EECS 598 Homework 2

Chenchen Ma

March 5, 2019

1 Transfer Learning [1]

- Performance of pre-trained model without finetuning: the validation accuracy is 57.5163%.
- Performance of pre-trained model with finetuning: the best validation accuracy is 92.1569%.

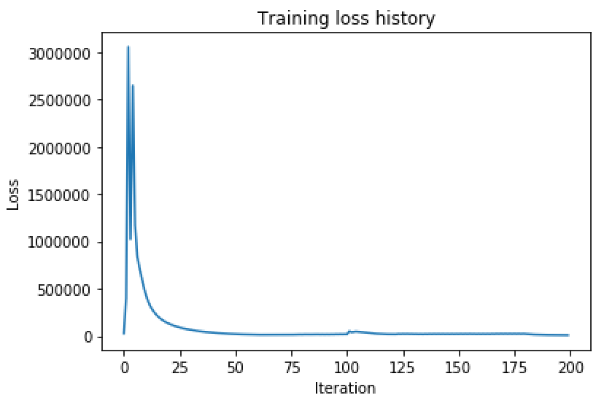
Freeze the parameters in pre-trained model and train the final fc layer

- Performance of pre-trained model without finetuning: the validation accuracy is 41.8301%.
- Performance of pre-trained model with finetuning: the best validation accuracy is 96.0784%.

2 Style Transfer

2.1 Composition VII + Tubingen

(a) Generated Image

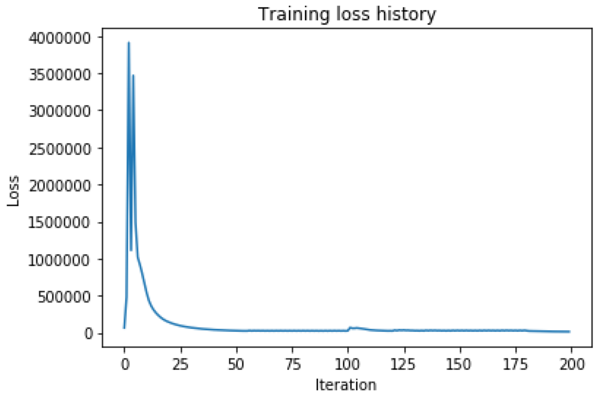


(b) Loss History

[1] For problem 1, I referred to the official pytorch tutorial https://pytorch.org/tutorials/beginner/transfer_learning_tutorial.html.

