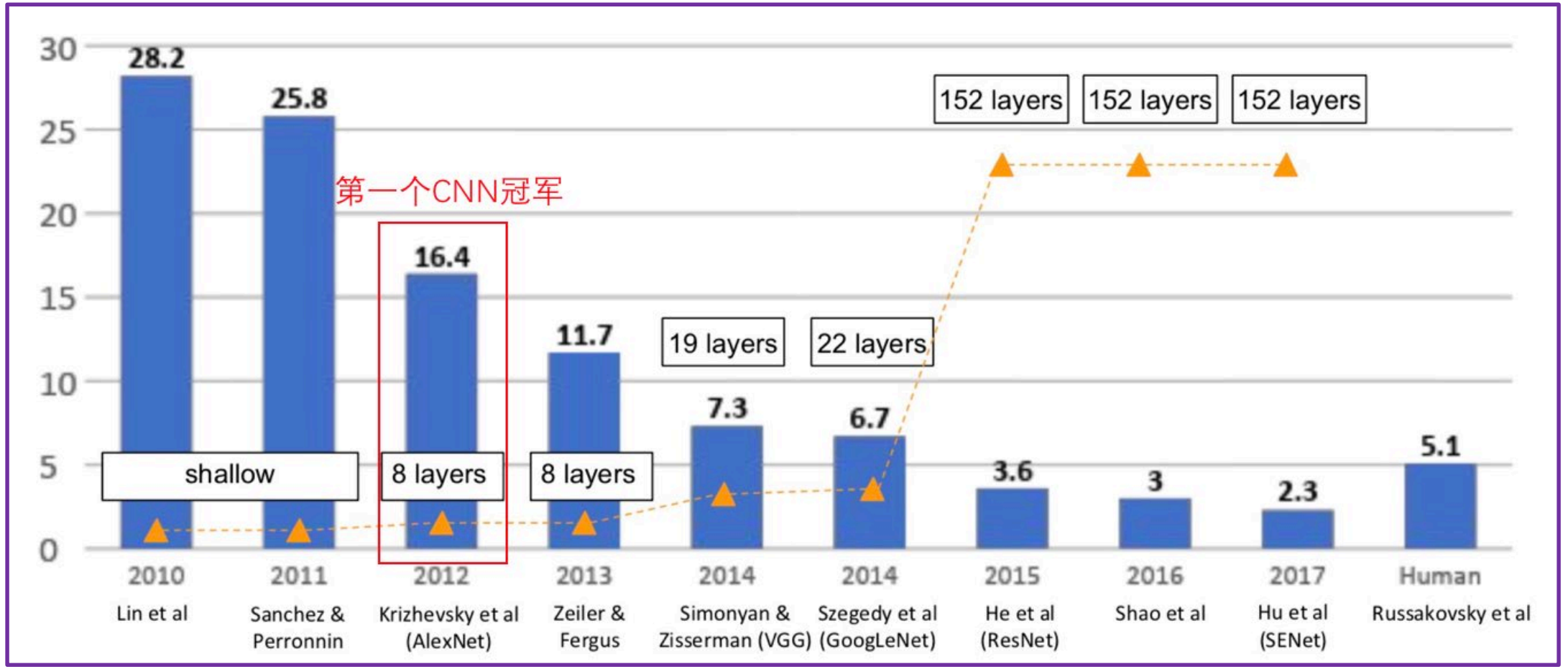
计算机视觉
Computer Vision

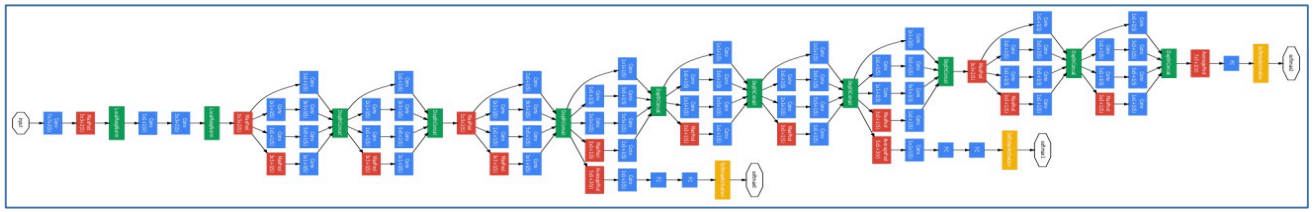
Lecture 10: 循环神经网络

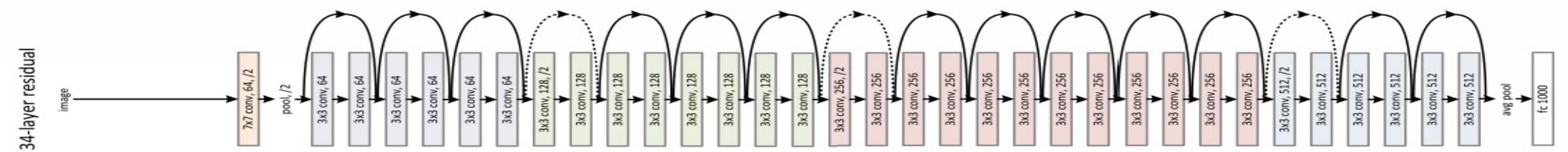
L09：卷积神经网络结构的演化



L09：卷积神经网络结构的演化

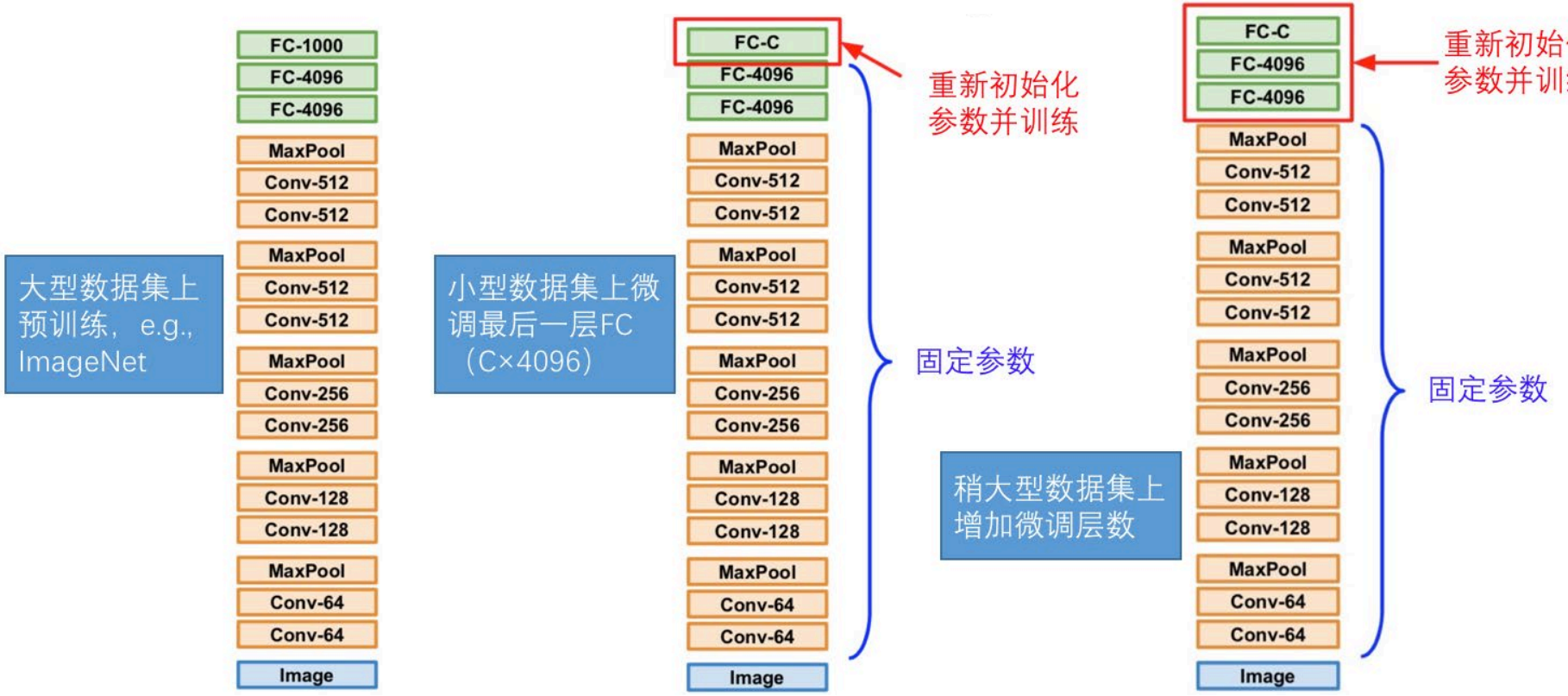
AlexNet→VGG→Inception→ResNet





L09：卷积神经网络结构的演化

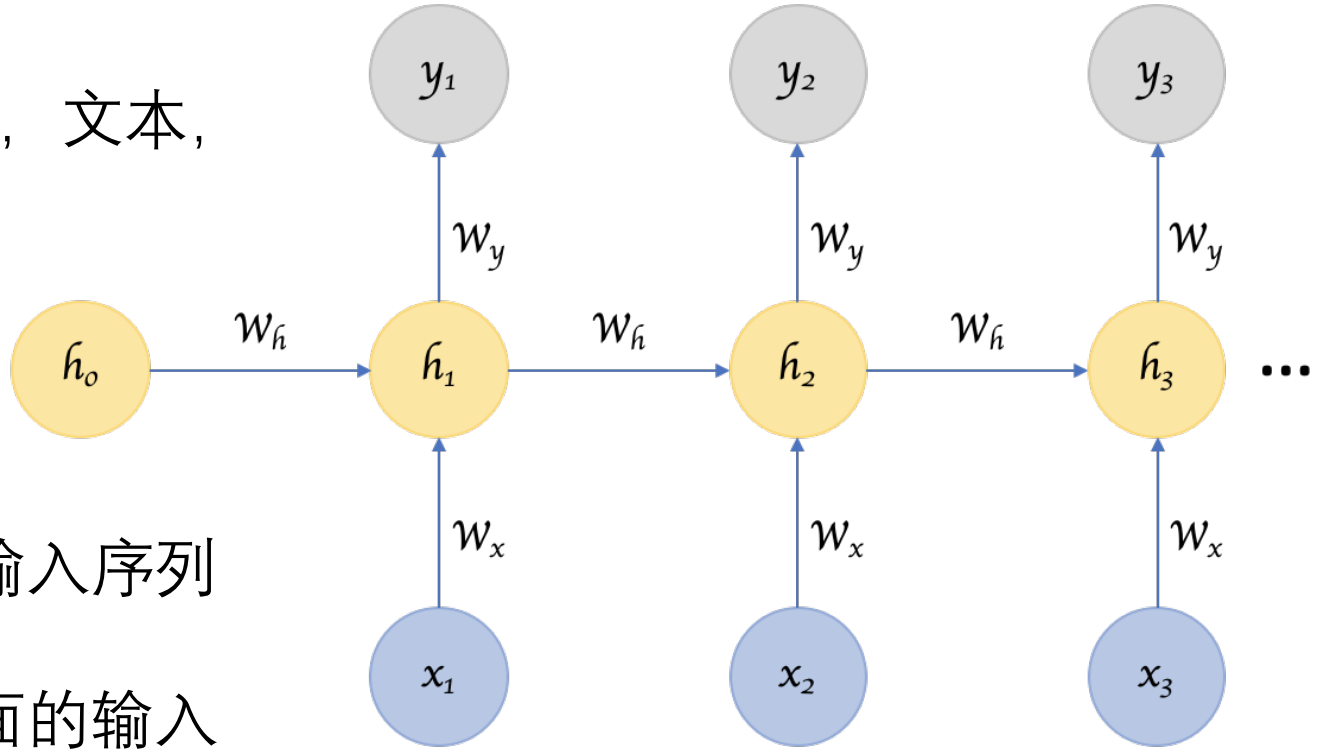
迁移学习



循环神经网络（Recurrent Neural Networks）

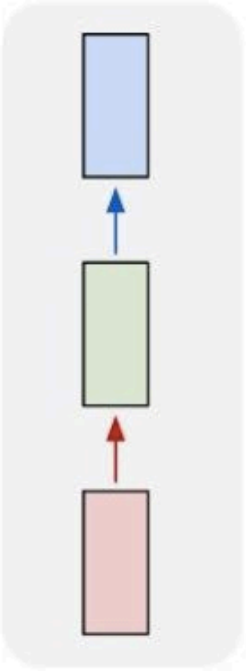
- Motivation
  - ✓很多数据都是序列数据，e.g.，文本，股票，交通轨迹
  - ✓序列没有固定的长度

