* 1. **简介**
     1. **残差网络和共享参数的残差网络**

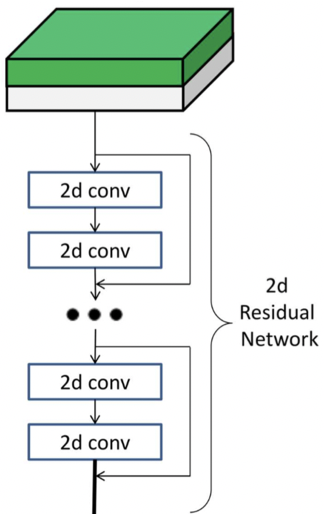


图 1. 深度残差网络

增加网络的宽度和深度可以很好的提高网络的表现，深的网络一般都比浅的的网络效果好，例如 VGG，该网络就是在AlexNex的基础上通过增加网络深度大幅度提高了网络性能4。

对于原来的网络，如果简单地增加深度，会导致梯度弥散或梯度爆炸，取决于网络结构和所选用的激活函数的特性。深度网络不能很简单地被很好地优化。如果深层网络的后面那些层是恒等映射，那么模型就退化为一个浅层网络。那当前要解决的就是学习恒等映射函数了。 但是直接让一些层去拟合一个潜在的恒等映射函数  ，比较困难，这可能就是深层网络难以训练的原因。但是，如果把网络设计为  ,如下图。我们可以转换为学习一个残差函数  。 只要  ，就构成了一个恒等映射  。 而且，拟合残差肯定更加容易。

根据多层的神经网络理论上可以拟合任意函数，那么可以利用一些层来拟合函数。问题是直接拟合  还是残差函数，拟合残差函数  更简单。虽然理论上两者都能得到近似拟合，但是后者学习起来显然更容易。

* + 1. **共享参数的残差网络：**

随着ResNet的广泛应用，研究者倾向于逐渐加深网路层数以获得更好的模型表现。但是很深的ResNet往往带来OOM 的问题，在一些应用场景下很难训练甚至inference。

将RNN按照时间步展开，很像是一个权重完全共享的Res，如图：

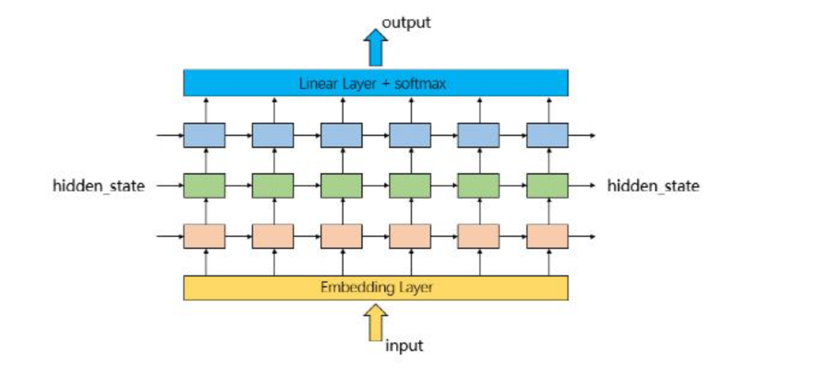


图2: RNN的时间步展开

无论是基本的res block 还是bottle neck形式，都可以在某种程度上被认为是：

第一个直接处理输入的conv是 “specific to this block”，是对输入最直接的abstraction，

第二个conv则是对第一个conv得到的结果在空间上的再一次整合。因此，在同一个stage中，第二个conv 在某种程度上是冗余的。如图所示。这里的stage是指两次down sampling （conv with stride larger than 1 or pooling）之间的所有block。

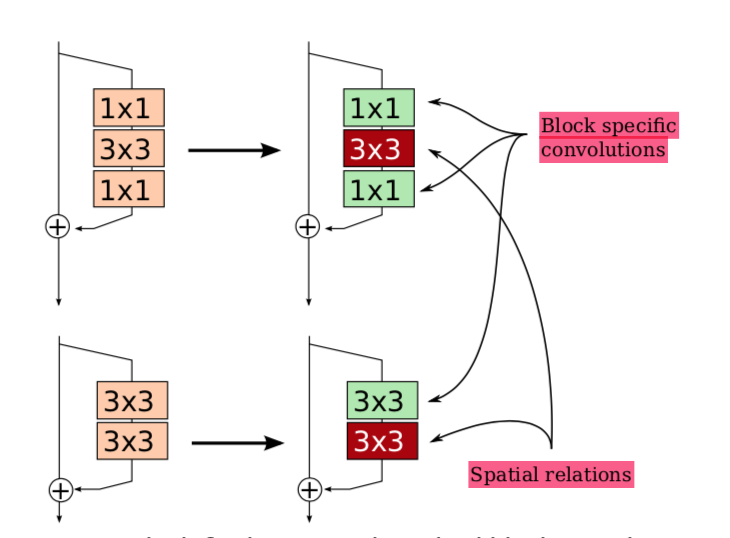


图3 ： 共享参数的残差网络的基本单元图示。

* + 1. **长短期记忆网络**

在序列相关的问题，如时间序列，文字序列，氨基酸、核苷酸序列等问题 时，我们不单单要考虑当前序列的情况，历史信息也是我们需要注意的地方。在 RNN 中，长短期记忆网络的应用较为广泛。长短期记忆 网络(long short-term memory, LSTM)的节点由一个单元、三个门组成。三个门包括:输入门、输出门和遗忘门组成。单元在任意时间的状态值由三个门通过各自的门控机制共同决定。