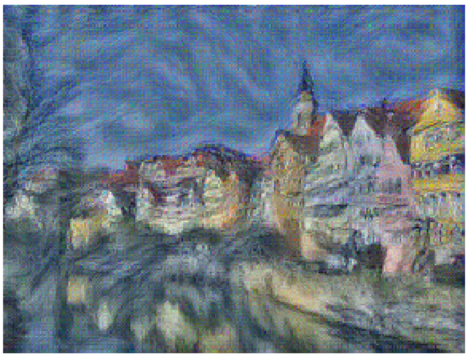
2.2 Scream + Tubingen

(c) Generated Image

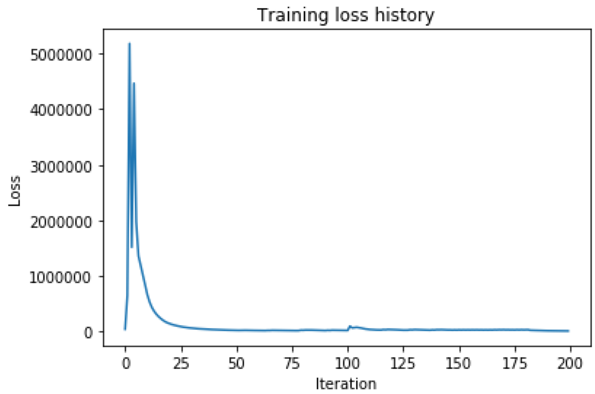


(d) Loss History

2.3 Starry Night + Tubingen



(e) Generated Image



(f) Loss History

5 Application to Image Captioning

RNN

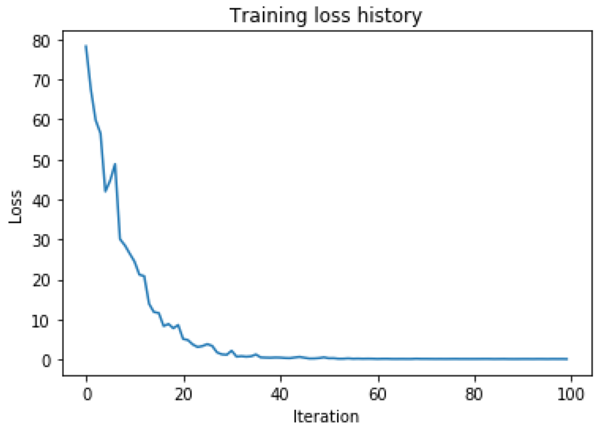


Figure 1: Training Loss for RNN

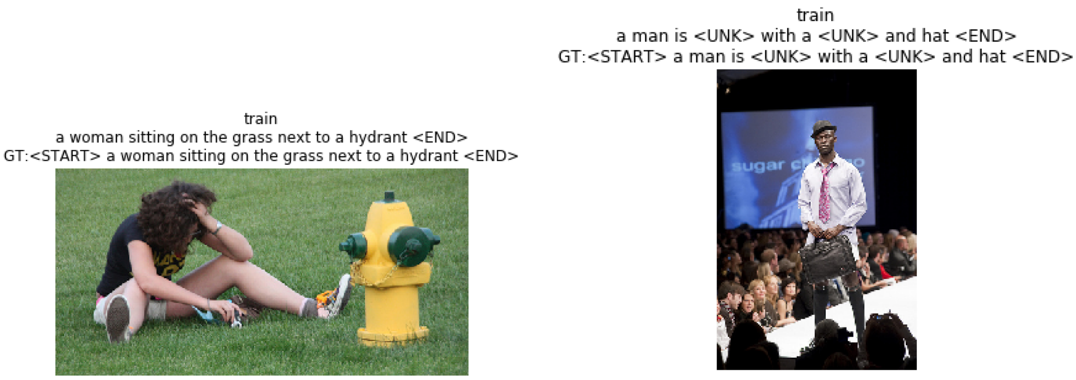


Figure 2: Results on Train Set

Figure 3: Results on Validation Set

LSTM

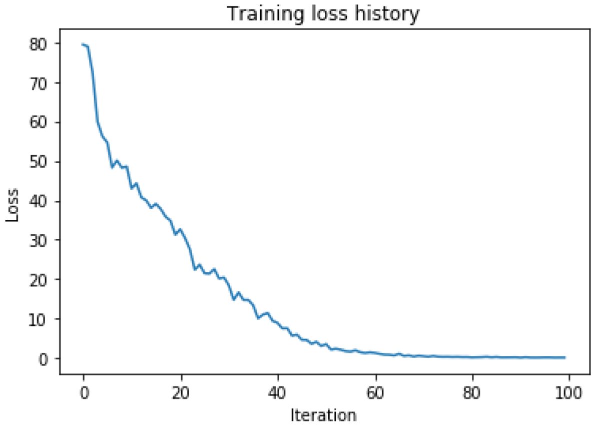


Figure 4: Training Loss for LSTM

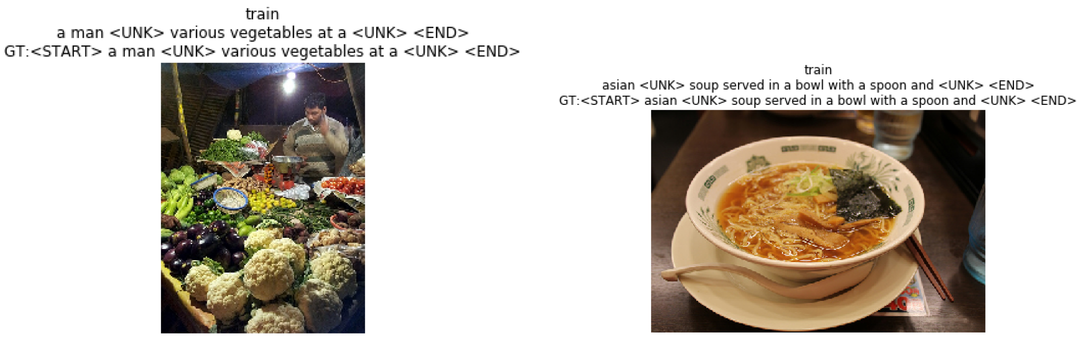


Figure 5: Results on Train Set

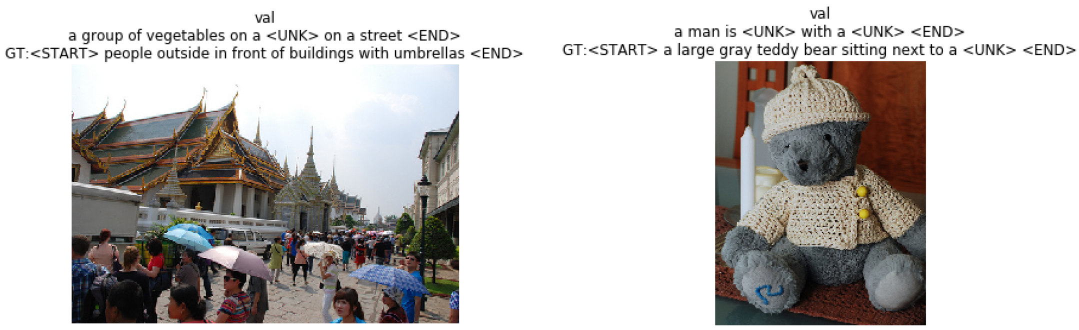


Figure 6: Results on Validation Set

6 Application to Text Classification

6.1

Bag of Words $\rightarrow$ *Linear* $\rightarrow$ *Sigmoid*

Train accuracy: 98.07%

Development accuracy: 96.06%

Test accuracy: 95.45%

6.2

Word Embedding $\rightarrow$ *AveragePooling* $\rightarrow$ *Linear* $\rightarrow$ *Sigmoid*

Train accuracy: 98.015%

Development accuracy: 95.83%

Test accuracy: 95.15%

6.3[2]

Word Embedding with GloVe $\rightarrow$ *AveragePooling* $\rightarrow$ *Linear* $\rightarrow$ *Sigmoid*

Train accuracy: 98.21%

Development accuracy: 95.79%

Test accuracy: 95.29%

6.4

Word Embedding with GloVe $\rightarrow$ *RNN* $\rightarrow$ *Linear* $\rightarrow$ *Sigmoid*

Train accuracy: 99.10%

Development accuracy: 94.17%

Test accuracy: 94.39%

6.5

Word Embedding with GloVe $\rightarrow$ *LSTM* $\rightarrow$ *Linear* $\rightarrow$ *Sigmoid*