- Solution
  - ✓采用重复的结构单元，每次输入序列里的一个数据
  - ✓后面的结构单元能够记忆前面的输入信息



RNNs：结构变化

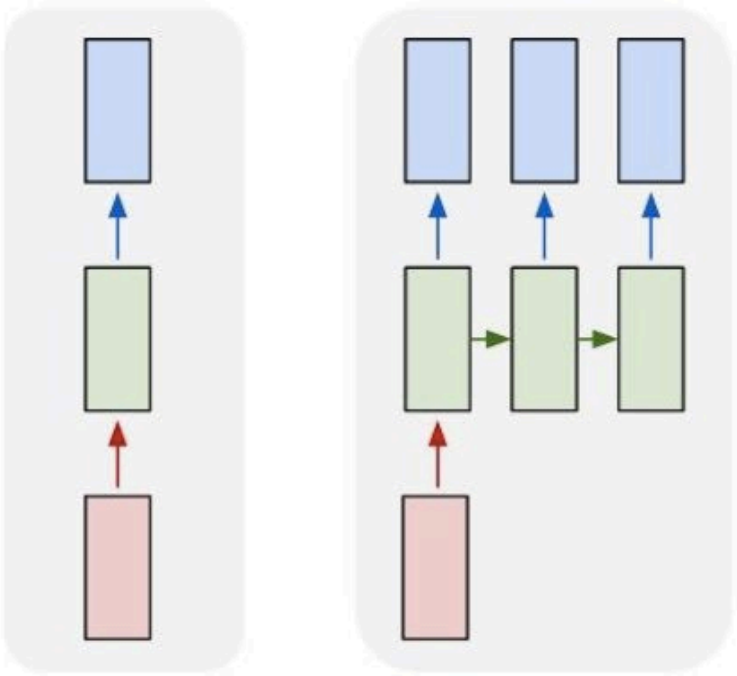
一对一



Vanilla NN

RNNs：结构变化

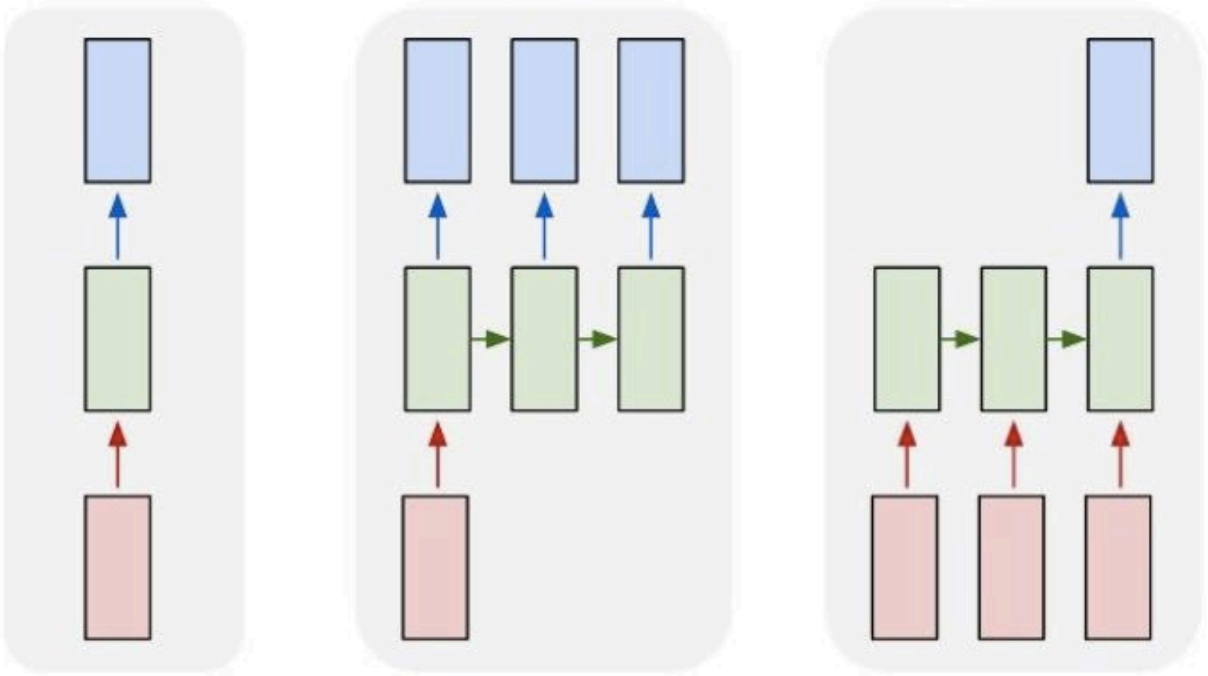
一对一　　一对多



图像→文本

RNNs：结构变化

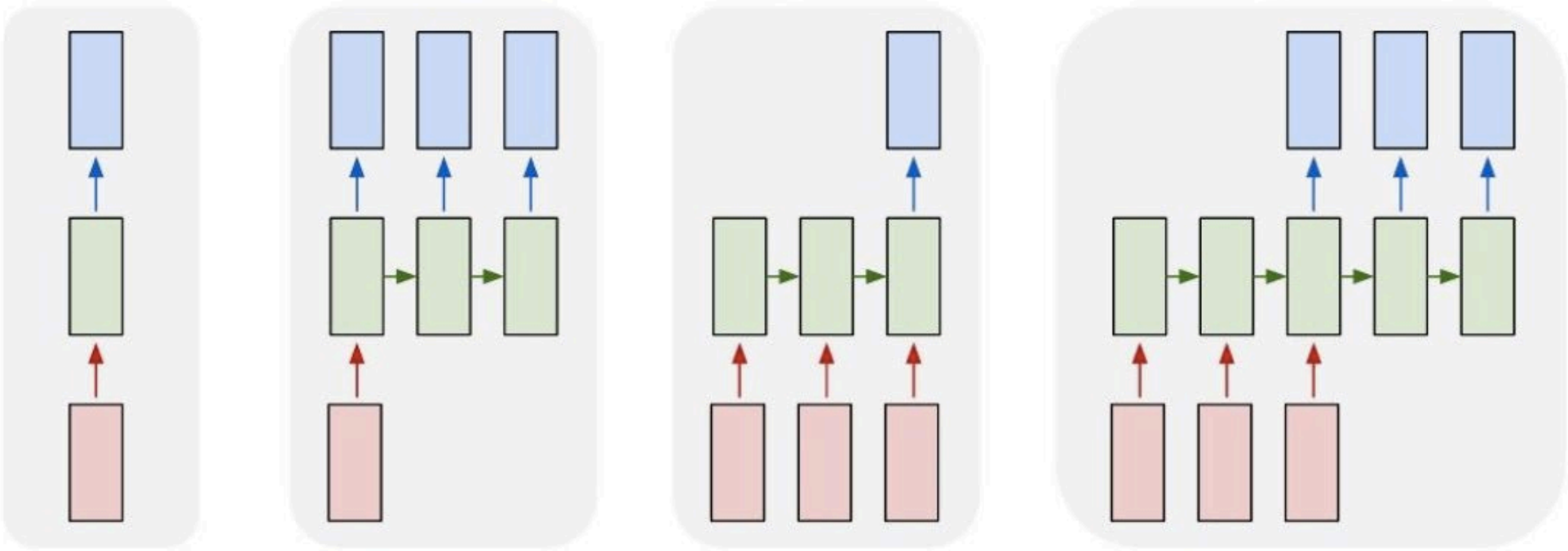
一对一　　一对多　　多对一



✓ 文本→图像
✓ 情感分析

RNNs：结构变化

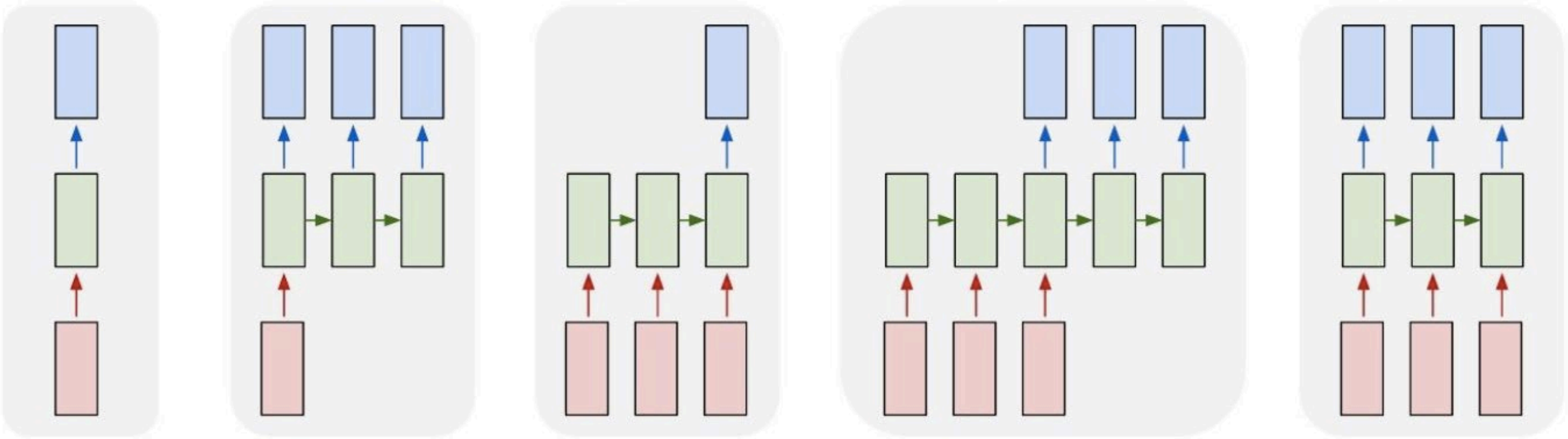
一对一　　一对多　　多对一　　多对多



- ✓ 机器翻译
- ✓ 文本改写

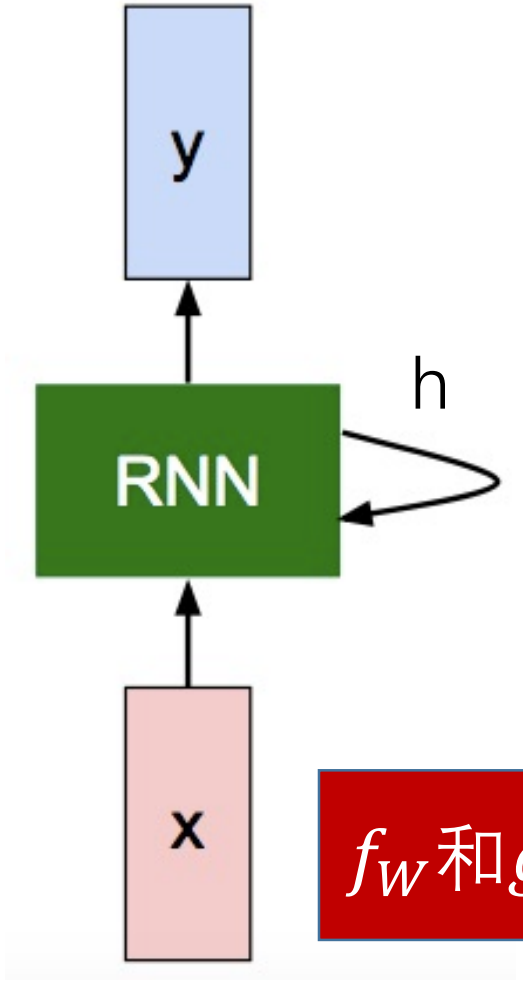
RNNs：结构变化

一对一　一对多　多对一　多对多　多对多



- ✓ 视频预测
- ✓ 股票分析

循环神经网络



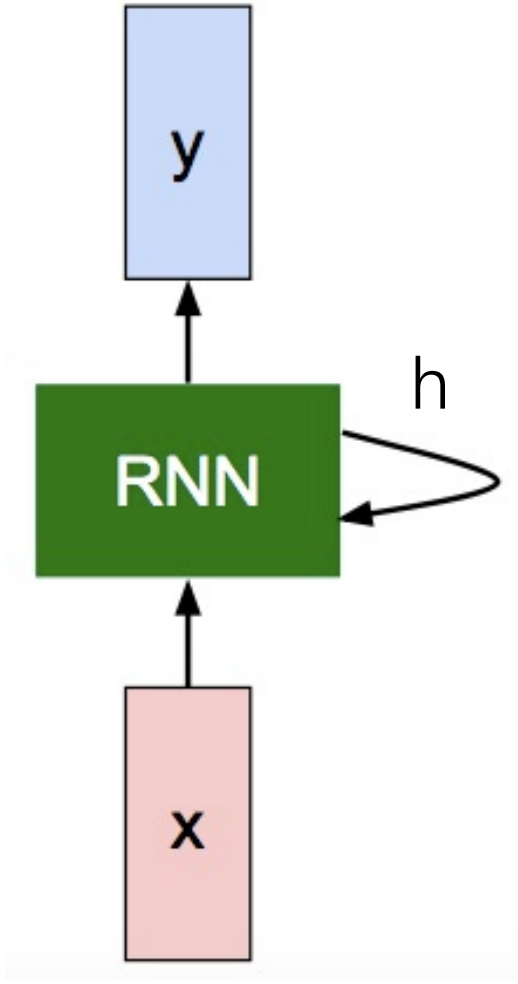
- 令t时刻的输入为x_t，输出为y_t
- 引入一个隐藏状态向量（hidden state vector）h
- 使用相同的结构顺序处理输入数据，并不断更新h
- t时刻的隐藏状态向量记为h_t
- 利用h_t计算t时刻的输出y_t

$$h_t = f_W(h_{t-1}, x_t)$$

$$y_t = g_W(h_t)$$

f_W和g_W对所有时刻相同

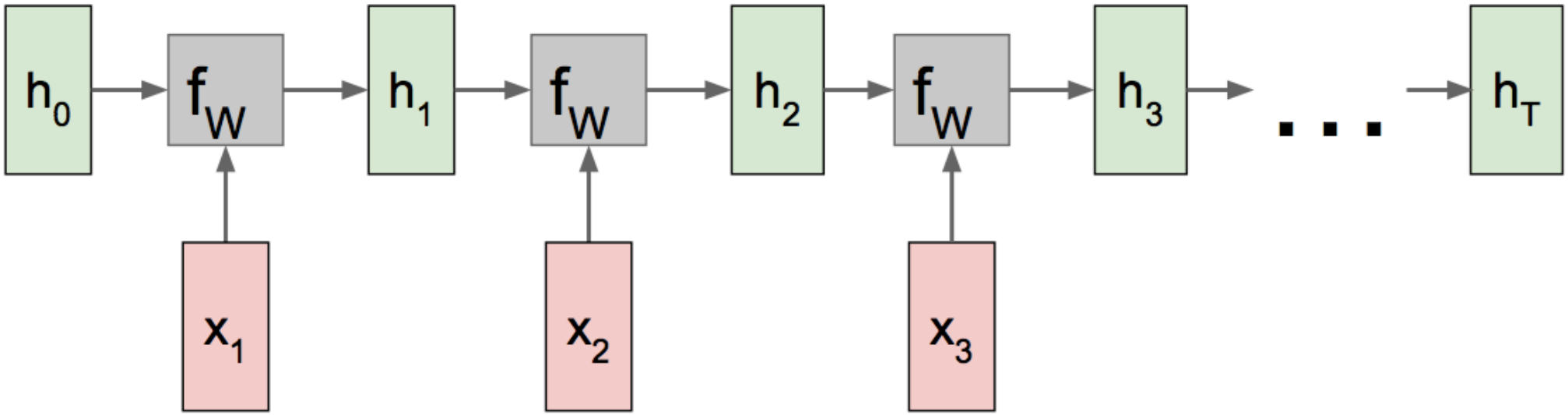
最简单的RNN（Vanilla RNN）



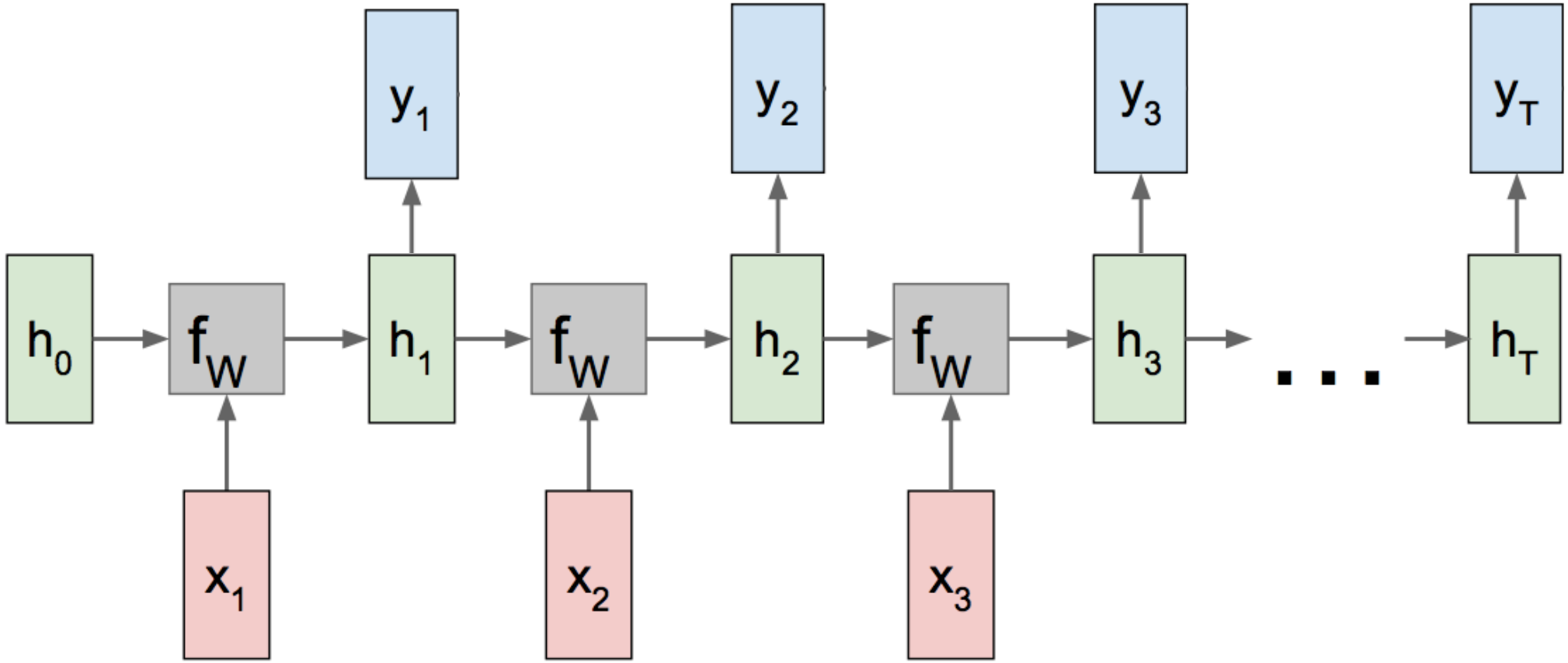
$$h_t = \tanh(W_{hh}h_{t-1} + W_{xh}x_t)$$

