对象信息。

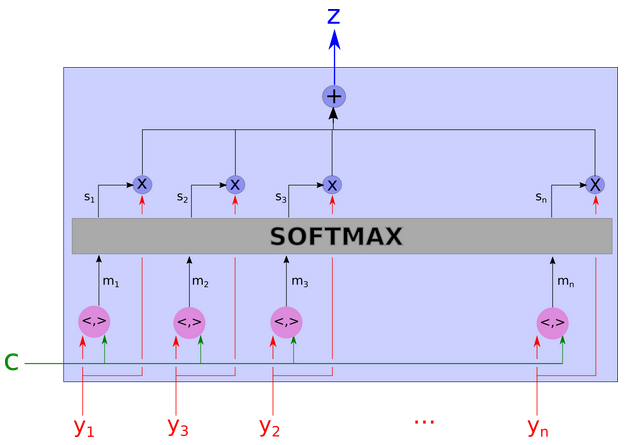
**聚焦机制用于图像描述**



上图是通常的解决图像描述生成问题的一般模型架构，首先由CNN提取图像信息，将图像进行encode，然后将图像特征（通常由一个向量表示）送入LSTM，由LSTM逐个产生描述单词。

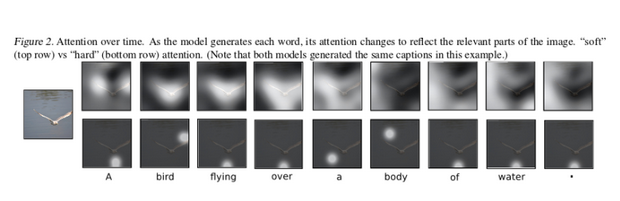
以上模型存在的一个缺陷是，每次生成新的描述词时，一个词应该集中于描述图像的一个object、patch或者region，所以在每个节点上将整个图作为输入是存在信息冗余的，应该找到一种机制使得可以有效地降低这种冗余，使得生成的描述更精确。

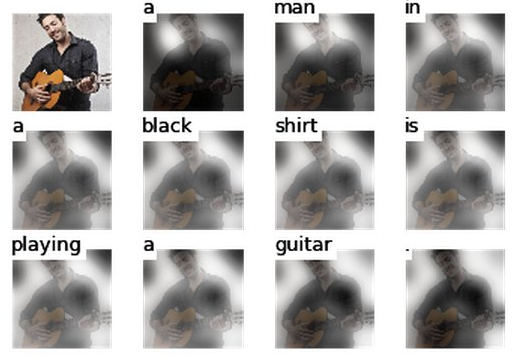




**Attention机制在图像描述上的实现**：

将输入图像分成不同的patch（上图中的y），对每个patch提取CNN特征，然后将其与上一时刻的隐状态输出C一起作一定运算后(比如内积)激活，再经过softmax层（当只有一个变量比较大的时候，就相当于在作argmax操作）后作为新的attention图像输出。







#### 评价

该方法将attention mechanism和encoder-decoder架构运用到图像描述的生成上来。相比每一步直接以整图输入产生描述单词更精确，能有效地去除冗余信息，提高描述的准确性，在几个常用数据集上都达到了state-of-art的效果。但是同时一定程度地增加了模型的复杂性，有更多的参数需要训练。需要更多的训练技巧（DropOut，选择不同梯度下降算法等）。

## （四）RNN在目标行进轨迹预测、目标跟踪上的应用

### **Deep Tracking: Seeing Beyond Seeing Using Recurrent Neural Networks**

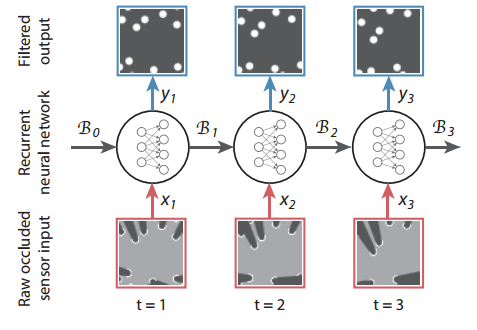
#### 作者机构

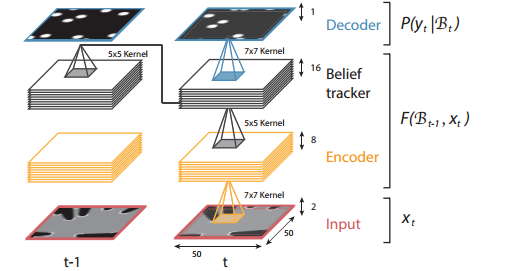
[Peter Ondruska](https://arxiv.org/find/cs/1/au:+Ondruska_P/0/1/0/all/0/1), [Ingmar Posner](https://arxiv.org/find/cs/1/au:+Posner_I/0/1/0/all/0/1)，Mobile Robotics Group, University of Oxford, United Kingdom，AAAI-16 conference, February 12-17, 2016