Train accuracy: 99.6925%

Development accuracy: 95.97%

Test accuracy: 95.79%

[2] To import pre-trained GloVe embedding, I referred to the online tutorial https://medium.com/@martinpella/how-to-use-pre-trained-word-embeddings-in-pytorch-71ca59249f76

$\boxed{3.2}$ Derive $\frac{\partial L}{\partial x_t}$, $\frac{\partial L}{\partial W_x}$, $\frac{\partial L}{\partial h_{t-1}}$, $\frac{\partial L}{\partial W_h}$, $\frac{\partial L}{\partial b}$ in terms of $\frac{\partial L}{\partial h_t}$.

Solutions: $h_t = \tanh(W_x x_t + W_h h_{t-1} + b)$

$h_t \in \mathbb{R}^m$, $W_x \in \mathbb{R}^{m\times d}$, $x_t \in \mathbb{R}^d$, $W_h \in \mathbb{R}^{m\times m}$, $b \in \mathbb{R}^m$

$\tanh'(x) = 1 - \tanh^2(x)$

$$\boxed{\frac{\partial L}{\partial x_t}} = \frac{\partial L}{\partial h_t}\frac{\partial h_t}{\partial x_t} = W_x^T\left(\frac{\partial L}{\partial h_t} \odot (1-h_t^2)\right)$$

$$\boxed{\frac{\partial L}{\partial W_x}} = \frac{\partial L}{\partial h_t}\frac{\partial h_t}{\partial W_x} = \left(\frac{\partial L}{\partial h_t} \odot (1-h_t^2)\right) x_t^T$$

$$\boxed{\frac{\partial L}{\partial h_{t-1}}} = W_h^T \frac{\partial L}{\partial h_t} \odot (1-h_t^2)$$

$$\boxed{\frac{\partial L}{\partial W_h}} = \frac{\partial L}{\partial h_t} \odot (1-h_t^2)\, h_{t-1}^T$$

$$\boxed{\frac{\partial L}{\partial b}} = \frac{\partial L}{\partial h_t} \odot (1-h_t^2)$$

I assume all the gradients are column vectors.