$$y_t = W_{hy}h_t$$

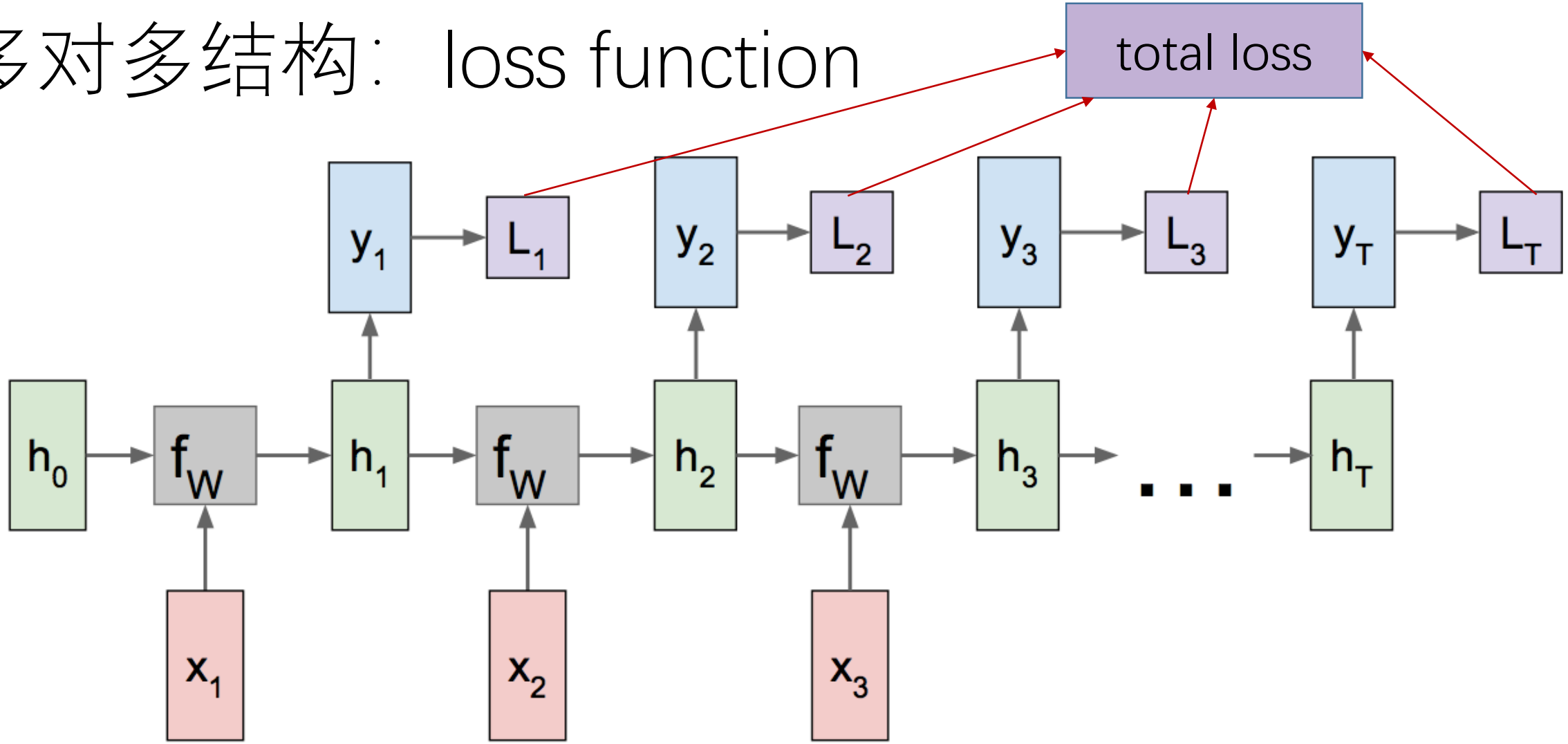
RNN：计算图



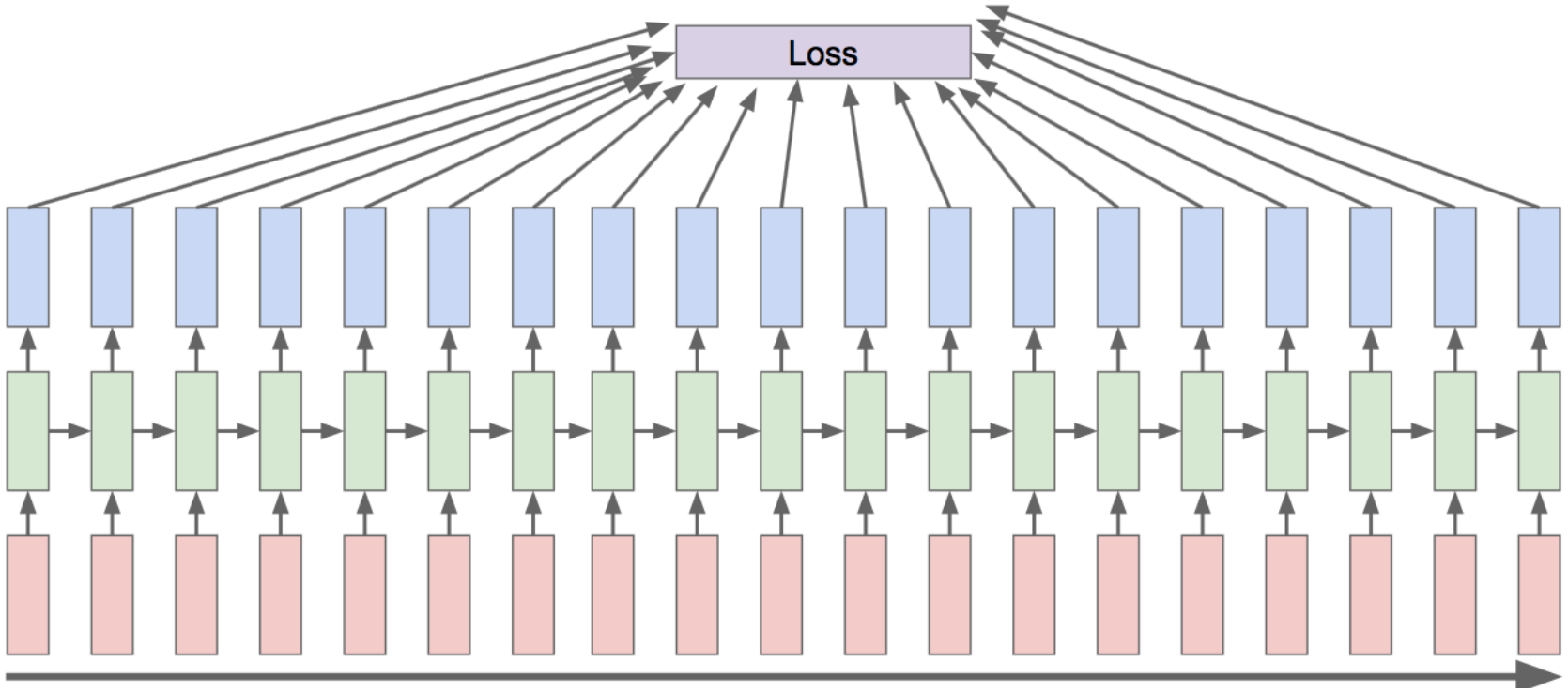
多对多结构



多对多结构：loss function



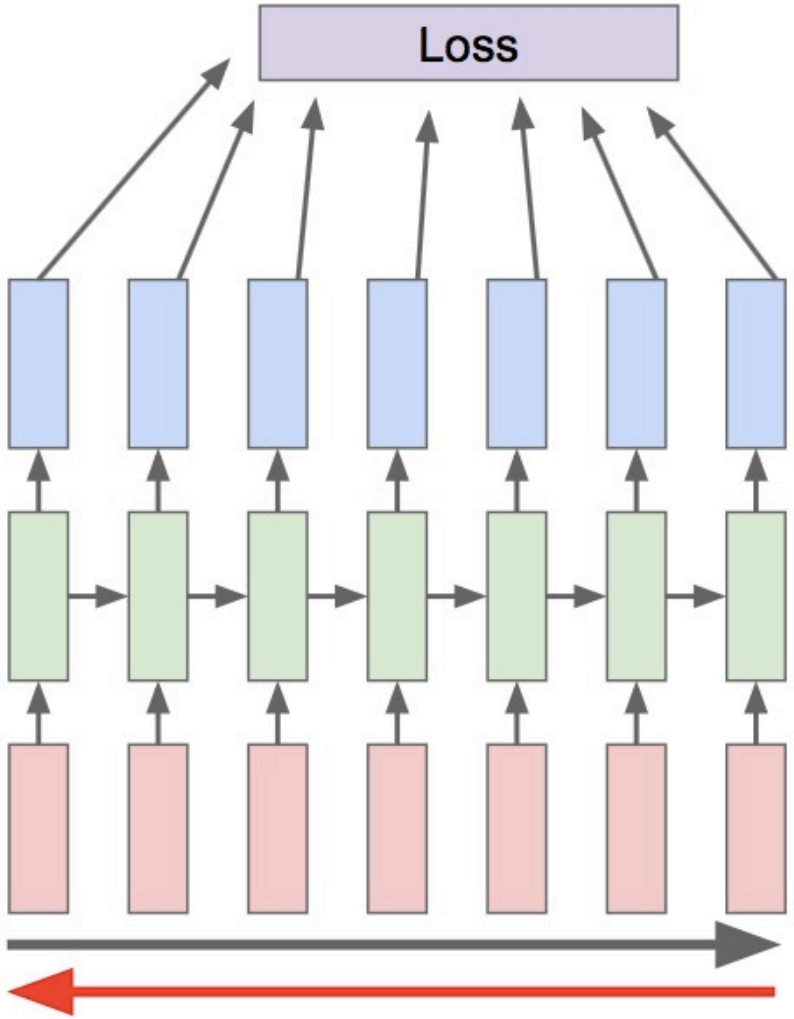
BPTT: backpropagation through time



- 反向传递Loss，并计算每一时刻的∇W_t
- 累加所有的∇W_t得到∇W
- 更新W_{hh}, W_{xh}, W_{hy}

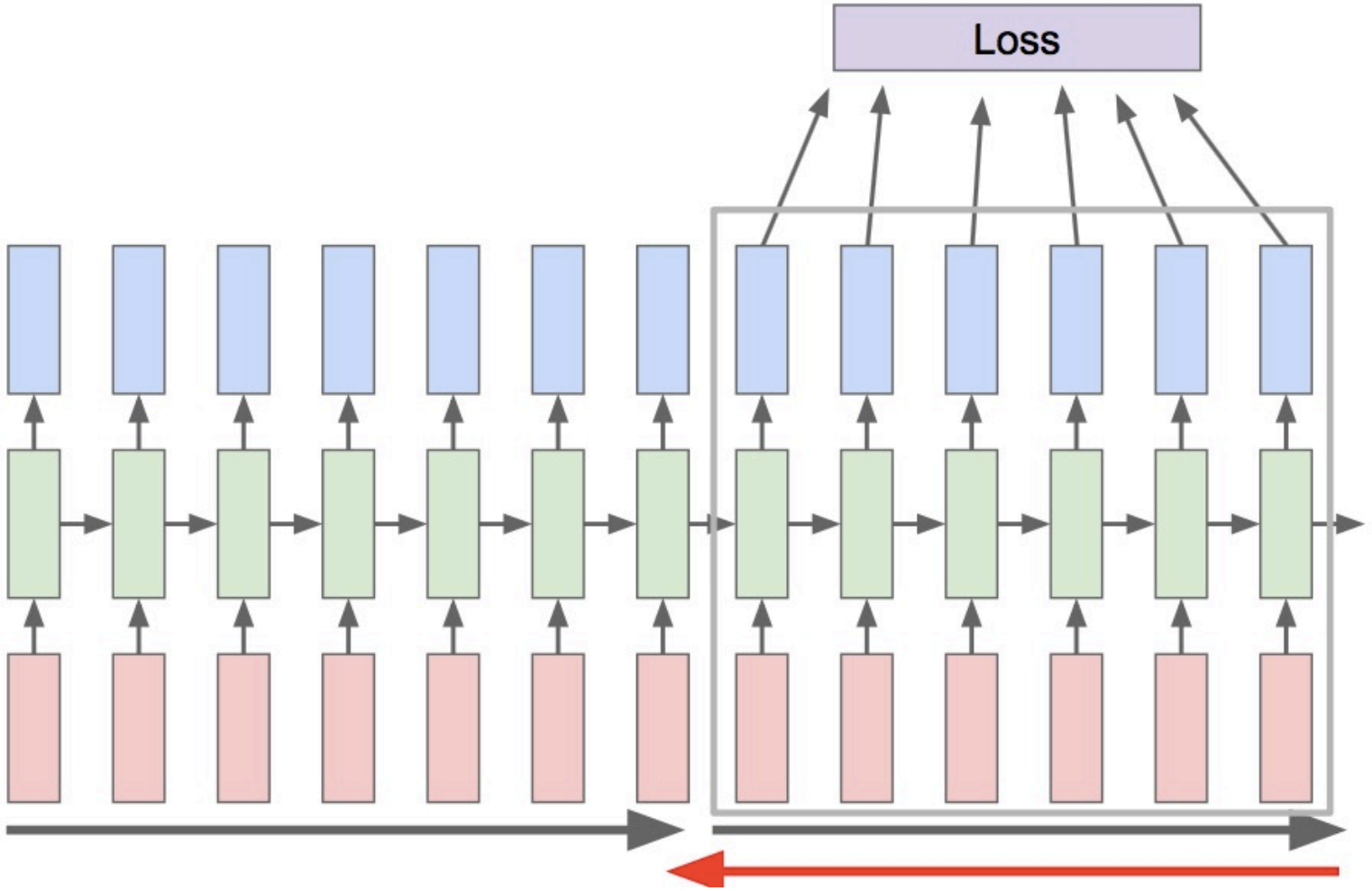
Werbos PJ. Generalization of backpropagation with application to a recurrent gas market model. Neural networks. 1988.

In practice：truncated BPTT（TBPTT）

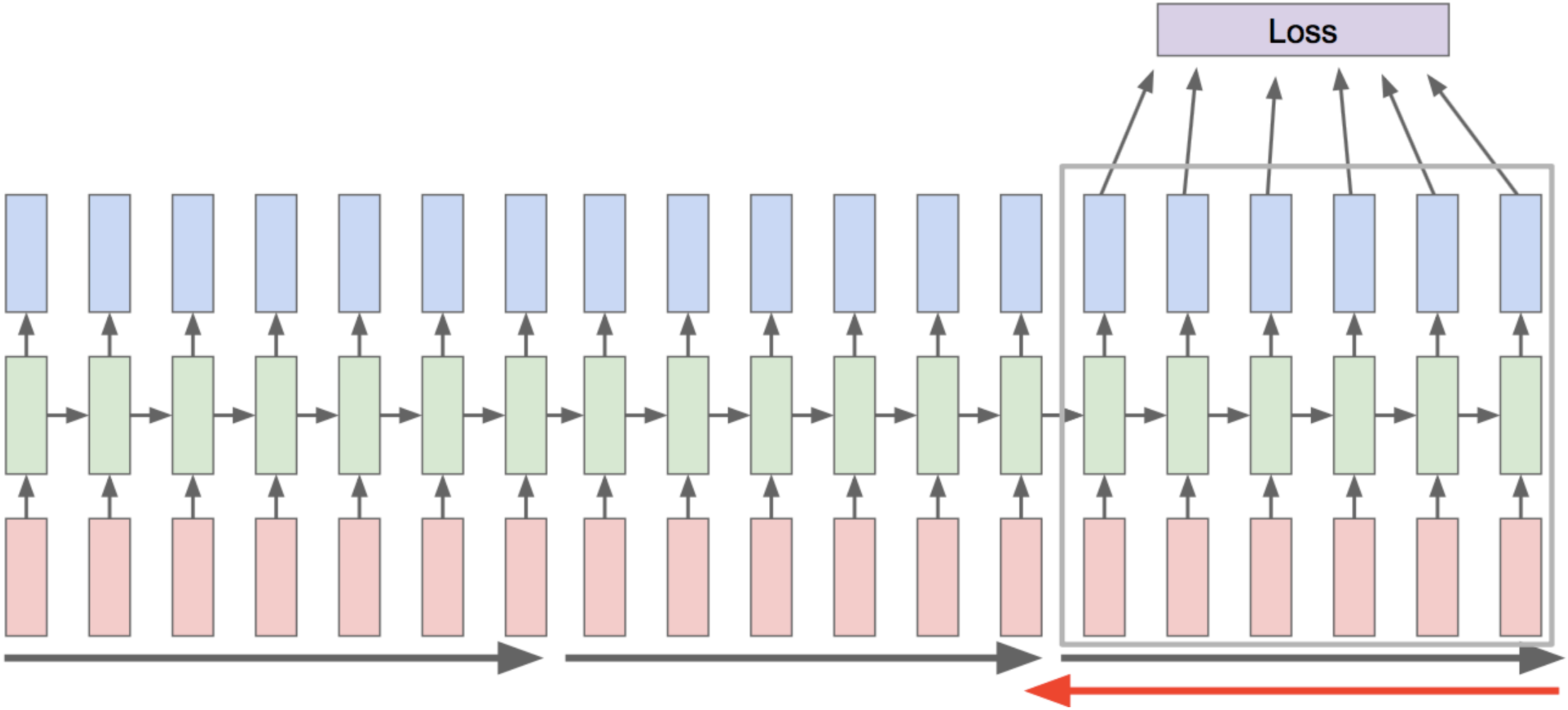


- 序列长度为n
- 周期性地，
  - ✓前向计算k_1步
  - ✓反向传播k_2步
- 可以设置$k_1 = k_2 < n$

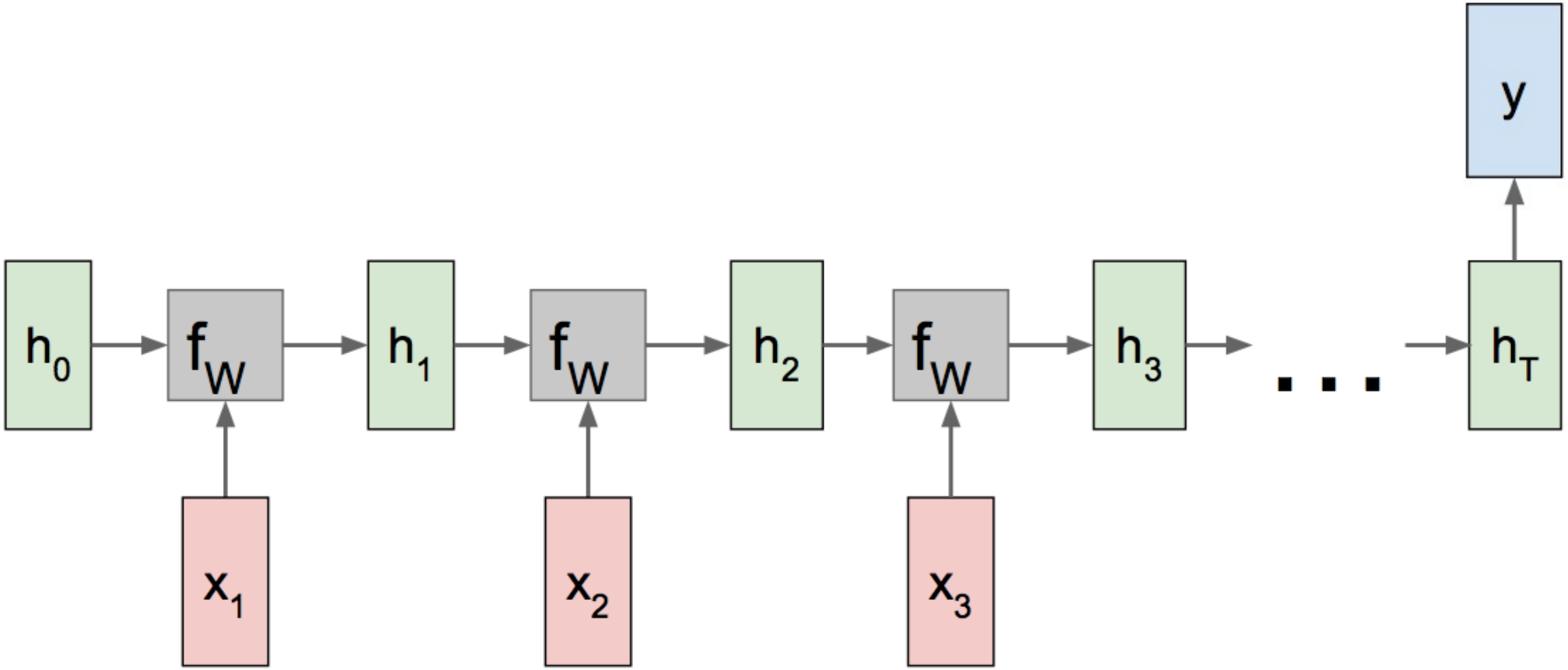
In practice：truncated BPTT（TBPTT）



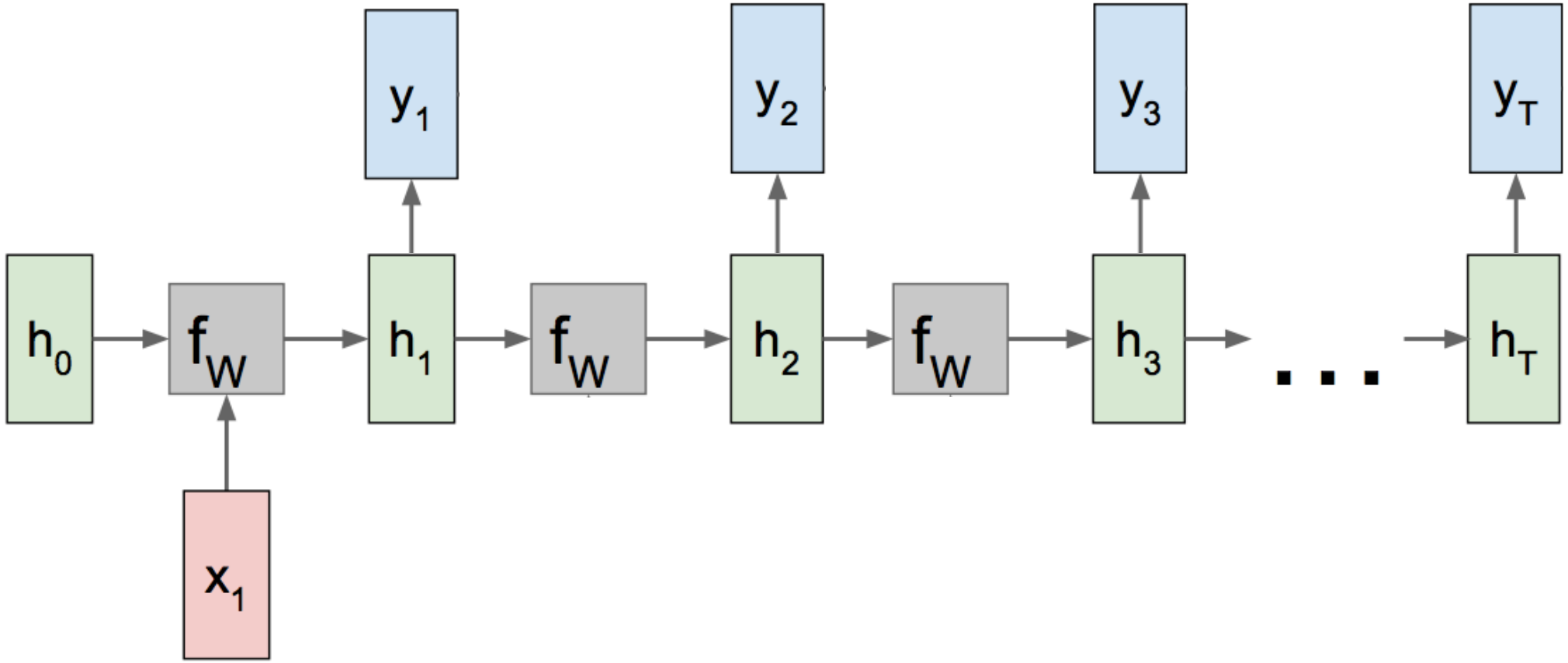
In practice：truncated BPTT（TBPTT）



多对一结构

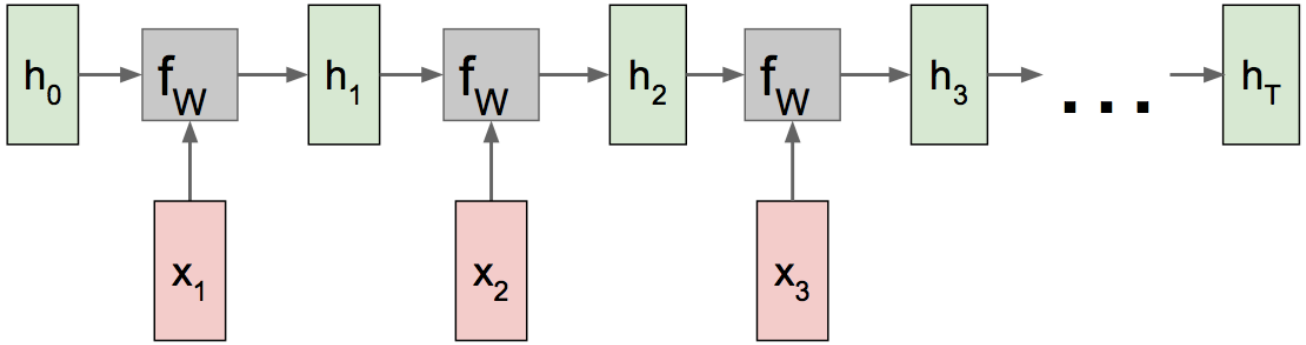


一对多结构



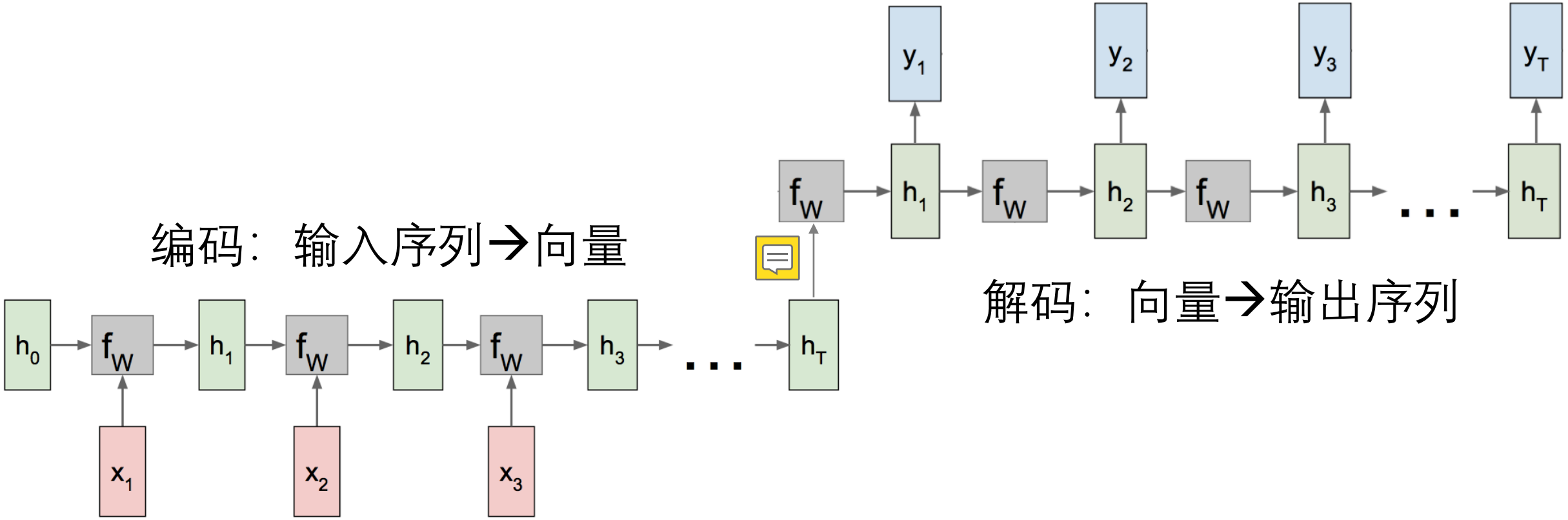
特殊多对多：Sequence-to-sequence

编码：输入序列→向量



Sutskever I, Vinyals O, Le QV. Sequence to sequence learning with neural networks. NIPS 2014.