#### 针对问题

利用RNN建立端到端的目标跟踪方法，做到实时地从原始的传感器信号输入产生整个环境状态的预测，能有效地处理遮挡问题。

#### 方法





**训练：**

1. 监督学习方式：

有每个时刻环境的真实状态yi（通过增设传感器）。

目标函数：



在监督学习框架下，将当前的传感器观测和上一时刻的隐状态作为输入，目标是预测当前的真实环境状态。这是此类问题的一般解决思路，然而在实际中，真实环境状态难以获得，通过增设传感器的方式虽然可以解决，但是代价巨大，不实际。

因此，转而寻求非监督的学习方式，即只从一系列的传感器观测中学习预测状态的模型。

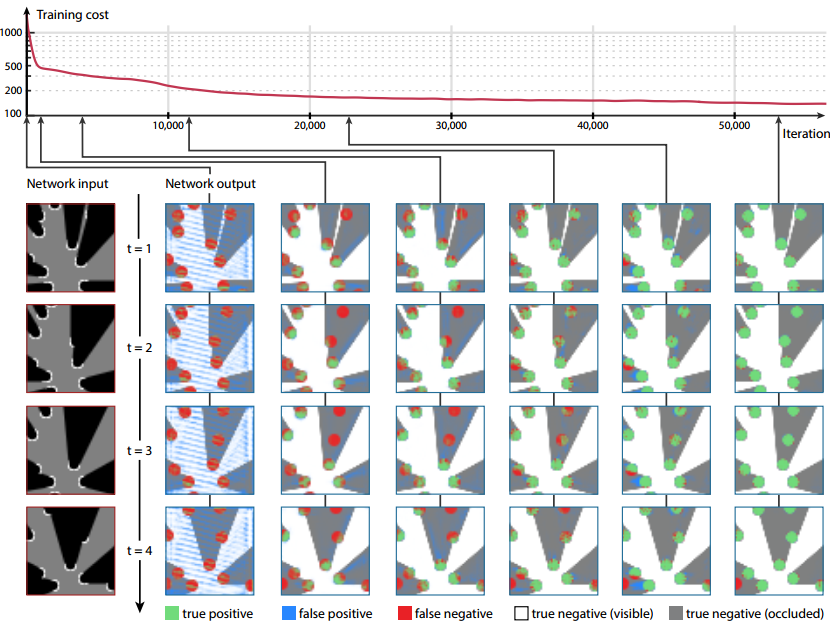
1. 非监督学习方式

与监督学习方式的不同点在于：

1. 输入数据只有传感器观测x，没有真实状态y;
2. 预测输出不再是当前的真实环境状态，变为若干步之后的观测状态；

通过这样的设定，可以使得模型区学习到预测若干步后的观测，从而具备轨迹推测和遮挡推测的能力。

**结果：**



## （五）RNN用于图像生成

### **DRAW: A Recurrent Neural Network For Image Generation**

#### 作者机构

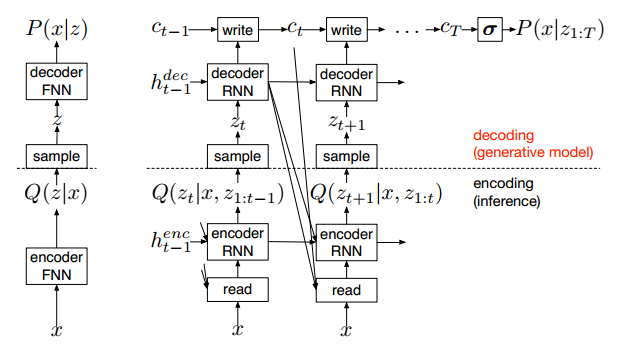
Karol Gregor，Ivo Danihelka，Alex Graves，Google DeepMind，2015.