$\boxed{3.4}$ Derive $\frac{\partial L}{\partial x_t}$ $(\forall 1 \le t \le T)$, $\frac{\partial L}{\partial W_x}$, $\frac{\partial L}{\partial h_0}$, $\frac{\partial L}{\partial W_h}$, $\frac{\partial L}{\partial b}$ in terms of $\frac{\partial L}{\partial h_t}$.

Solutions: Since gradients flow through all unrolled steps in RNN, we can define the aggregated Loss as in the lecture,

$$\mathcal{L}_t = \sum_{\tau=t}^{T} D(y_\tau, \hat{y}_\tau)$$

where $D(y_\tau, \hat{y}_\tau)$ is the loss computed at step τ.

The notations are pretty confusing in this question. According to the answer on Piazza, we should interprete $\frac{\partial L}{\partial x_t}$ as $\frac{\partial \mathcal{L}_t}{\partial x_t}$, and $\frac{\partial L}{\partial h_t}$ is actually $\frac{\partial D(y_t, \hat{y}_t)}{\partial h_t}$.

We can derive the gradients recursively.

① First we have

$$\boxed{\frac{\partial h_t}{\partial h_{t-1}}} = \mathrm{diag}(1 - h_t^2)\, W_h.$$

$$\frac{\partial \mathcal{L}_t}{\partial h_{t,m}} = \frac{\partial L}{\partial h_{t,m}} + \sum_{m'} \frac{\partial h_{t+1,m'}}{\partial h_{t,m}} \frac{\partial \mathcal{L}_{t+1}}{\partial h_{t+1,m'}}$$

$$\Rightarrow \boxed{\frac{\partial \mathcal{L}_t}{\partial h_t}} = \frac{\partial L}{\partial h_t} + \sum_{\tau=t+1}^{T} \left(\prod_{r=t+1}^{\tau} \mathrm{diag}(1-h_r^2) W_h \right) \frac{\partial L}{\partial h_\tau}$$

② Based on ①, we have

$$\boxed{\frac{\partial L}{\partial W_h^{(m,m')}}} = \sum_t \frac{\partial h_{t,m}}{\partial W_h^{(m,m')}} \frac{\partial \mathcal{L}_t}{\partial h_{t,m}} = \sum_t (1 - h_{t,m}^2)\, h_{t-1,m'} \frac{\partial \mathcal{L}_t}{\partial h_{t,m}}$$

(Here I use $W_h^{(m,m')}$ to denote the (m,m') th element of W_h)

$$\Rightarrow \boxed{\frac{\partial L}{\partial W_h}} = \sum_t (1-h_t^2) \odot \frac{\partial \mathcal{L}_t}{\partial h_t}\, h_{t-1}^T$$

Similarly, we have

$$\boxed{\frac{\partial L}{\partial W_x^{(m,d)}}} = \sum_t \frac{\partial h_{t,m}}{\partial W_x^{(m,d)}} \frac{\partial \mathcal{L}_t}{\partial h_{t,m}} = \sum_t (1 - h_{t,m}^2) x_{t,d} \frac{\partial \mathcal{L}_t}{\partial h_{t,m}}$$

$$\Rightarrow \boxed{\frac{\partial L}{\partial W_x}} = \sum_t (1-h_t^2) \odot \frac{\partial \mathcal{L}_t}{\partial h_t} x_t^T$$

$$\boxed{\frac{\partial L}{\partial b_m}} = \sum_t \frac{\partial h_{t,m}}{\partial b_m} \frac{\partial \mathcal{L}_t}{\partial h_{t,m}} = \sum_t (1 - h_{t,m}^2) \frac{\partial \mathcal{L}_t}{\partial h_{t,m}}$$

$$\Rightarrow \boxed{\frac{\partial L}{\partial b}} = \sum_t (1-h_t^2) \odot \frac{\partial \mathcal{L}_t}{\partial h_t}$$

③ Similarly, we can get $\frac{\partial L}{\partial x_t}$ recursively.