特殊多对多：Sequence-to-sequence



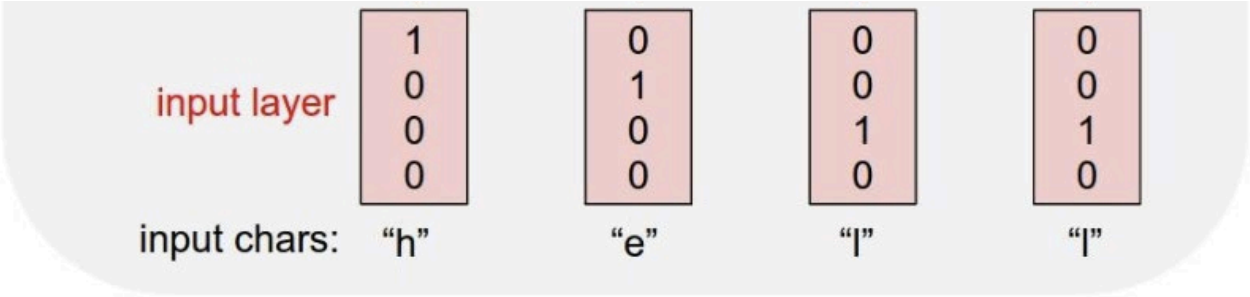
Sutskever I, Vinyals O, Le QV. Sequence to sequence learning with neural networks. NIPS 2014.

RNN实战：训练语言模型

- Character-level Language Model

- 语料库：hello

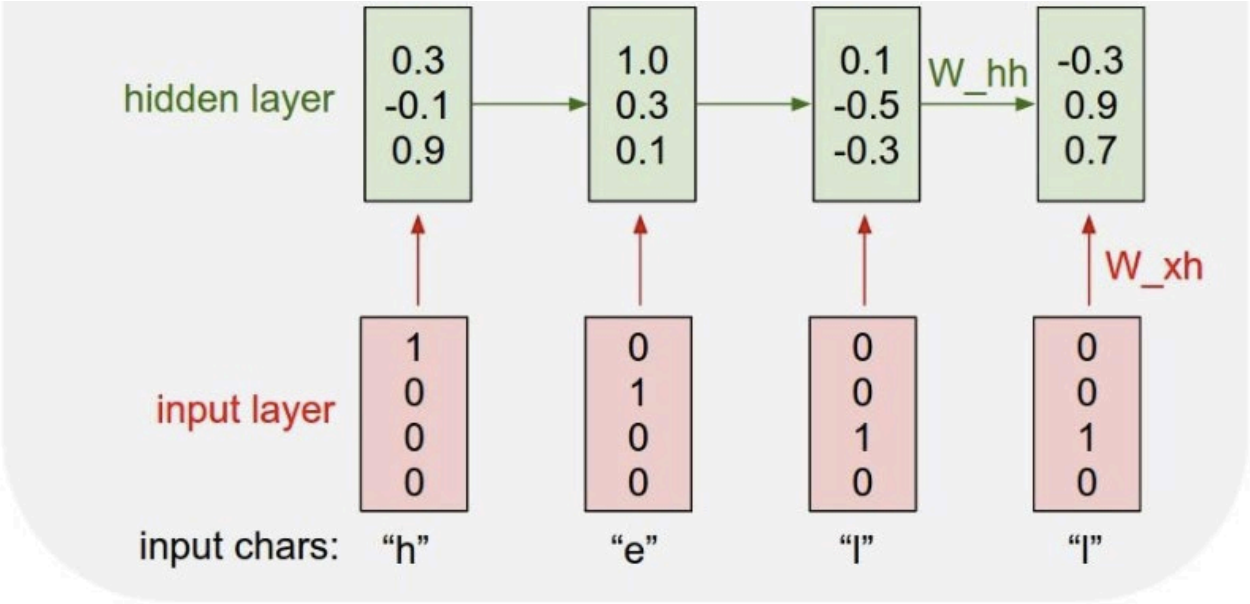
- Vocabulary：[h, e, l, o]



RNN实战：训练语言模型

- Character-level Language Model
- 语料库：hello
- Vocabulary：[h, e, l, o]

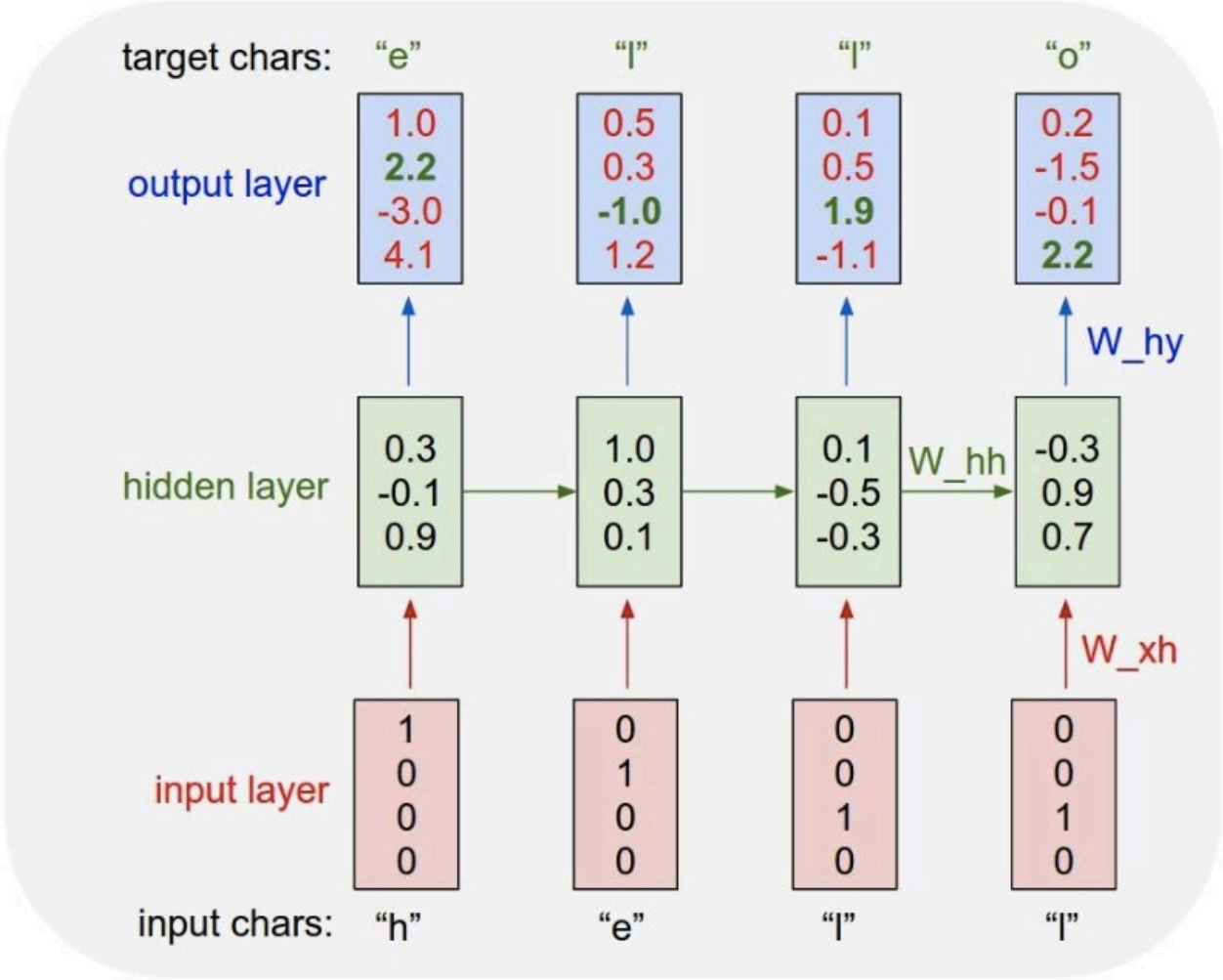
$$h_t = \tanh(W_{hh}h_{t-1} + W_{xh}x_t)$$



RNN实战：训练语言模型

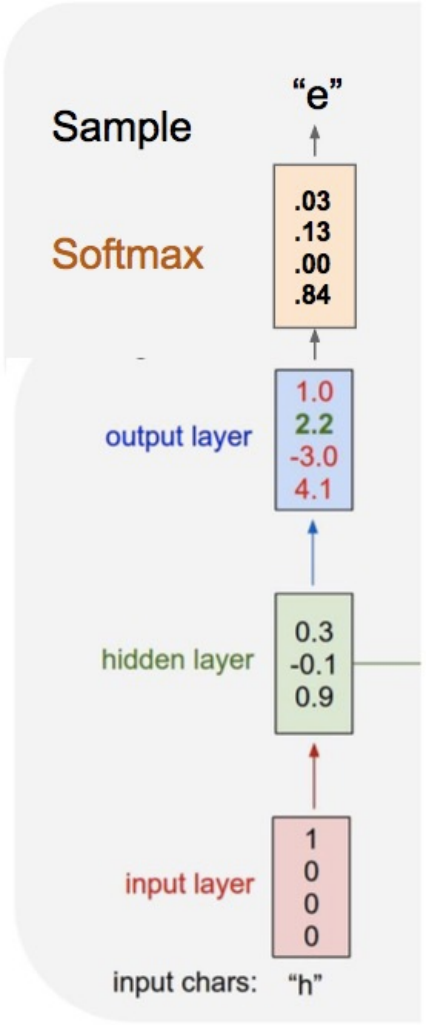
$$y_t = W_{hy} h_t$$

- Character-level Language Model
- 语料库：hello
- Vocabulary：[h, e, l, o]



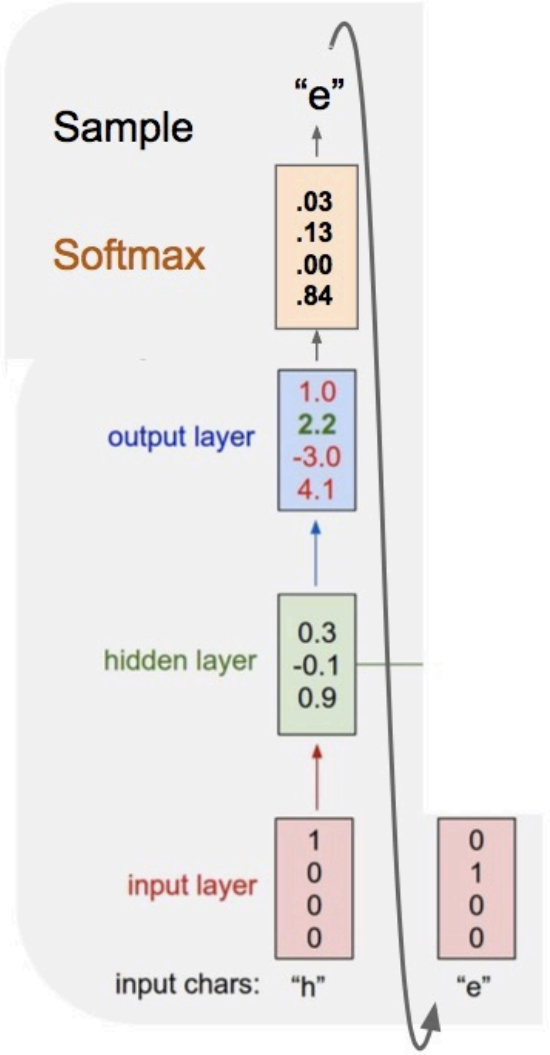
RNN实战：使用语言模型推理

- Character-level Language Model
- 语料库：hello
- Vocabulary：[h, e, l, o]
- 每次采样输出一个character



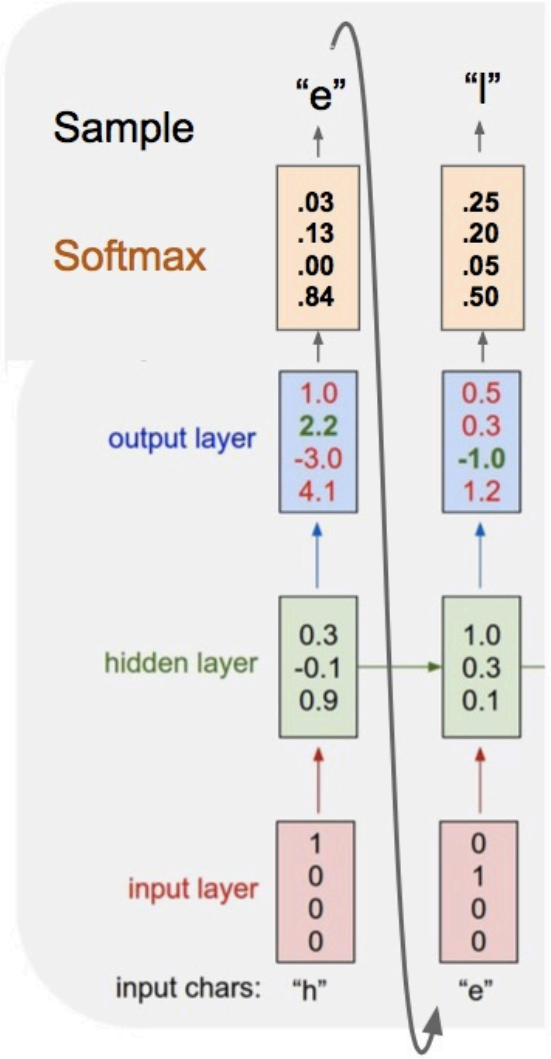
RNN实战：使用语言模型推理

- Character-level Language Model
- 语料库：hello
- Vocabulary：[h, e, l, o]
- 每次采样输出一个character，将前一时刻输出作为下一时刻输入



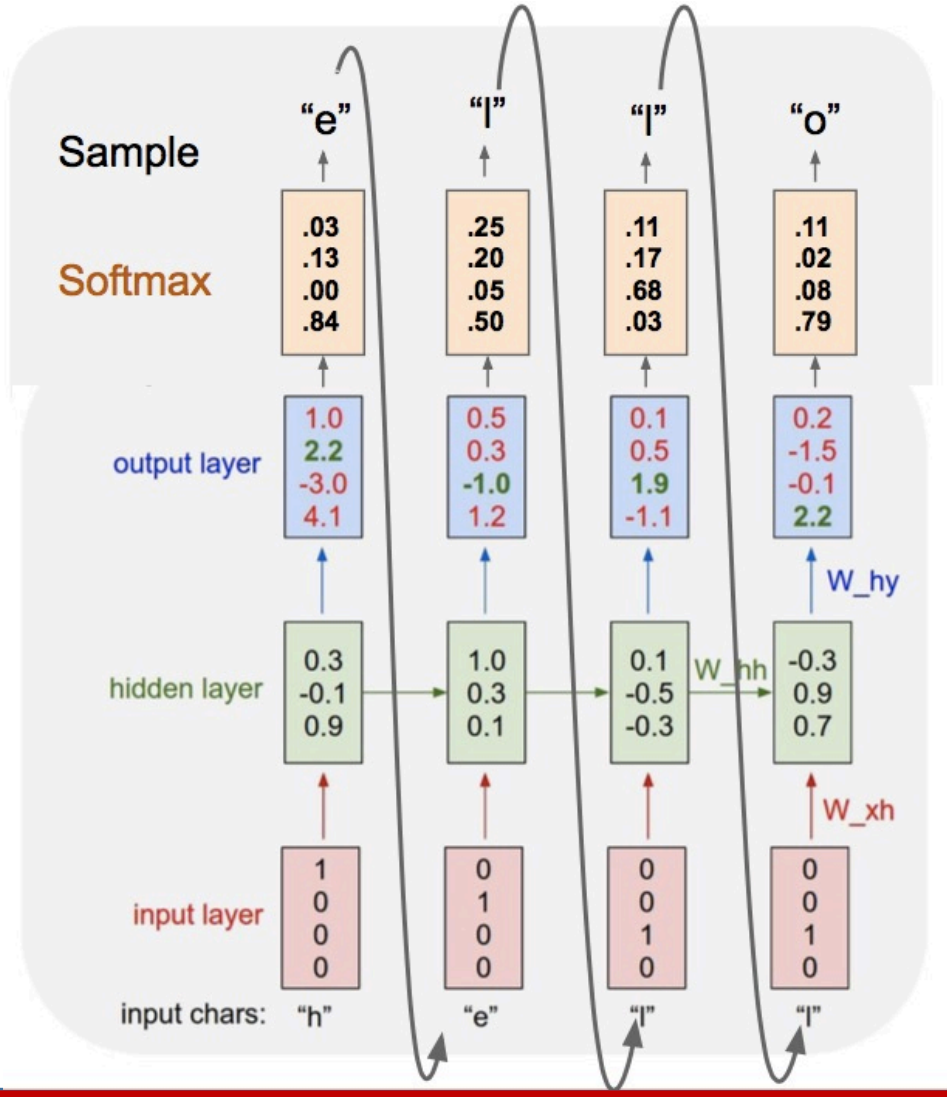
RNN实战：使用语言模型推理

- Character-level Language Model
- 语料库：hello
- Vocabulary：[h, e, l, o]
- 每次采样输出一个character，将前一时刻输出作为下一时刻输入



RNN实战：使用语言模型推理

- Character-level Language Model
- 语料库：hello
- Vocabulary：[h, e, l, o]
- 每次采样输出一个character，将前一时刻输出作为下一时刻输入



RNN应用：写诗

七步成诗 RNN @TangPoemGen · 2015年10月26日

【绝句】

扶苏堤上嫌龙雀，歌断近村偷酒蚕。
一吉旋书暴窈窕，春初娇语杀花春。
陈池听梦望墙门，事好宁知竞守秋。
谁不惯言归梦后，但将跳急逐人想。
愁鞮径著看窗中，忽觉圆光物外寒。
今日此心无住地，画帘休入暮云弦。

RNN应用：写论文

For $\bigoplus_{n=1,\ldots,m}$ where $\mathcal{L}_{m_\bullet} = 0$, hence we can find a closed subset $\mathcal{H}$ in $\mathcal{H}$ and any sets $\mathcal{F}$ on X, U is a closed immersion of S, then $U \to T$ is a separated algebraic space.

Proof. Proof of (1). It also start we get

$$S = \mathrm{Spec}(R) = U \times_X U \times_X U$$

and the compariciely in the fibre product covering we have to prove the lemma generated by $\coprod Z \times_U U \to V$. Consider the maps M along the set of points Sch_{fppf} and $U \to U$ is the fibre category of S in U in Section, ?? and the fact that any U affine, see Morphisms, Lemma ??. Hence we obtain a scheme S and any open subset $W \subset U$ in $Sh(G)$ such that $\mathrm{Spec}(R') \to S$ is smooth or an

$$U = \bigcup U_i \times_{S_i} U_i$$

which has a nonzero morphism we may assume that f_i is of finite presentation over S. We claim that $\mathcal{O}_{X,x}$ is a scheme where $x, x', s'' \in S'$ such that $\mathcal{O}_{X,x'} \to \mathcal{O}'_{X',x'}$ is separated. By Algebra, Lemma ?? we can define a map of complexes $\mathrm{GL}_{S'}(x'/S'')$ and we win. □

To prove study we see that $\mathcal{F}|_U$ is a covering of $\mathcal{X}'$, and $\mathcal{T}_i$ is an object of $\mathcal{F}_{X/S}$ for $i > 0$ and $\mathcal{F}_p$ exists and let $\mathcal{F}_i$ be a presheaf of $\mathcal{O}_X$-modules on $\mathcal{C}$ as a $\mathcal{F}$-module. In particular $\mathcal{F} = U/\mathcal{F}$ we have to show that

$$\widetilde{M}^\bullet = \mathcal{I}^\bullet \otimes_{\mathrm{Spec}(k)} \mathcal{O}_{S,s} - i_X^{-1}\mathcal{F})$$

is a unique morphism of algebraic stacks. Note that

$$\mathrm{Arrows} = (Sch/S)^{opp}_{fppf}, (Sch/S)_{fppf}$$

and

$$V = \Gamma(S, \mathcal{O}) \longmapsto (U, \mathrm{Spec}(A))$$

is an open subset of X. Thus U is affine. This is a continuous map of X is the inverse, the groupoid scheme S.