#### 针对问题和创新点

RNN用作图像生成；将attention mechanism和VAE(variational auto-encoder)作结合。

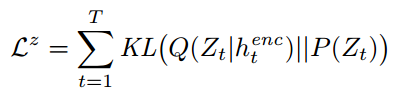
#### 方法

**GAN and VAE**



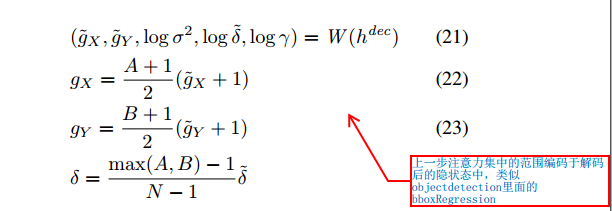
**损失函数包括两个部分：**



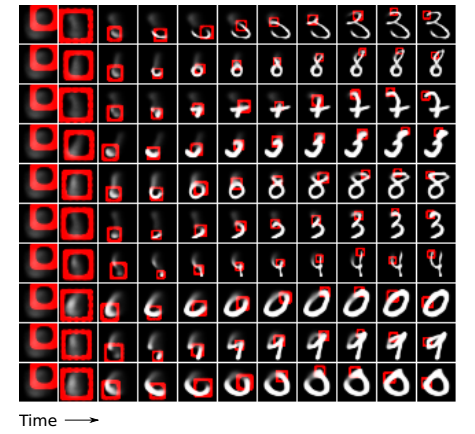


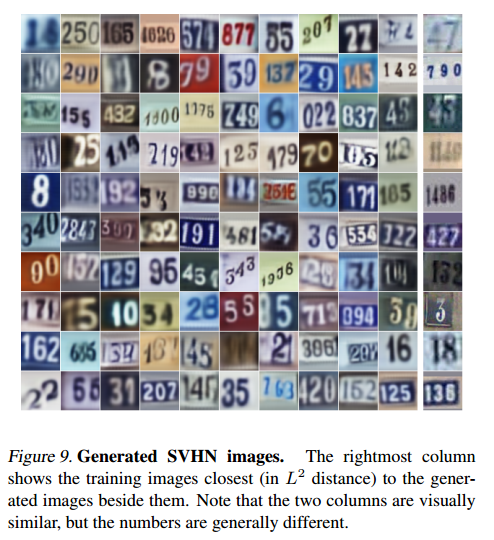
**Attention和Without Attention的比较:**

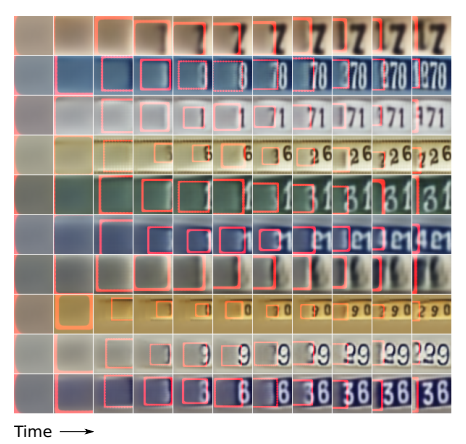
其attention机制的运用跟前面文章中所述稍有区别，本文中采用的方法类似于RCNN中使用的BoundingBox回归，它从上一时刻的隐变量中直接学习控制下一时刻焦点窗口的五个参数。

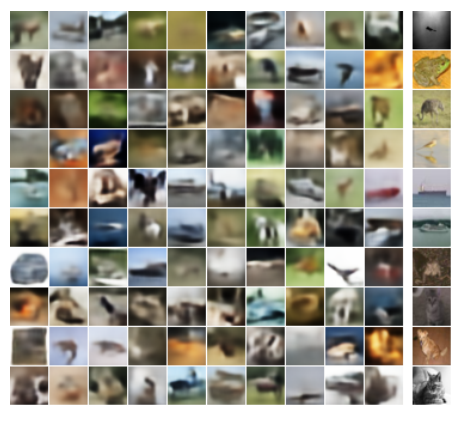


实验结果证明，增加Attention机制后效果有明显的提升。









#### 评价

利用VAN对隐变量的分布加以限制可以获得更为有效的隐变量特征。在其他针对特定的学习任务时，也可以考虑增加先验知识。通过改变模型结构或者添加目标函数限制得到更好的描述模型。