$$\boxed{\frac{\partial L}{\partial x_t}} = \frac{\partial \mathcal{L}_t}{\partial x_t} = \frac{\partial \mathcal{L}_t}{\partial h_t} \frac{\partial h_t}{\partial x_t}$$

$$= W_x^T (1-h_t^2) \odot \frac{\partial \mathcal{L}_t}{\partial h_t}, \quad \forall\ 1 \le t \le T$$

$$\boxed{\frac{\partial L}{\partial h_0}} = \frac{\partial \mathcal{L}_1}{\partial h_0} = \frac{\partial \mathcal{L}_1}{\partial h_1} \frac{\partial h_1}{\partial h_0}$$

$$= W_h^T (1-h_1^2) \odot \frac{\partial \mathcal{L}_1}{\partial h_1}$$

4.2 For the simplicity of notations, all the products of two column vectors are element-wise.

$$\boxed{\frac{\partial L}{\partial c_t}} = \frac{\partial L}{\partial h_t}\frac{\partial h_t}{\partial c_t} + \frac{\partial L}{\partial c_{t+1}}\frac{\partial c_{t+1}}{\partial c_t}$$

$$= \frac{\partial L}{\partial h_t} o_t (1-\tanh^2(c_t)) + \frac{\partial L}{\partial c_{t+1}} f_{t+1}$$

Note the abuse of notations here. I use $\frac{\partial L}{\partial c_t}$ to denote the gradient of L w.r.t c_t from two paths: $L \to h_t \to c_t$ and $L \to c_{t+1} \to c_t$.

$$\boxed{\frac{\partial L}{\partial x_t}} = \frac{\partial L}{\partial f_t}\frac{\partial f_t}{\partial x_t} + \frac{\partial L}{\partial i_t}\frac{\partial i_t}{\partial x_t} + \frac{\partial L}{\partial \tilde{c}_t}\frac{\partial \tilde{c}_t}{\partial x_t} + \frac{\partial L}{\partial o_t}\frac{\partial o_t}{\partial x_t}$$

$$= \frac{\partial L}{\partial c_t}\frac{\partial c_t}{\partial f_t}\frac{\partial f_t}{\partial x_t} + \frac{\partial L}{\partial c_t}\frac{\partial c_t}{\partial i_t}\frac{\partial i_t}{\partial x_t}$$

$$+ \frac{\partial L}{\partial c_t}\frac{\partial c_t}{\partial \tilde{c}_t}\frac{\partial \tilde{c}_t}{\partial x_t} + \frac{\partial L}{\partial h_t}\frac{\partial h_t}{\partial o_t}\frac{\partial o_t}{\partial x_t}$$

$$= (W_x^f)^T \frac{\partial L}{\partial c_t} c_{t-1} f_t (1-f_t) + (W_x^i)^T \frac{\partial L}{\partial c_t} \tilde{c}_t i_t (1-i_t)$$

$$+ (W_x^c)^T \frac{\partial L}{\partial c_t} i_t (1-\tilde{c}_t^2) + (W_x^o)^T \frac{\partial L}{\partial h_t} \tanh(c_t) o_t (1-o_t)$$

$$\boxed{\frac{\partial L}{\partial h_{t-1}}} = (W_h^f)^T \frac{\partial L}{\partial c_t} c_{t-1} f_t (1-f_t) + (W_h^i)^T \frac{\partial L}{\partial c_t} \tilde{c}_t i_t (1-i_t)$$

$$+ (W_h^c)^T \frac{\partial L}{\partial c_t} i_t (1-\tilde{c}_t^2) + (W_h^o)^T \frac{\partial L}{\partial h_t} \tanh(c_t) o_t (1-o_t)$$

$$\boxed{\frac{\partial L}{\partial c_{t-1}}} = \frac{\partial L}{\partial c_t}\frac{\partial c_t}{\partial c_{t-1}} = \frac{\partial L}{\partial c_t} f_t$$

where $\frac{\partial L}{\partial c_t} = \frac{\partial L}{\partial h_t} o_t (1-\tanh^2(c_t)) + \frac{\partial L}{\partial c_{t+1}} f_{t+1}$

$$\boxed{\frac{\partial L}{\partial W_x^f}} = \frac{\partial L}{\partial C_t} \frac{\partial C_t}{\partial f_t} \frac{\partial f_t}{\partial W_x^f}$$

$$= \frac{\partial L}{\partial C_t} C_{t-1} f_t (1-f_t)\, x_t^T$$

$$\boxed{\frac{\partial L}{\partial W_h^f}} = \frac{\partial L}{\partial C_t} \frac{\partial C_t}{\partial f_t} \frac{\partial f_t}{\partial W_h^f}$$