Proof. See discussion of sheaves of sets. □

The result for prove any open covering follows from the less of Example ??. It may replace S by $X_{spaces,étale}$ which gives an open subspace of X and T equal to S_{Zar}, see Descent, Lemma ??. Namely, by Lemma ?? we see that R is geometrically regular over S.

Lemma 0.1. *Assume (3) and (3) by the construction in the description.*

Suppose $X = \lim |X|$ *(by the formal open covering* X *and a single map* $\underline{Proj}_X(\mathcal{A}) = \mathrm{Spec}(B)$ *over* U *compatible with the complex*

$$Set(\mathcal{A}) = \Gamma(X, \mathcal{O}_{X,\mathcal{O}_X}).$$

When in this case of to show that $\mathcal{Q} \to \mathcal{C}_{Z/X}$ *is stable under the following result in the second conditions of (1), and (3). This finishes the proof. By Definition ?? (without element is when the closed subschemes are catenary. If* T *is surjective we may assume that* T *is connected with residue fields of* S*. Moreover there exists a closed subspace* $Z \subset X$ *of* X *where* U *in* X' *is proper (some defining as a closed subset of the uniqueness it suffices to check the fact that the following theorem*

(1) f *is locally of finite type. Since* $S = \mathrm{Spec}(R)$ *and* $Y = \mathrm{Spec}(R)$.

Proof. This is form all sheaves of sheaves on X. But given a scheme U and a surjective étale morphism $U \to X$. Let $U \cap U = \coprod_{i=1,\ldots,n} U_i$ be the scheme X over S at the schemes $X_i \to X$ and $U = \lim_i X_i$. □

The following lemma surjective restrocomposes of this implies that $\mathcal{F}_{x_0} = \mathcal{F}_{x_0} = \mathcal{F}_{\mathcal{X},\ldots,0}$.

Lemma 0.2. *Let* X *be a locally Noetherian scheme over* S, $E = \mathcal{F}_{X/S}$. *Set* $\mathcal{I} = \mathcal{J}_1 \subset \mathcal{I}'_n$. *Since* $\mathcal{I}^n \subset \mathcal{I}^n$ *are nonzero over* $i_0 \leq \mathfrak{p}$ *is a subset of* $\mathcal{J}_{n,0} \circ \overline{A}_2$ *works.*

Lemma 0.3. *In Situation ??. Hence we may assume* $\mathfrak{q}' = 0$.

Proof. We will use the property we see that $\mathfrak{p}$ is the mext functor (??). On the other hand, by Lemma ?? we see that

$$D(\mathcal{O}_{X'}) = \mathcal{O}_X(D)$$

where K is an F-algebra where δ_{n+1} is a scheme over S. □

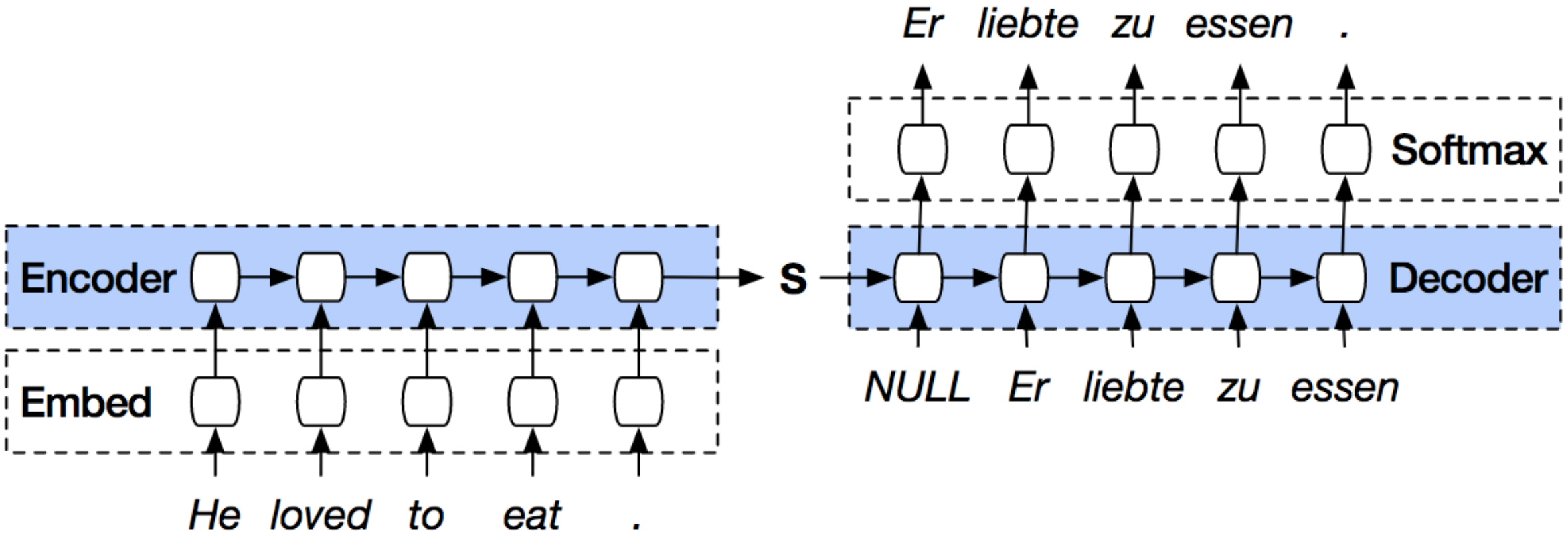
- 生成latex源码

RNN应用：写代码

```
static void do_command(struct seq_file *m, void *v)
{
  int column = 32 << (cmd[2] & 0x80);
  if (state)
    cmd = (int)(int_state ^ (in_8(&ch->ch_flags) & Cmd) ? 2 : 1);
  else
    seq = 1;
  for (i = 0; i < 16; i++) {
    if (k & (1 << 1))
      pipe = (in_use & UMXTHREAD_UNCCA) +
        ((count & 0x00000000fffffff8) & 0x000000f) << 8;
    if (count == 0)
      sub(pid, ppc_md.kexec_handle, 0x20000000);
    pipe_set_bytes(i, 0);
  }
  /* Free our user pages pointer to place camera if all dash */
  subsystem_info = &of_changes[PAGE_SIZE];
  rek_controls(offset, idx, &soffset);
  /* Now we want to deliberately put it to device */
  control_check_polarity(&context, val, 0);
  for (i = 0; i < COUNTER; i++)
    seq_puts(s, "policy ");
}
```

- 利用整个Linux内核训练

RNN应用：机器翻译

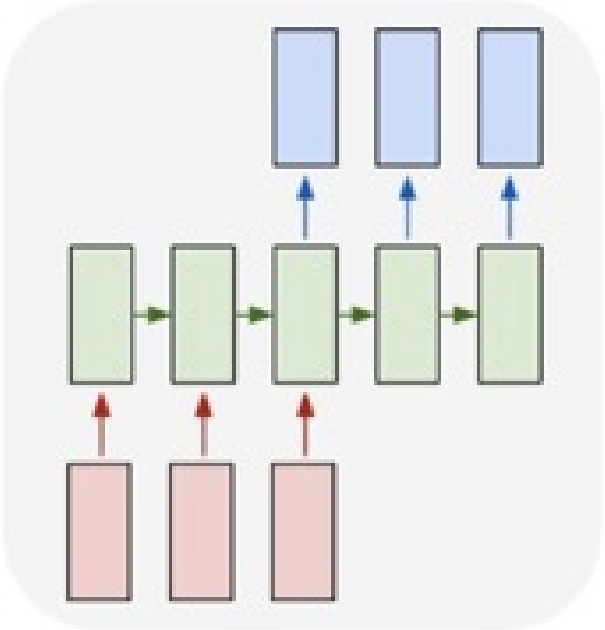


Sutskever I, Vinyals O, Le QV. Sequence to sequence learning with neural networks. NIPS 2014.

RNN应用：代码纠错

```
1  #include <stdio.h>
2  #include <stdlib.h>
3  int pow(int a, int b);
4  int main(){
5   int n;
6   scanf("%d",&n);
7   int i, j;
8   for(i=1;i<=n;i++){
9    for(j=0;j<=n;j++){
10    if(j<i){
11     printf("%d ",pow(i,j));}}
12   printf("\n");}
13  return 0;
14 int pow(int a, int b){
15  int i, res=1;
16  for(i=0;i<b;i++)
17   res = a * res;
18  return res;}
```

(a) Input program `p.c` with a missing closing brace at line 13



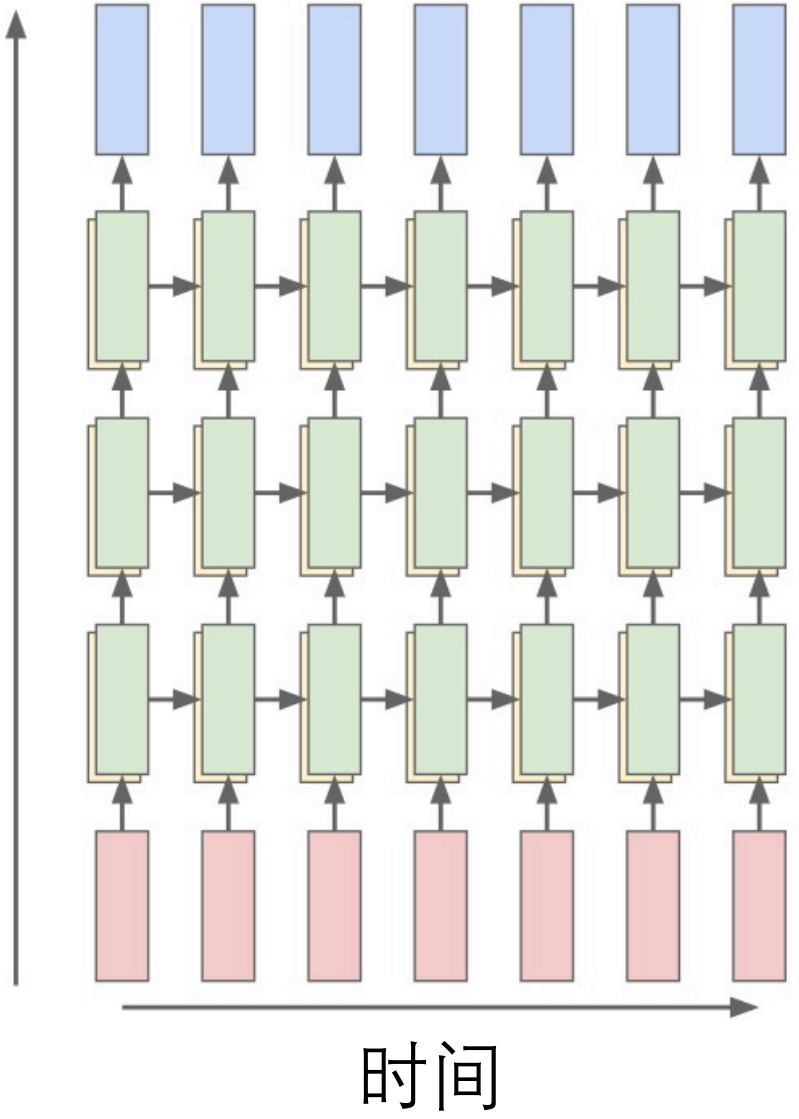
```
1  #include <stdio.h>
2  #include <stdlib.h>
3  int pow(int a, int b);
4  int main(){
5   int n;
6   scanf("%d",&n);
7   int i, j;
8   for(i=1;i<=n;i++){
9    for(j=0;j<=n;j++){
10    if(j<i){
11     printf("%d ",pow(i,j));}}
12   printf("\n");}
13  return 0;}
14 int pow(int a, int b){
15  int i, res=1;
16  for(i=0;i<b;i++)
17   res = a * res;
18  return res;}
```

(c) The program after the fix (shaded) suggested by DeepFix is applied

多层RNN结构

$$h_t^l = \tanh W^l \begin{pmatrix} h_t^{l-1} \\ h_{t-1}^l \end{pmatrix}$$

$h \in \mathbb{R}^n$ W^l $[n \times 2n]$



图像描述：image captioning

A cat sitting on a suitcase on the floor

A cat is sitting on a tree branch

A dog is running in the grass with a frisbee

A white teddy bear sitting in the grass

Two people walking on the beach with surfboards

A tennis player in action on the court

Two giraffes standing in a grassy field

A man riding a dirt bike on a dirt track

Mao J, Xu W, et al. Deep captioning with multimodal recurrent neural networks. 2014.

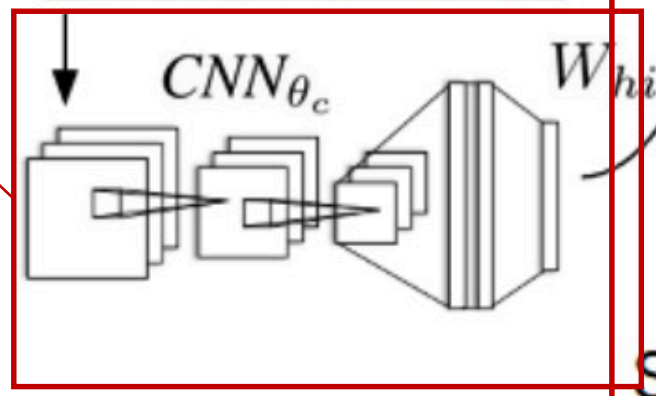
Karpathy A, Fei-Fei L. Deep visual-semantic alignments for generating image descriptions. CVPR 2015.

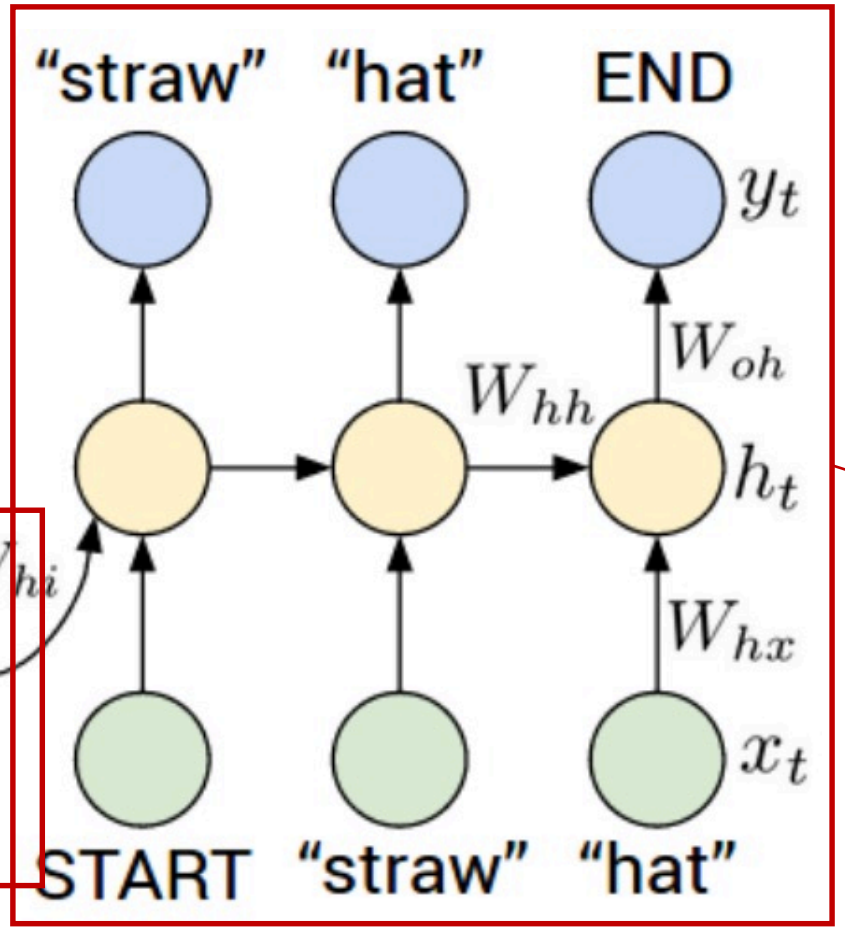
Vinyals O, Toshev A, Bengio S, Erhan D. Show and tell: A neural image caption generator. CVPR 2015.

Xu K, Ba J, Kiros R, Cho K, et al. Show, attend and tell: Neural image caption generation with visual attention. ICML 2015.

图像描述

使用ImageNet
预训练CNN





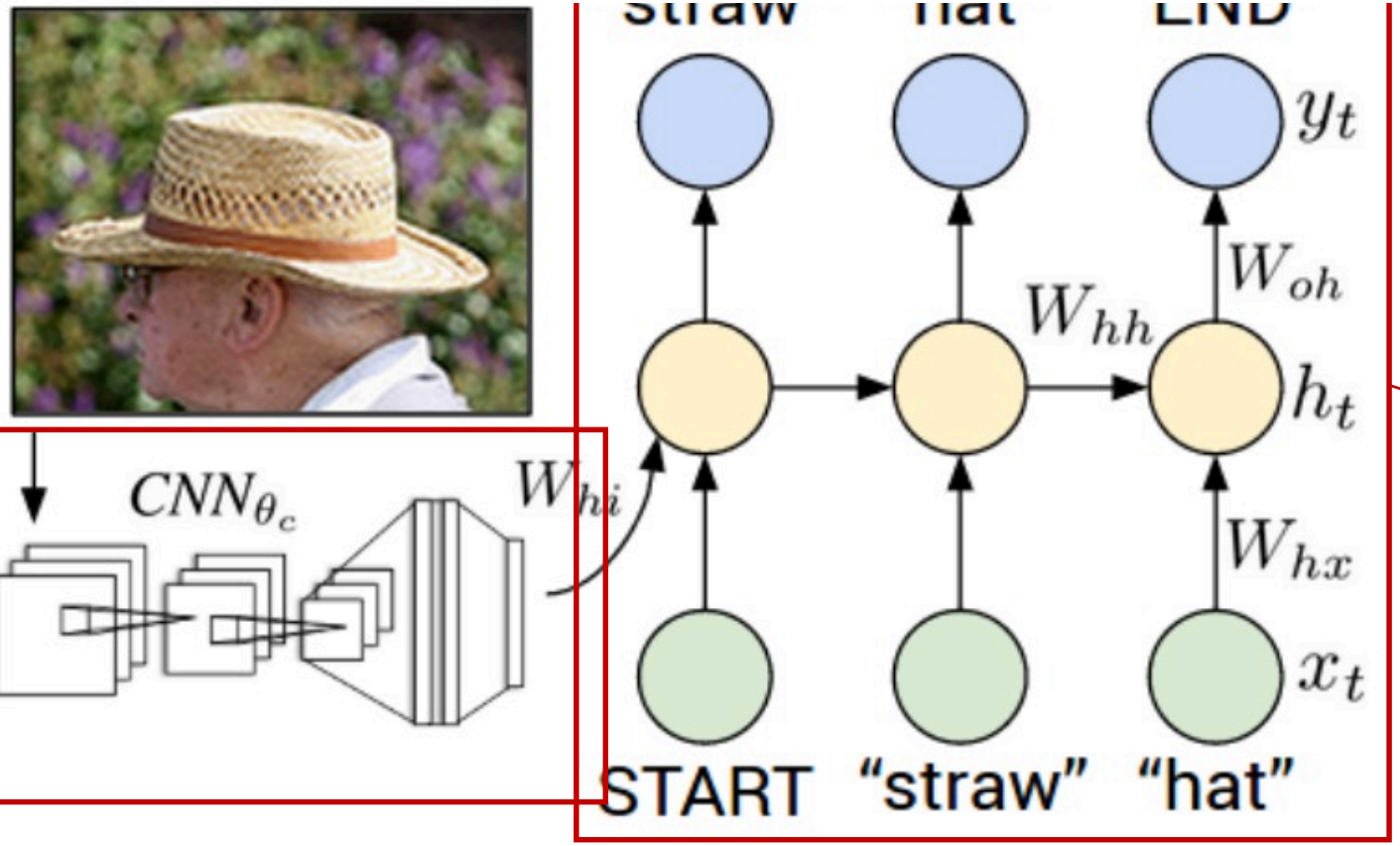
RNN

Karpathy A, Fei-Fei L. Deep visual-semantic alignments for generating image descriptions. CVPR 2015.

图像描述

$$b_v = W_{hi}[CNN_{\theta_c}(I)]$$
$$h_t = f(W_{hx}x_t + W_{hh}h_{t-1} + b_h + \mathbb{1}(t=1) \odot b_v)$$
$$y_t = softmax(W_{oh}h_t + b_o).$$

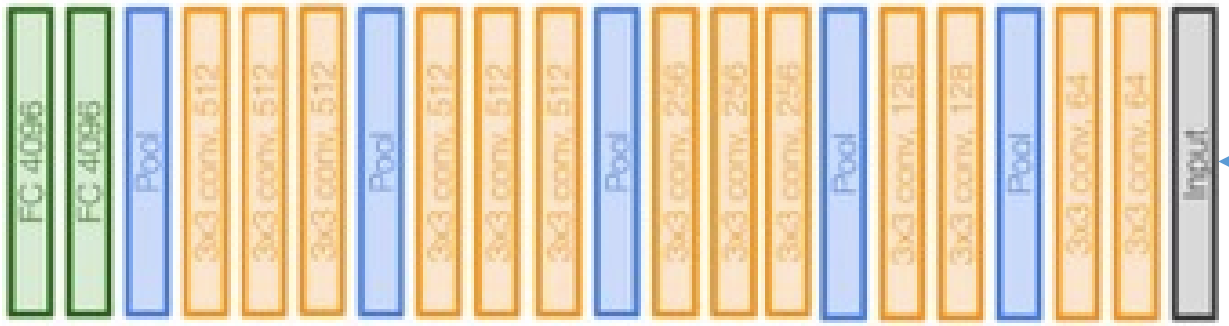
使用ImageNet预训练CNN

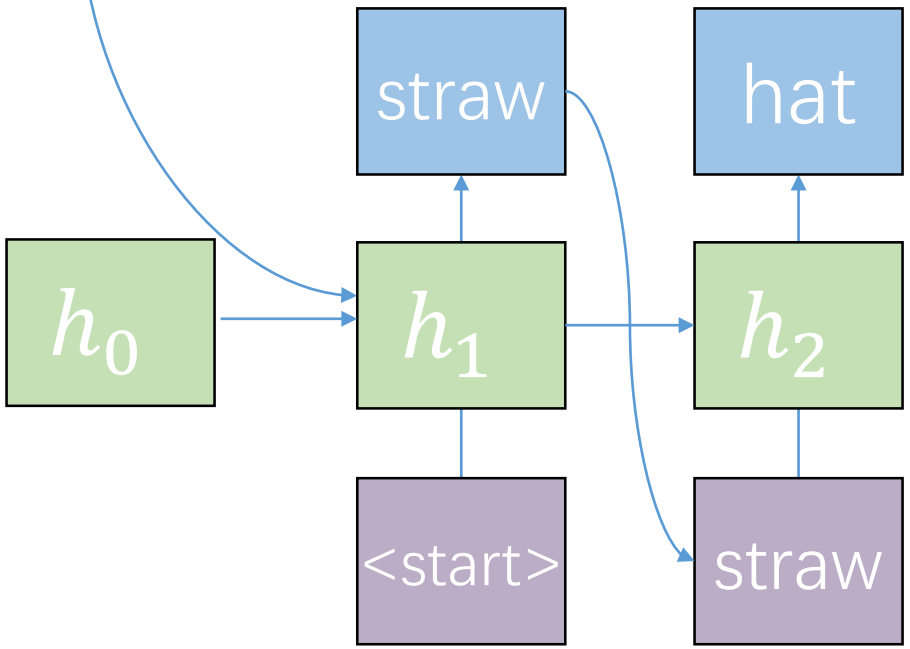


RNN

Karpathy A, Fei-Fei L. Deep visual-semantic alignments for generating image descriptions. CVPR 2015.

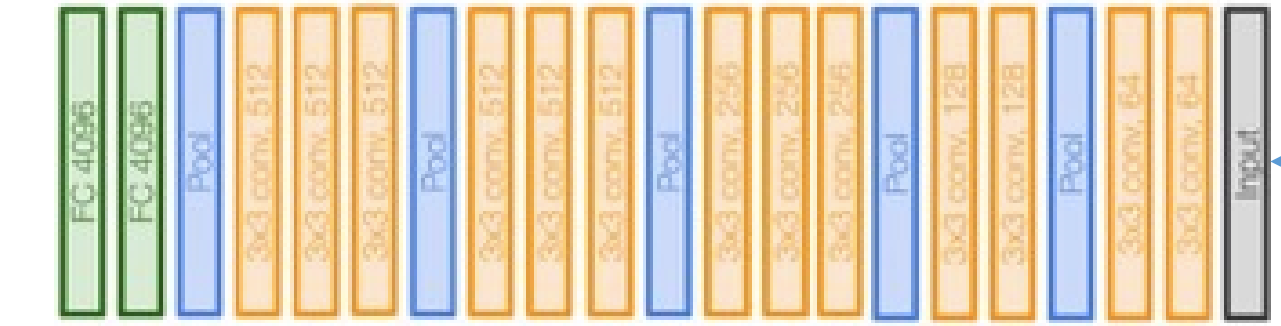
生成图像描述

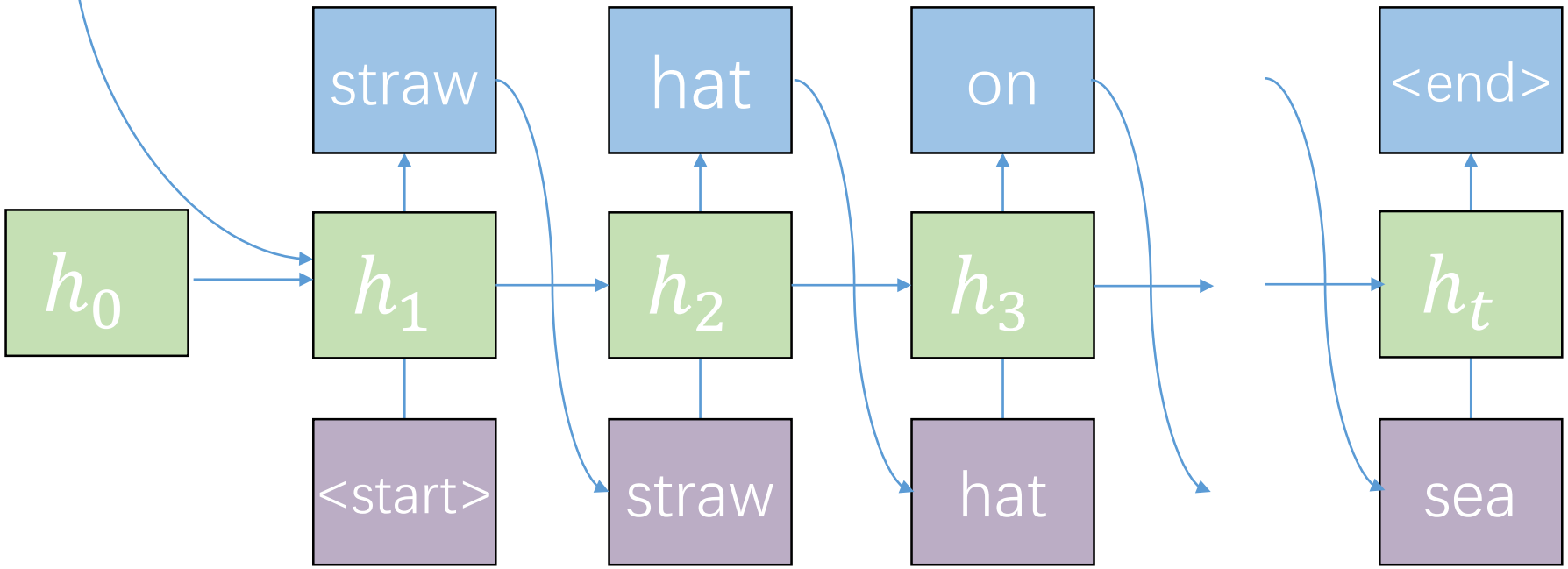




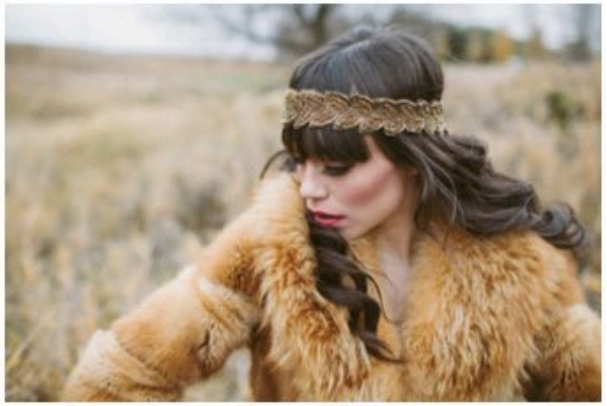
$$b_v = W_{hi}[CNN_{\theta_c}(I)]$$
$$h_t = f(W_{hx}x_t + W_{hh}h_{t-1} + b_h + \mathbb{1}(t=1) \odot b_v)$$
$$y_t = softmax(W_{oh}h_t + b_o).$$

生成图像描述


FC 4096
FC 4096
Pool
3x3 conv, 512
3x3 conv, 512
3x3 conv, 512
Pool
3x3 conv, 512
3x3 conv, 512
3x3 conv, 512
Pool
3x3 conv, 256
3x3 conv, 256
3x3 conv, 256
Pool
3x3 conv, 128
3x3 conv, 128
Pool
3x3 conv, 64
3x3 conv, 64
Input

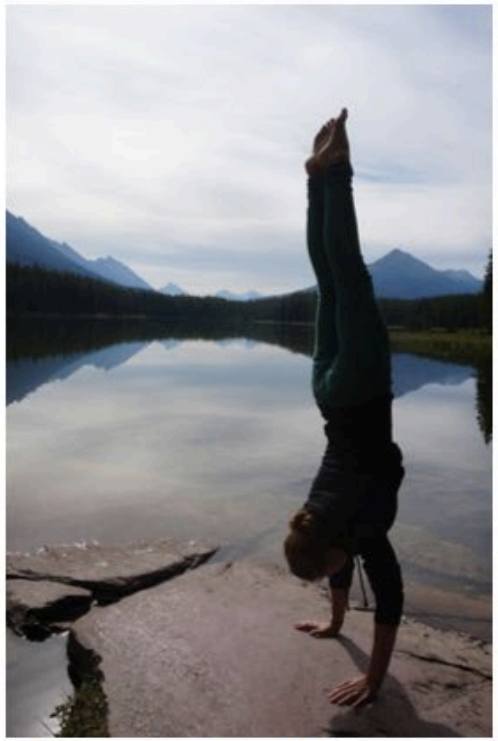

straw
hat
on
<end>
h_0
h_1
h_2
h_3
h_t
<start>
straw
hat
sea

生成描述不正确的例子



A woman is holding a cat in her hand

A person holding a computer mouse on a desk

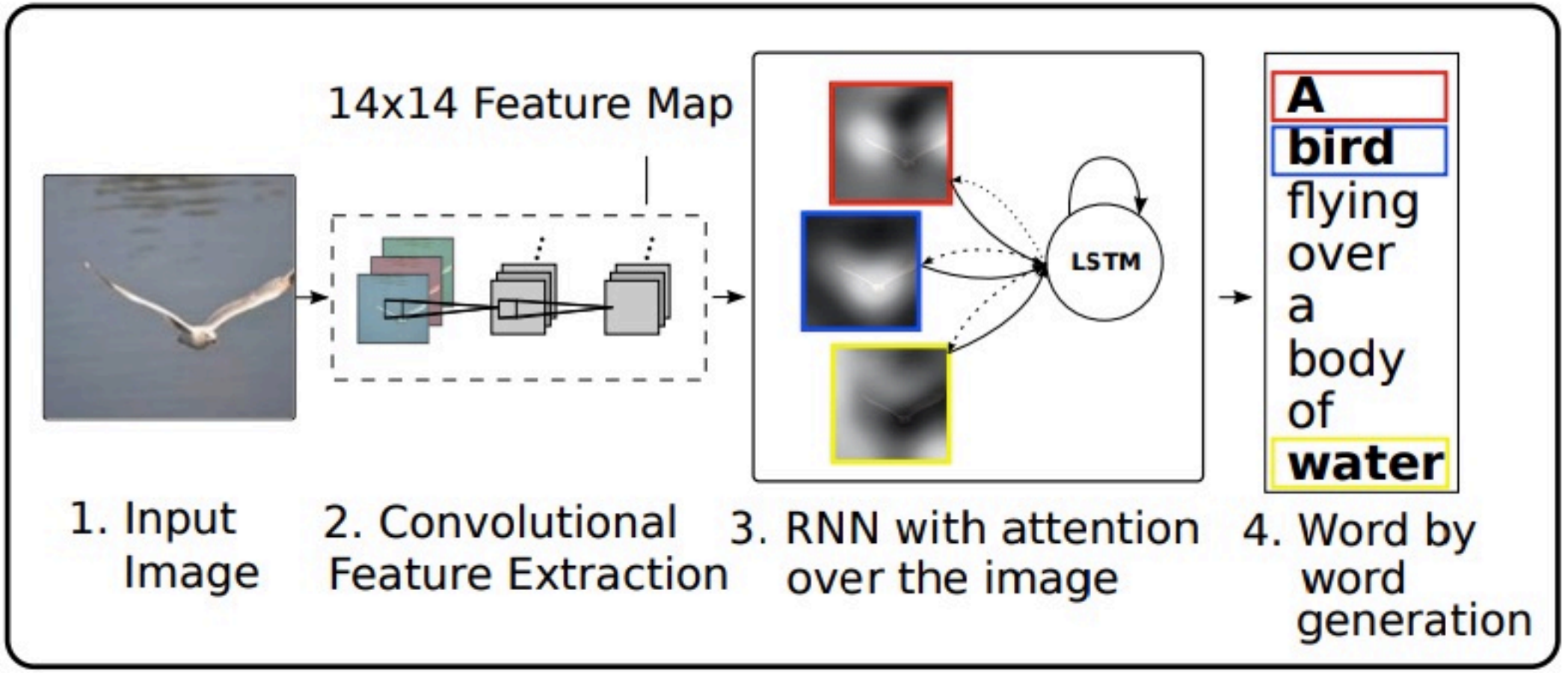


A woman standing on a beach holding a surfboard

A bird is perched on a tree branch

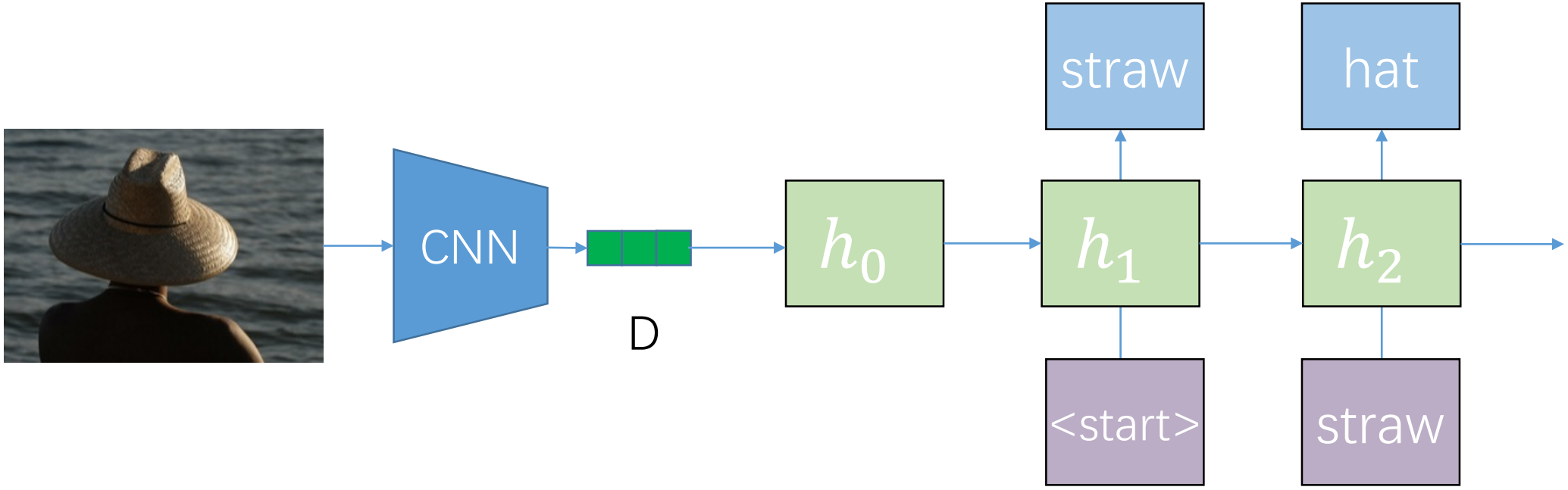
A man in a baseball uniform throwing a ball

使用注意力机制生成描述

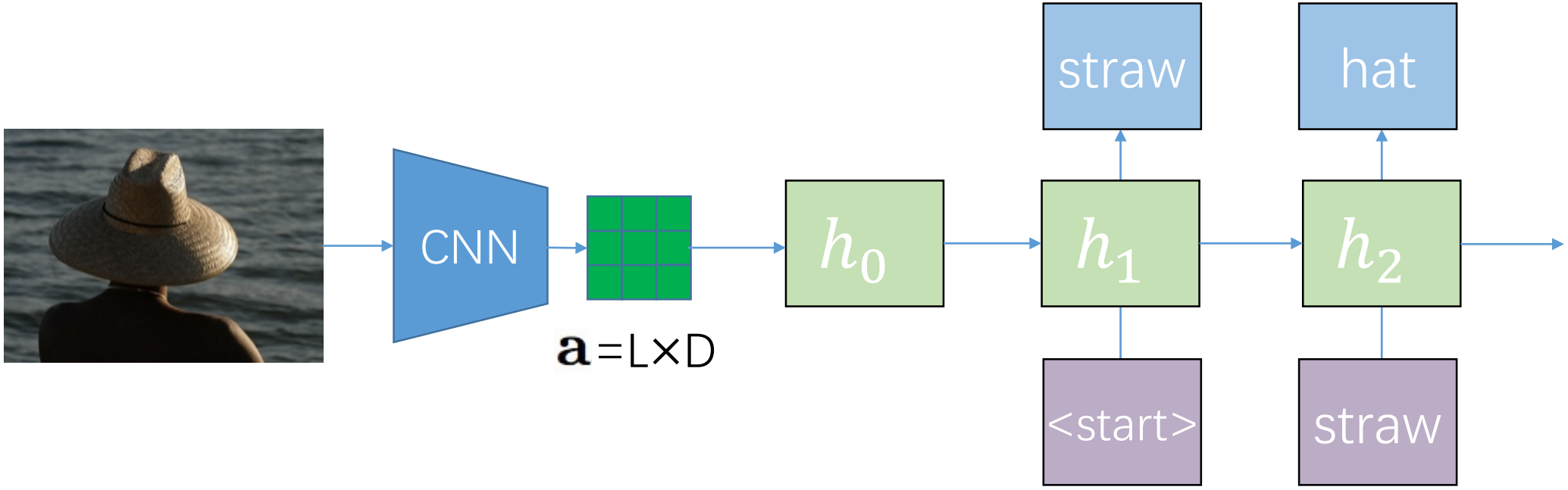


Xu K, Ba J, Kiros R, Cho K, et al. Show, attend and tell: Neural image caption generation with visual attention. ICML 2015.