$$= \frac{\partial L}{\partial C_t} C_{t-1} f_t (1-f_t)\, h_{t-1}^T$$

$$\boxed{\frac{\partial L}{\partial b^f}} = \frac{\partial L}{\partial C_t} C_{t-1} f_t (1-f_t)$$

where $\frac{\partial L}{\partial C_t} = \frac{\partial L}{\partial h_t} O_t (1 - \tanh^2(C_t)) + \frac{\partial L}{\partial C_{t+1}} f_{t+1}$

$$\boxed{\frac{\partial L}{\partial W_x^i}} = \frac{\partial L}{\partial C_t} \frac{\partial C_t}{\partial i_t} \frac{\partial i_t}{\partial W_x^i}$$

$$= \frac{\partial L}{\partial C_t} \tilde{C}_t\, i_t (1-i_t)\, x_t^T$$

$$\boxed{\frac{\partial L}{\partial W_h^i}} = \frac{\partial L}{\partial C_t} \frac{\partial C_t}{\partial i_t} \frac{\partial i_t}{\partial W_h^i}$$

$$= \frac{\partial L}{\partial C_t} \tilde{C}_t\, i_t (1-i_t)\, h_{t-1}^T$$

$$\boxed{\frac{\partial L}{\partial b^i}} = \frac{\partial L}{\partial C_t} \frac{\partial C_t}{\partial i_t} \frac{\partial i_t}{\partial b^i}$$

$$= \frac{\partial L}{\partial C_t} \tilde{C}_t\, i_t (1-i_t)$$

where $\frac{\partial L}{\partial C_t} = \frac{\partial L}{\partial h_t} O_t (1 - \tanh^2(C_t)) + \frac{\partial L}{\partial C_{t+1}} f_{t+1}$

$$\boxed{\frac{\partial L}{\partial W_x^c}} = \frac{\partial L}{\partial C_t} \frac{\partial C_t}{\partial \tilde{C}_t} \frac{\partial \tilde{C}_t}{\partial W_x^c}$$

$$= \frac{\partial L}{\partial C_t} i_t (1 - \tilde{C}_t^2) x_t^T$$

$$\boxed{\frac{\partial L}{\partial W_h^c}} = \frac{\partial L}{\partial C_t} \frac{\partial C_t}{\partial \tilde{C}_t} \frac{\partial \tilde{C}_t}{\partial W_h}$$

$$= \frac{\partial L}{\partial C_t} i_t (1 - \tilde{C}_t^2) h_{t-1}^T$$

$$\boxed{\frac{\partial L}{\partial b^c}} = \frac{\partial L}{\partial C_t} i_t (1 - \tilde{C}_t^2)$$

where $\frac{\partial L}{\partial C_t} = \frac{\partial L}{\partial h_t} O_t (1 - \tanh^2(C_t)) + \frac{\partial L}{\partial C_{t+1}} f_{t+1}$

$$\boxed{\frac{\partial L}{\partial W_x^o}} = \frac{\partial L}{\partial h_t} \frac{\partial h_t}{\partial O_t} \frac{\partial O_t}{\partial W_x^o}$$

$$= \frac{\partial L}{\partial h_t} \tanh(C_t) O_t (1 - O_t) x_t^T$$

$$\boxed{\frac{\partial L}{\partial W_h^o}} = \frac{\partial L}{\partial h_t} \frac{\partial h_t}{\partial O_t} \frac{\partial O_t}{\partial W_h^o}$$

$$= \frac{\partial L}{\partial h_t} \tanh(C_t) O_t (1 - O_t) h_{t-1}^T$$

$$\boxed{\frac{\partial L}{\partial b^o}} = \frac{\partial L}{\partial h_t} \frac{\partial h_t}{\partial O_t} \frac{\partial O_t}{\partial b^o}$$

$$= \frac{\partial L}{\partial h_t} \tanh(C_t) O_t (1 - O_t)$$

$\boxed{4.4}$ Similar to 3.4, we can define the loss from t up to T:

$$\mathcal{L}_t = \sum_{\tau=t}^{T} D(y_\tau, \hat{y}_\tau)$$

① First we have

$$\boxed{\frac{\partial h_t}{\partial h_{t-1}}} = \frac{\partial h_t}{\partial o_t}\frac{\partial o_t}{\partial h_{t-1}} + \frac{\partial h_t}{\partial f_t}\frac{\partial f_t}{\partial h_{t-1}} + \frac{\partial h_t}{\partial i_t}\frac{\partial i_t}{\partial h_{t-1}} + \frac{\partial h_t}{\partial \tilde{C}_t}\frac{\partial \tilde{C}_t}{\partial h_{t-1}}$$