不使用attention



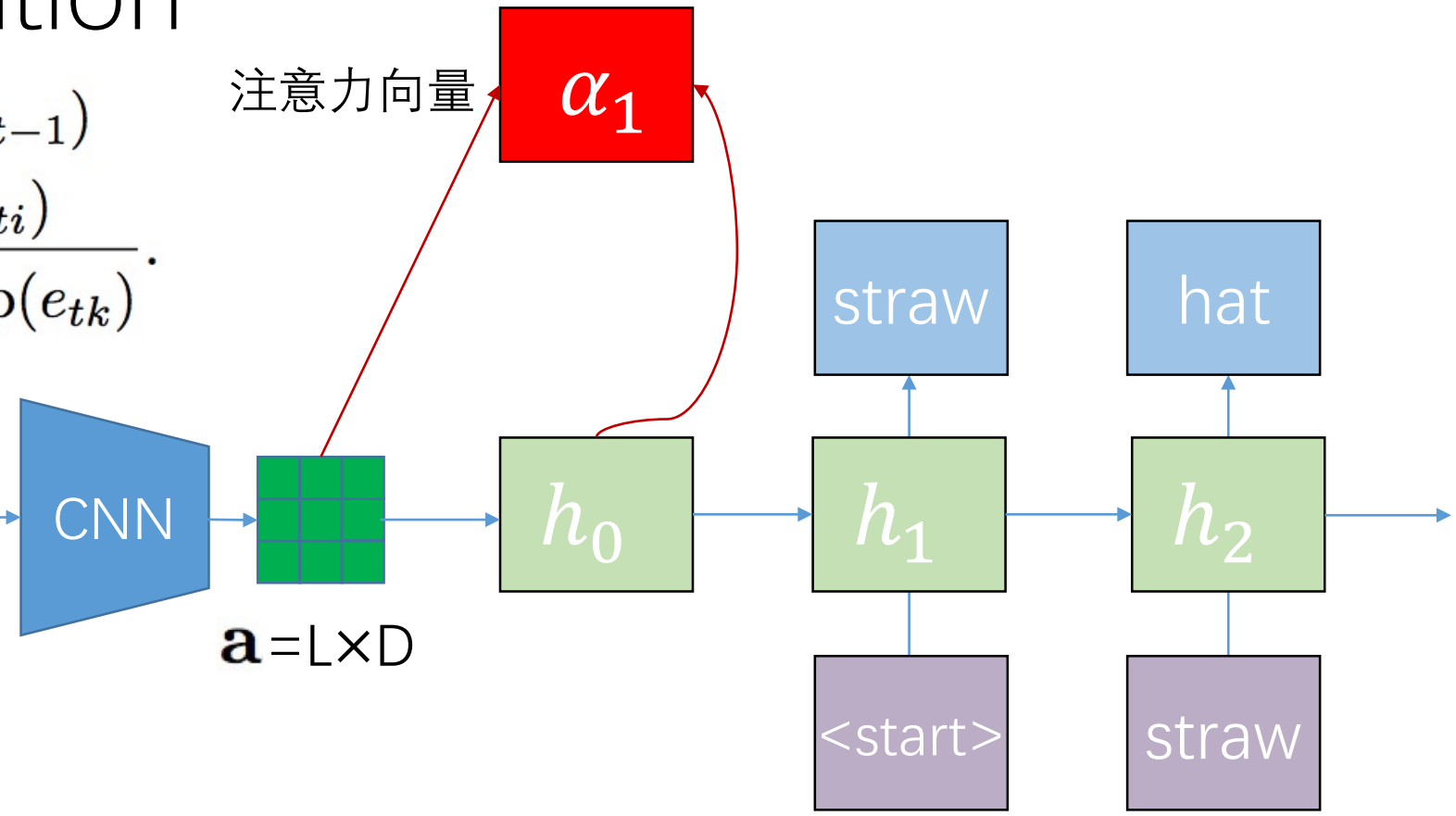
使用attention



使用attention

$$e_{ti} = f_{\text{att}}(\mathbf{a}_i, \mathbf{h}_{t-1})$$

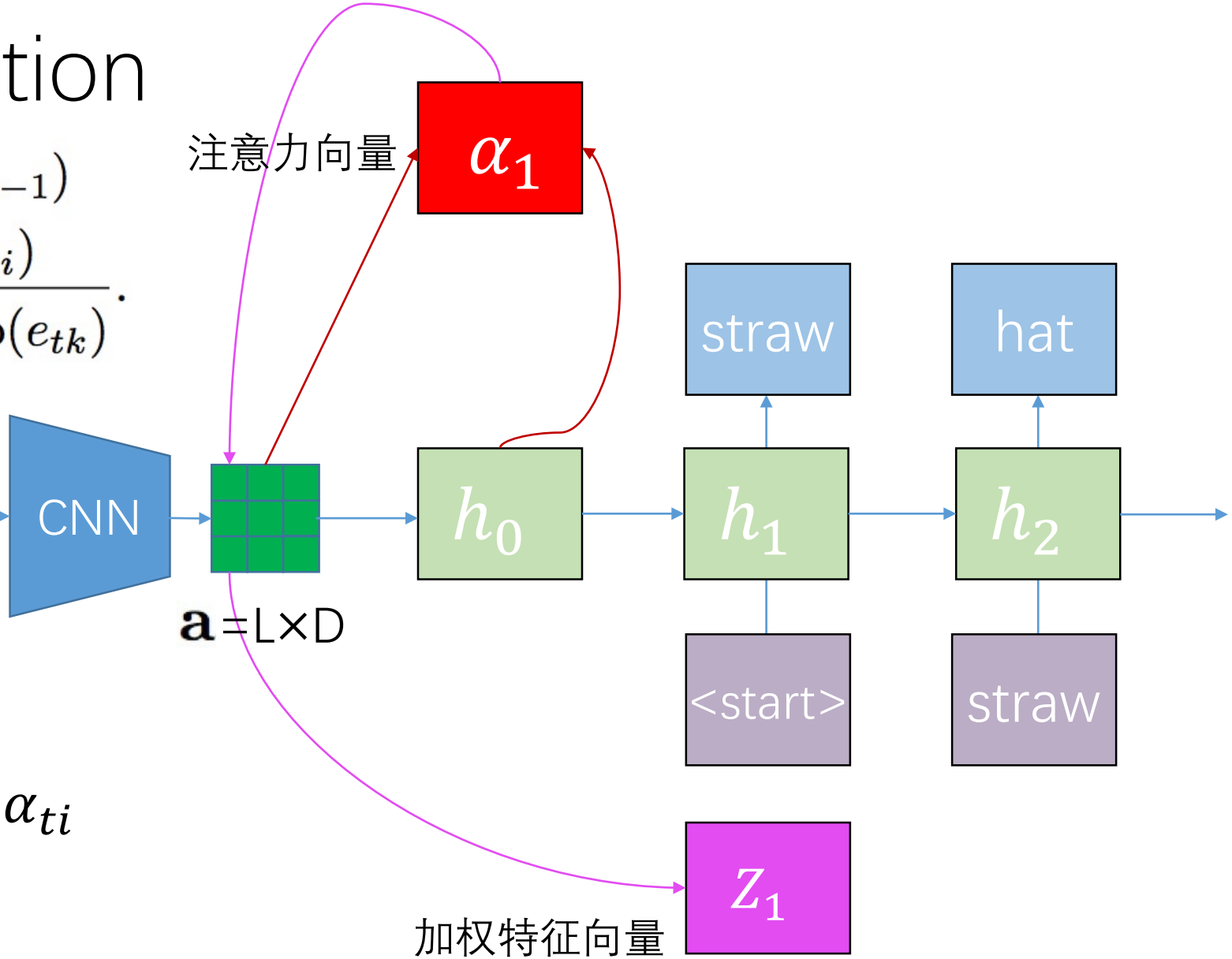
$$\alpha_{ti} = \frac{\exp(e_{ti})}{\sum_{k=1}^{L} \exp(e_{tk})}.$$



使用attention

$$e_{ti} = f_{\text{att}}(\mathbf{a}_i, \mathbf{h}_{t-1})$$

$$\alpha_{ti} = \frac{\exp(e_{ti})}{\sum_{k=1}^{L} \exp(e_{tk})}.$$

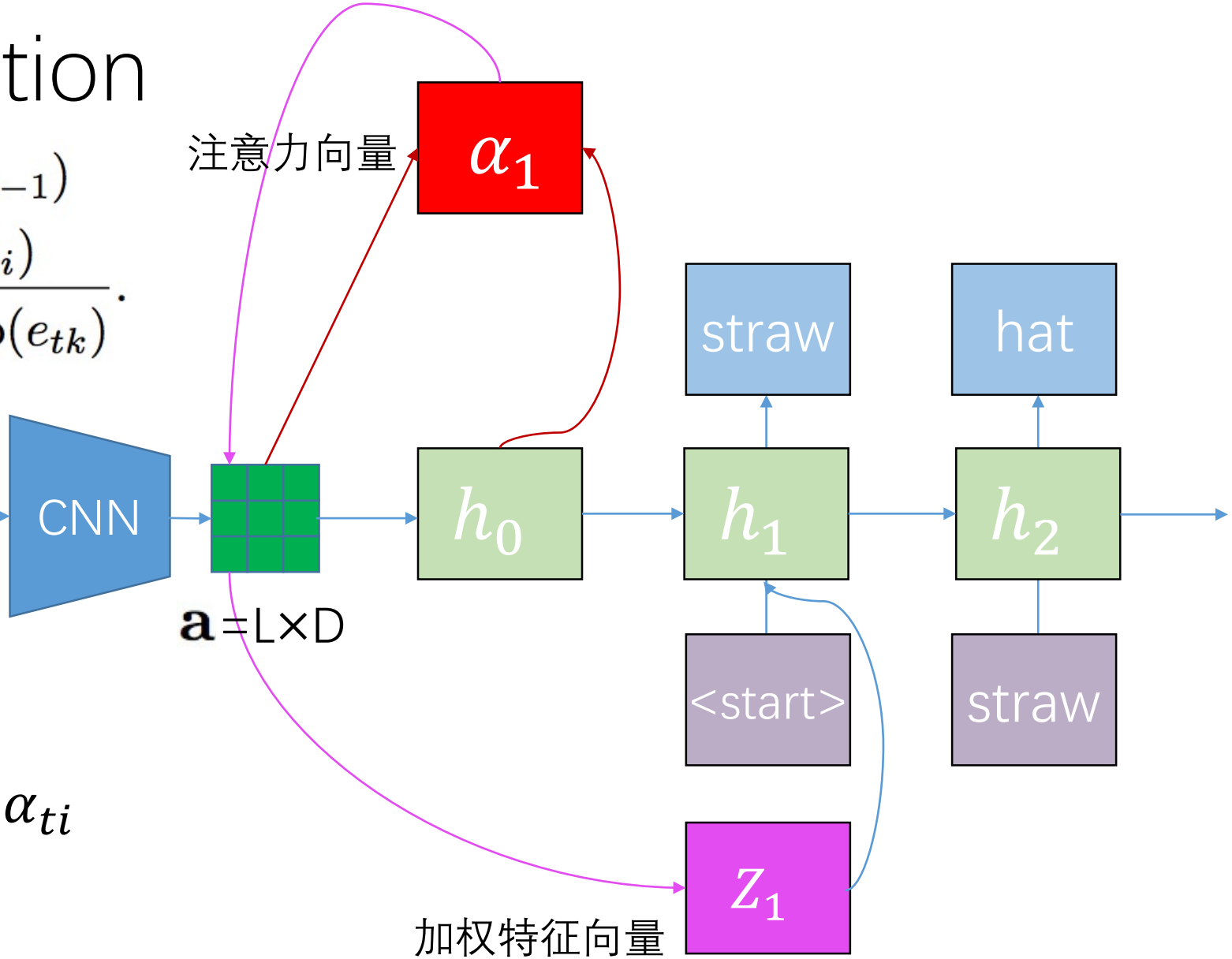


$$Z_t = \sum_{i=1}^{L} \mathbf{a}_i * \alpha_{ti}$$

使用attention

$$e_{ti} = f_{\text{att}}(\mathbf{a}_i, \mathbf{h}_{t-1})$$

$$\alpha_{ti} = \frac{\exp(e_{ti})}{\sum_{k=1}^{L} \exp(e_{tk})}.$$



$$Z_t = \sum_{i=1}^{L} \mathbf{a}_i * \alpha_{ti}$$

使用attention

$$e_{ti} = f_{\text{att}}(\mathbf{a}_i, \mathbf{h}_{t-1})$$

$$\alpha_{ti} = \frac{\exp(e_{ti})}{\sum_{k=1}^{L} \exp(e_{tk})}.$$

CNN

$\mathbf{a}$=L×D

注意力向量

α_1 α_2

h_0 h_1 h_2

straw hat

<start> straw

Z_1 Z_2

加权特征向量

$$Z_t = \sum_{i=1}^{L} \mathbf{a}_i * \alpha_{ti}$$

使用attention

$$e_{ti} = f_{\text{att}}(\mathbf{a}_i, \mathbf{h}_{t-1})$$

$$\alpha_{ti} = \frac{\exp(e_{ti})}{\sum_{k=1}^{L} \exp(e_{tk})}.$$

$$Z_t = \sum_{i=1}^{L} \mathbf{a}_i * \alpha_{ti}$$

注意力向量

α_1 α_2 α_3

CNN

$\mathbf{a}$=L×D

h_0 h_1 h_2

straw hat

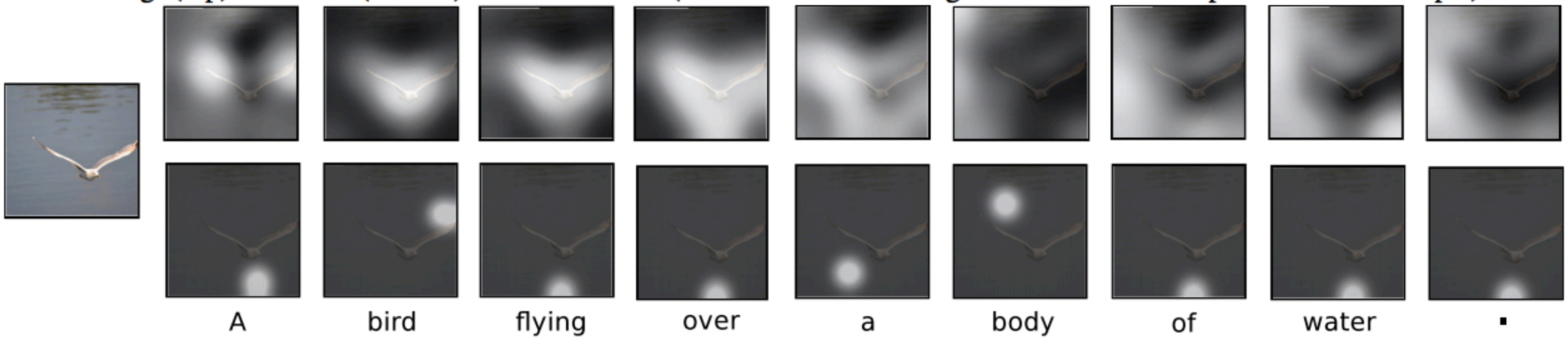
<start> straw

Z_1 Z_2 Z_3

加权特征向量

使用注意力机制生成描述

Figure 3. Visualization of the attention for each generated word. The rough visualizations obtained by upsampling the attention weights and smoothing. (top)"soft" and (bottom) "hard" attention (note that both models generated the same captions in this example).



Xu K, Ba J, Kiros R, Cho K, et al. Show, attend and tell: Neural image caption generation with visual attention. ICML 2015.

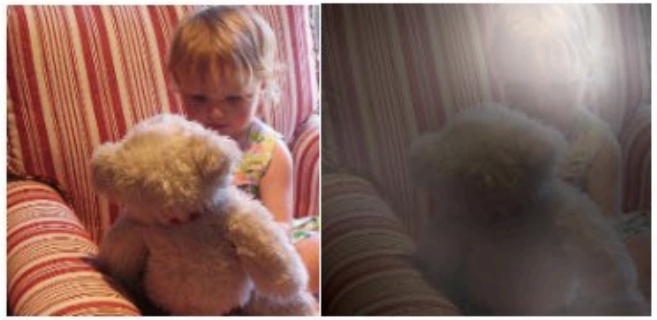
使用注意力机制生成描述

Figure 4. Examples of attending to the correct object (*white* indicates the attended regions, *underlines* indicated the corresponding word)

A woman is throwing a frisbee in a park.

A dog is standing on a hardwood floor.

A stop sign is on a road with a mountain in the background.



A little girl sitting on a bed with a teddy bear.

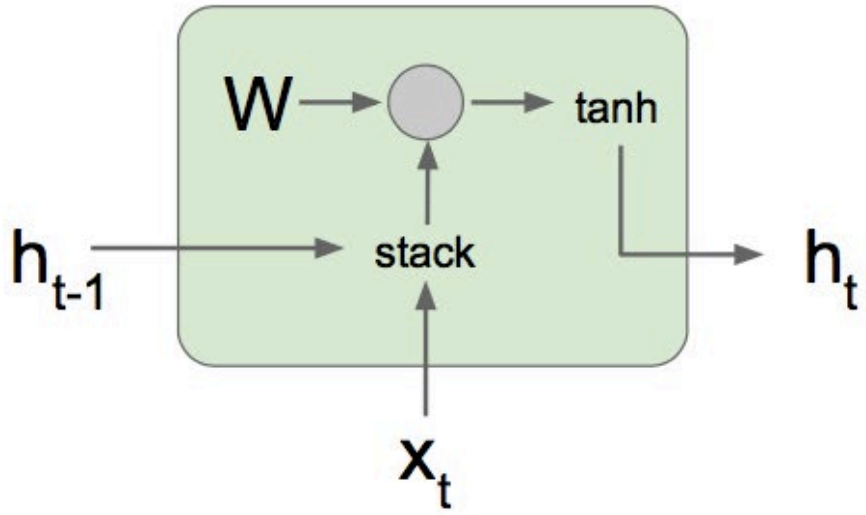
A group of people sitting on a boat in the water.

A giraffe standing in a forest with trees in the background.

Xu K, Ba J, Kiros R, Cho K, et al. Show, attend and tell: Neural image caption generation with visual attention. ICML 2015.

RNNs的缺点

- 前向计算

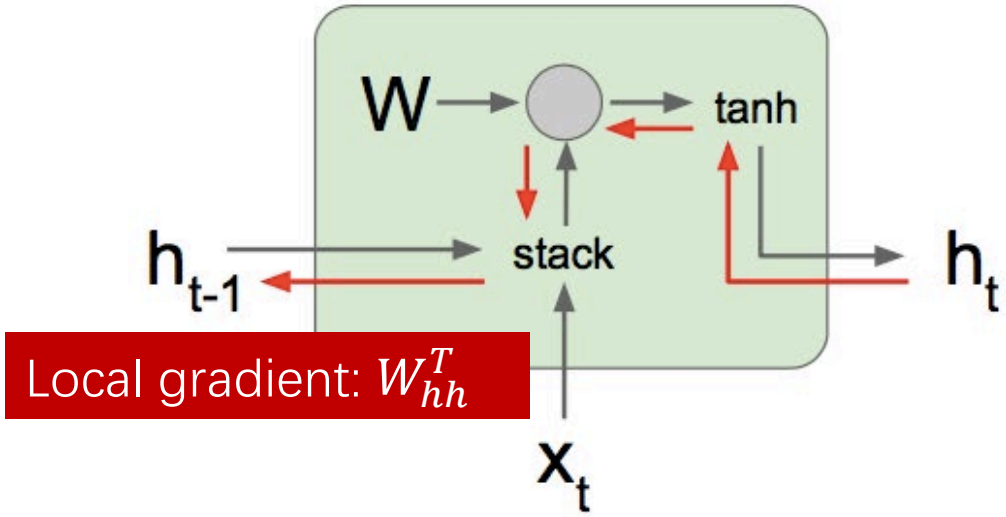


$$h_t = \tanh(W_{hh}h_{t-1} + W_{xh}x_t)$$
$$= \tanh\left((W_{hh} \quad W_{hx})\begin{pmatrix} h_{t-1} \\ x_t \end{pmatrix}\right)$$
$$= \tanh\left(W\begin{pmatrix} h_{t-1} \\ x_t \end{pmatrix}\right)$$

Bengio Y, Simard P, Frasconi P. Learning long-term dependencies with gradient descent is difficult. IEEE transactions on neural networks. 1994.

RNNs的缺点

- 反向计算梯度流（gradient flow）



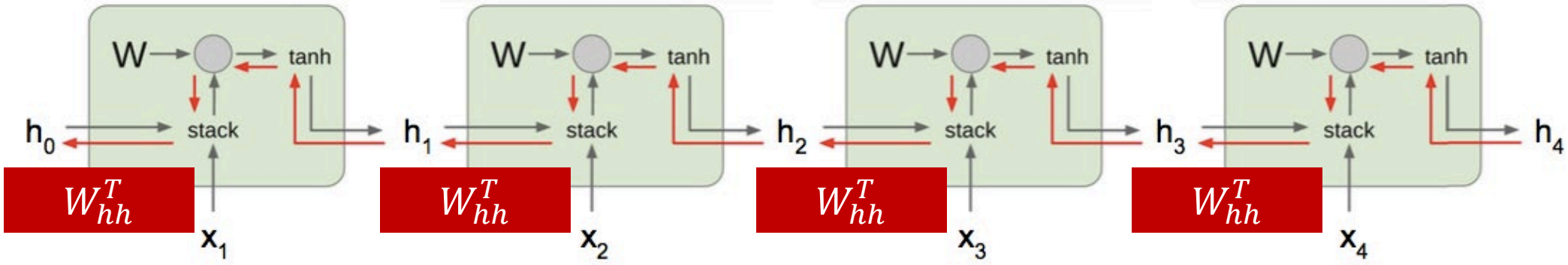
Local gradient: W_{hh}^T

$$h_t = \tanh(W_{hh}h_{t-1} + W_{xh}x_t)$$
$$= \tanh\left((W_{hh} \quad W_{hx})\begin{pmatrix} h_{t-1} \\ x_t \end{pmatrix}\right)$$
$$= \tanh\left(W\begin{pmatrix} h_{t-1} \\ x_t \end{pmatrix}\right)$$

Bengio Y, Simard P, Frasconi P. Learning long-term dependencies with gradient descent is difficult. IEEE transactions on neural networks. 1994.

RNNs的缺点

- 反向计算梯度流（gradient flow）



特征向量组成的矩阵

特征值组成的对角线矩阵

链式法则：反复乘以同一个矩阵W_{hh}^T

$$W_{hh}^T = Q\Lambda Q^{-1} \rightarrow (W_{hh}^T)^n = Q\Lambda^n Q^{-1}$$

特征值>1：梯度爆炸

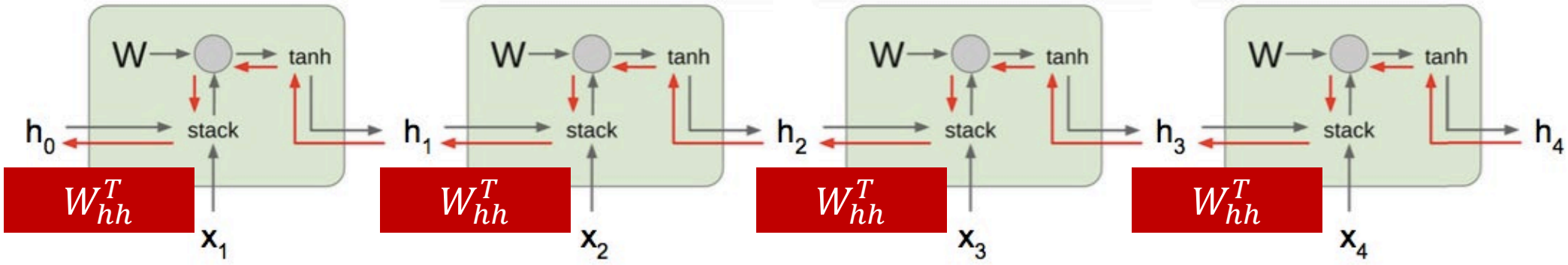
特征值<1：梯度消失

Bengio Y, Simard P, Frasconi P. Learning long-term dependencies with gradient descent is difficult. IEEE transactions on neural networks. 1994.

梯度裁剪

```
grad_norm = np.sum(grad * grad)
if grad_norm > threshold:
  grad *= (threshold / grad_norm)
```

- 反向计算梯度流（gradient flow）



特征值>1：梯度爆炸

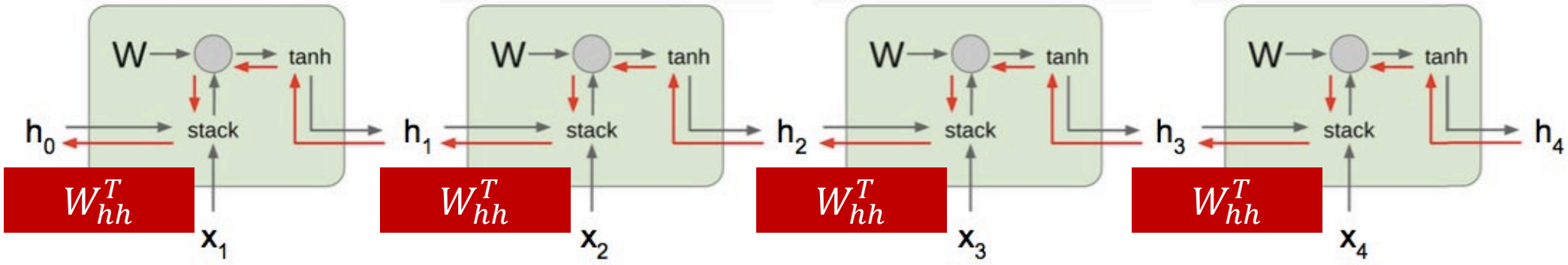
梯度裁剪： Gradient clipping

✓ 如果梯度的L2 norm过大，则缩小梯度

Pascanu R, Mikolov T, Bengio Y. On the difficulty of training recurrent neural networks.ICML 2013.

使用更高级结构

- 反向计算梯度流（gradient flow）



特征值<1：梯度消失 → LSTM、GRU、etc.

Pascanu R, Mikolov T, Bengio Y. On the difficulty of training recurrent neural networks.ICML 2013.

Long Short-Term Memory (长短期记忆)

Vanilla RNN

$$h_t = \tanh\left(W \begin{pmatrix} h_{t-1} \\ x_t \end{pmatrix}\right)$$

VS

LSTM

四个向量

$$\begin{pmatrix} i \\ f \\ o \\ g \end{pmatrix} = \begin{pmatrix} \sigma \\ \sigma \\ \sigma \\ \tanh \end{pmatrix} W \begin{pmatrix} h_{t-1} \\ x_t \end{pmatrix}$$

$$c_t = f \odot c_{t-1} + i \odot g$$

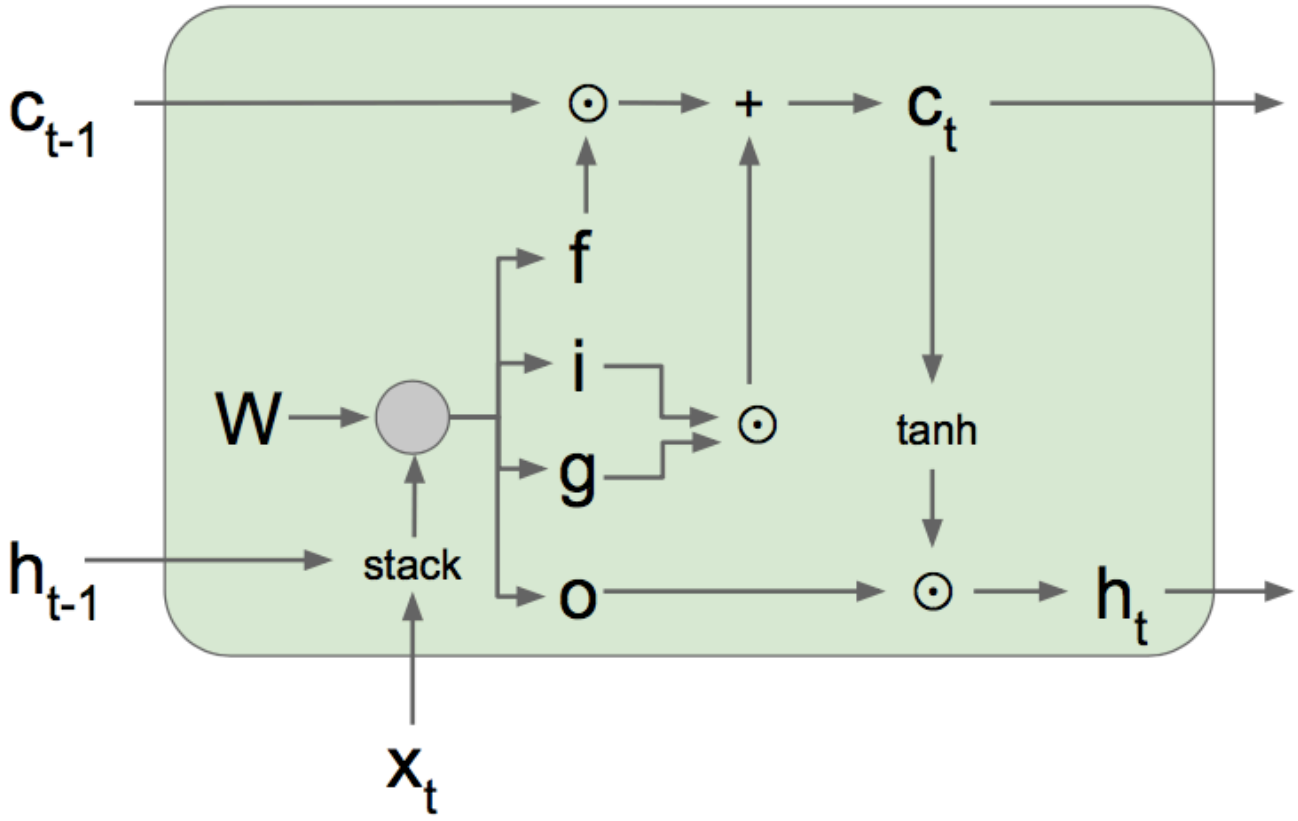
$$h_t = o \odot \tanh(c_t)$$

cell state vector:存储历史信息

Hochreiter S, Schmidhuber J. Long short-term memory. Neural computation. 1997.

LSTM：前向传播

i: input gate，控制在c_t中写入哪些信息
f: forget gate，控制从c_{t-1}中擦除哪些信息
o: output gate，控制从c_t输出哪些信息
g: 激活输入



$$\begin{pmatrix} i \\ f \\ o \\ g \end{pmatrix} = \begin{pmatrix} \sigma \\ \sigma \\ \sigma \\ \tanh \end{pmatrix} W \begin{pmatrix} h_{t-1} \\ x_t \end{pmatrix}$$

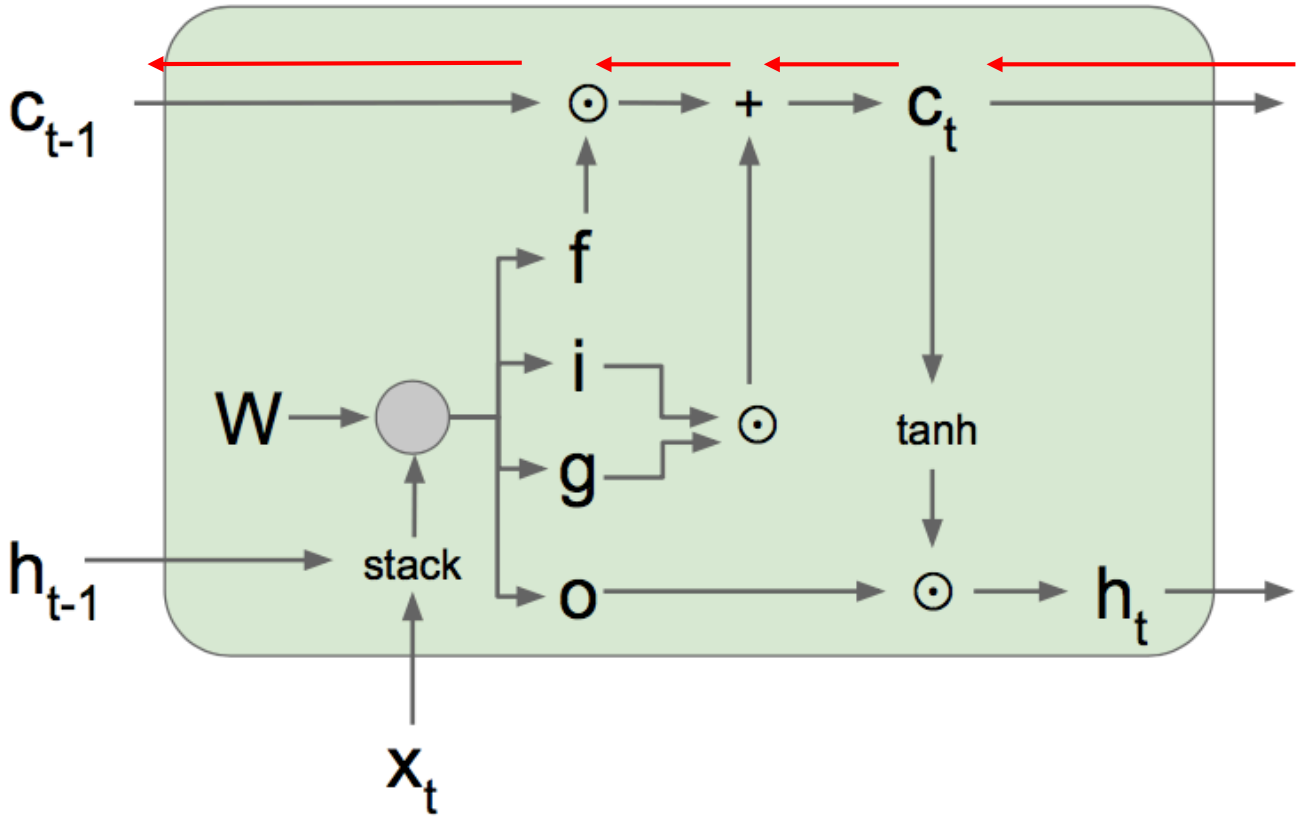
$$c_t = f \odot c_{t-1} + i \odot g$$

$$h_t = o \odot \tanh(c_t)$$

Hochreiter S, Schmidhuber J. Long short-term memory. Neural computation. 1997.

LSTM：反向传播梯度流

回传部分局部梯度：和f有关！



i, f, g的梯度：从c_t向c_{t-1}反向传播梯度流，局部梯度为向量f

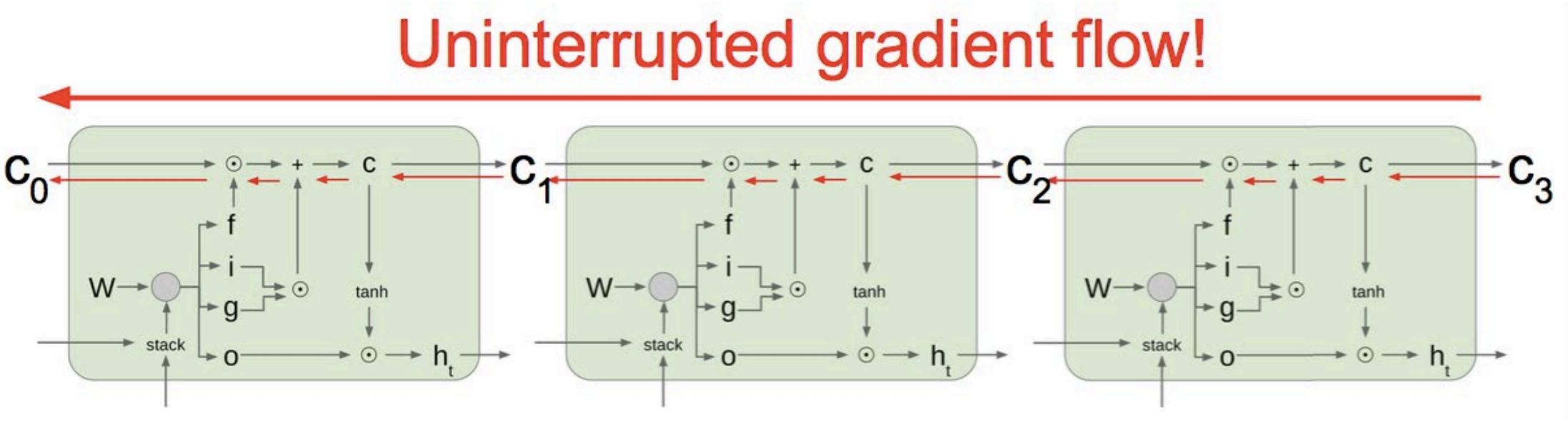
$$\begin{pmatrix} i \\ f \\ o \\ g \end{pmatrix} = \begin{pmatrix} \sigma \\ \sigma \\ \sigma \\ \tanh \end{pmatrix} W \begin{pmatrix} h_{t-1} \\ x_t \end{pmatrix}$$

$$c_t = f \odot c_{t-1} + i \odot g$$

$$h_t = o \odot \tanh(c_t)$$

Hochreiter S, Schmidhuber J. Long short-term memory. Neural computation. 1997.

LSTM：反向传播梯度流



Gated recurrent unit （GRU）

$$h_j^{\langle t \rangle} = z_j h_j^{\langle t-1 \rangle} + (1 - z_j)\tilde{h}_j^{\langle t \rangle},$$

$$\tilde{h}_j^{\langle t \rangle} = \tanh\left([\mathbf{W} e(\mathbf{x}_t)]_j + [\mathbf{U}\left(\mathbf{r} \odot \mathbf{h}_{\langle t-1 \rangle}\right)]_j \right)$$

更新门

$$z_j = \sigma\left([\mathbf{W}_z e(\mathbf{x}_t)]_j + [\mathbf{U}_z \mathbf{h}_{\langle t-1 \rangle}]_j \right),$$

重置门

$$r_j = \sigma\left([\mathbf{W}_r e(\mathbf{x}_t)]_j + [\mathbf{U}_r \mathbf{h}_{\langle t-1 \rangle}]_j \right).$$

Cho K, et al.. Learning phrase representations using RNN encoder-decoder for statistical machine translation. 2014.

小结

- RNN的结构
  - ✓一对多，多对一，多对多
- RNN的应用：图像描述
- RNN的缺点：梯度爆炸和梯度消失
  - ✓梯度裁剪
  - ✓LSTM，GRU

L11

- 生成模型