$$= \mathrm{diag}(\tanh(C_t) o_t (1-o_t)) W_h^o + \mathrm{diag}(o_t(1-\tanh^2(C_t)) C_{t-1} f_t (1-f_t)) W_h^f$$

$$+ \mathrm{diag}(o_t(1-\tanh^2(C_t)) \tilde{C}_t i_t (1-i_t)) W_h^i + \mathrm{diag}(o_t(1-\tanh^2(C_t)) i_t (1-\tilde{C}_t^2)) W_h^c$$

$$\boxed{\frac{\partial \mathcal{L}_t}{\partial h_t}} = \frac{\partial L}{\partial h_t} + \frac{\partial h_{t+1}}{\partial h_t}\frac{\partial \mathcal{L}_{t+1}}{\partial h_{t+1}}$$

Therefore, we can derive $\frac{\partial \mathcal{L}_t}{\partial h_t}$ recursively based on $\frac{\partial L}{\partial h_t}$, $t=1,\cdots,T$.

② We can also compute $\frac{\partial \mathcal{L}_t}{\partial C_t}$

$$\frac{\partial \mathcal{L}_T}{\partial C_T} = \frac{\partial L}{\partial h_T} o_T (1-\tanh^2(C_T))$$

$$\boxed{\frac{\partial \mathcal{L}_t}{\partial C_t}} = \frac{\partial L}{\partial h_t}\frac{\partial h_t}{\partial C_t} + \frac{\partial \mathcal{L}_{t+1}}{\partial C_{t+1}} f_{t+1}$$

$$= \frac{\partial L}{\partial h_t} o_t (1-\tanh^2(C_t)) + \sum_{\tau=t+1}^{T} \frac{\partial L}{\partial h_\tau} \prod_{r=t+1}^{\tau} [o_r (1-\tanh^2(C_r)) f_r]$$

③ From ① and ②, we have derived $\frac{\partial \mathcal{L}_t}{\partial h_t}$ and $\frac{\partial \mathcal{L}_t}{\partial C_t}$, $\forall\ 1 \le t \le T$.

$$\boxed{\frac{\partial L}{\partial W_x^f}} = \sum_t \frac{\partial \mathcal{L}_t}{\partial C_t}\frac{\partial C_t}{\partial f_t}\frac{\partial f_t}{\partial W_x^f}$$

$$= \sum_t \frac{\partial \mathcal{L}_t}{\partial C_t} C_{t-1} f_t (1-f_t) x_t^T$$

Similarly, we have

$$\boxed{\frac{\partial L}{\partial W_h^f}} = \sum_t \frac{\partial \mathcal{L}_t}{\partial C_t} C_{t-1} f_t (1-f_t) h_{t-1}^T$$

$$\boxed{\frac{\partial L}{\partial b^f}} = \sum_t \frac{\partial \mathcal{L}_t}{\partial C_t} C_{t-1} f_t (1-f_t)$$

$$\boxed{\frac{\partial L}{\partial W_x^i}} = \sum_t \frac{\partial \mathcal{L}_t}{\partial C_t} \frac{\partial C_t}{\partial i_t} \frac{\partial i_t}{\partial W_x^i}$$

$$= \sum_t \frac{\partial \mathcal{L}_t}{\partial C_t} \tilde{C}_t \, i_t (1 - i_t) x_t^T$$

Similarly, $$\boxed{\frac{\partial L}{\partial W_h^i}} = \sum_t \frac{\partial \mathcal{L}_t}{\partial C_t} \tilde{C}_t \, i_t (1 - i_t) h_{t-1}^T$$

$$\boxed{\frac{\partial L}{\partial b^i}} = \sum_t \frac{\partial \mathcal{L}_t}{\partial C_t} \tilde{C}_t \, i_t (1 - i_t)$$

$$\boxed{\frac{\partial L}{\partial W_x^c}} = \sum_t \frac{\partial \mathcal{L}_t}{\partial C_t} i_t \left(1 - \tilde{C}_t^2\right) x_t^T$$

$$\boxed{\frac{\partial L}{\partial W_h^c}} = \sum_t \frac{\partial \mathcal{L}_t}{\partial C_t} i_t \left(1 - \tilde{C}_t^2\right) h_{t-1}^T$$

$$\boxed{\frac{\partial L}{\partial b^c}} = \sum_t \frac{\partial \mathcal{L}_t}{\partial C_t} i_t \left(1 - \tilde{C}_t^2\right)$$

$$\boxed{\frac{\partial L}{\partial W_x^o}} = \sum_t \frac{\partial \mathcal{L}_t}{\partial h_t} \frac{\partial h_t}{\partial O_t} \frac{\partial O_t}{\partial W_x^o}$$

$$= \sum_t \frac{\partial \mathcal{L}_t}{\partial h_t} \tanh(C_t) \, O_t (1 - O_t) x_t^T$$

$$\boxed{\frac{\partial L}{\partial W_h^o}} = \sum_t \frac{\partial \mathcal{L}_t}{\partial h_t} \tanh(C_t) \, O_t (1 - O_t) h_{t-1}^T$$

$$\boxed{\frac{\partial L}{\partial b^o}} = \sum_t \frac{\partial \mathcal{L}_t}{\partial h_t} \tanh(C_t) \, O_t (1 - O_t)$$

④ Now let's derive $\frac{\partial L}{\partial x_t}$ and $\frac{\partial L}{\partial h_0}$.

$$\boxed{\frac{\partial L}{\partial x_t}} = \frac{\partial \mathcal{L}_t}{\partial x_t} = \frac{\partial \mathcal{L}_t}{\partial h_t}\frac{\partial h_t}{\partial O_t}\frac{\partial O_t}{\partial x_t} + \frac{\partial \mathcal{L}_t}{\partial C_t}\frac{\partial C_t}{\partial x_t}$$

$$= \frac{\partial \mathcal{L}_t}{\partial h_t}\frac{\partial h_t}{\partial O_t}\frac{\partial O_t}{\partial x_t} + \frac{\partial \mathcal{L}_t}{\partial C_t}\frac{\partial C_t}{\partial f_t}\frac{\partial f_t}{\partial x_t}$$

$$+ \frac{\partial \mathcal{L}_t}{\partial C_t}\frac{\partial C_t}{\partial i_t}\frac{\partial i_t}{\partial x_t} + \frac{\partial \mathcal{L}_t}{\partial C_t}\frac{\partial C_t}{\partial \tilde{C}_t}\frac{\partial \tilde{C}_t}{\partial x_t}$$

$$= W_x^{oT}\left(\frac{\partial \mathcal{L}_t}{\partial h_t}\tanh(C_t)O_t(1-O_t)\right) + W_x^{f\,T}\left(\frac{\partial \mathcal{L}_t}{\partial C_t}C_{t-1}f_t(1-f_t)\right)$$

$$+ W_x^{i\,T}\left(\frac{\partial \mathcal{L}_t}{\partial C_t}\tilde{C}_t i_t(1-i_t)\right) + W_x^{c\,T}\left(\frac{\partial \mathcal{L}_t}{\partial C_t} i_t(1-\tilde{C}_t^2)\right).$$

$$\boxed{\frac{\partial L}{\partial h_0}} = \frac{\partial \mathcal{L}_1}{\partial h_1}\frac{\partial h_1}{\partial O_1}\frac{\partial O_1}{\partial h_0} + \frac{\partial \mathcal{L}_1}{\partial C_1}\frac{\partial C_1}{\partial f_1}\frac{\partial f_1}{\partial h_0}$$

$$+ \frac{\partial \mathcal{L}_1}{\partial C_1}\frac{\partial C_1}{\partial i_1}\frac{\partial i_1}{\partial h_0} + \frac{\partial \mathcal{L}_1}{\partial C_1}\frac{\partial C_1}{\partial \tilde{C}_1}\frac{\partial \tilde{C}_1}{\partial h_0}$$

$$= W_h^{o\,T}\left(\frac{\partial \mathcal{L}_1}{\partial h_1}\tanh(C_1)O_1(1-O_1)\right) + W_h^{f\,T}\left(\frac{\partial \mathcal{L}_1}{\partial C_1}C_0 f_1(1-f_1)\right)$$

$$+ W_h^{i\,T}\left(\frac{\partial \mathcal{L}_1}{\partial C_1}\tilde{C}_1 i_1(1-i_1)\right) + W_h^{c\,T}\left(\frac{\partial \mathcal{L}_1}{\partial C_1} i_1(1-\tilde{C}_1^2)\right).$$