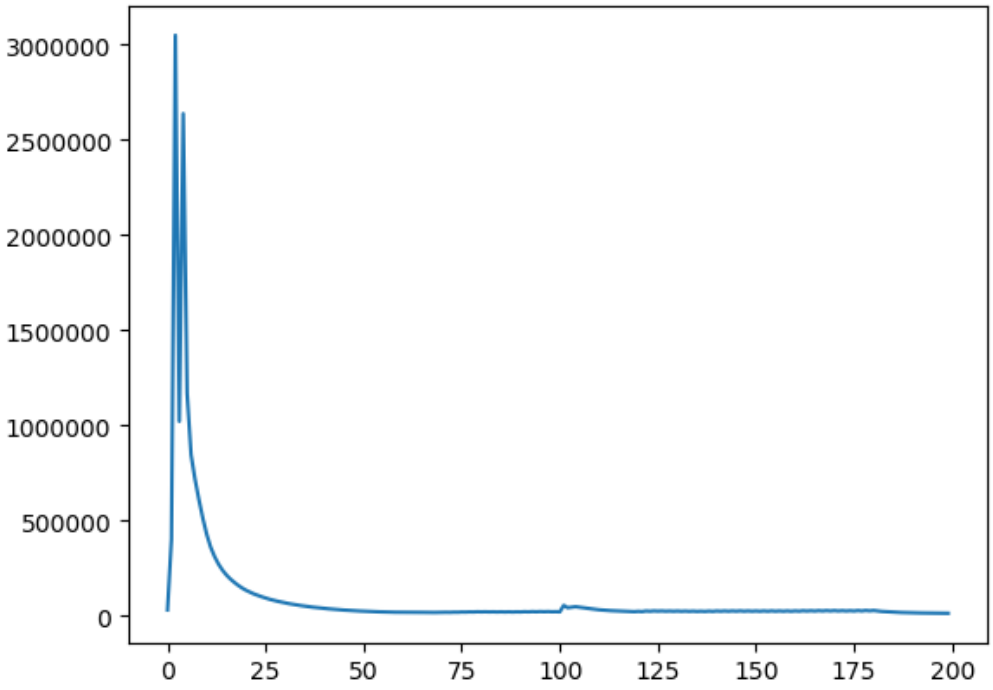
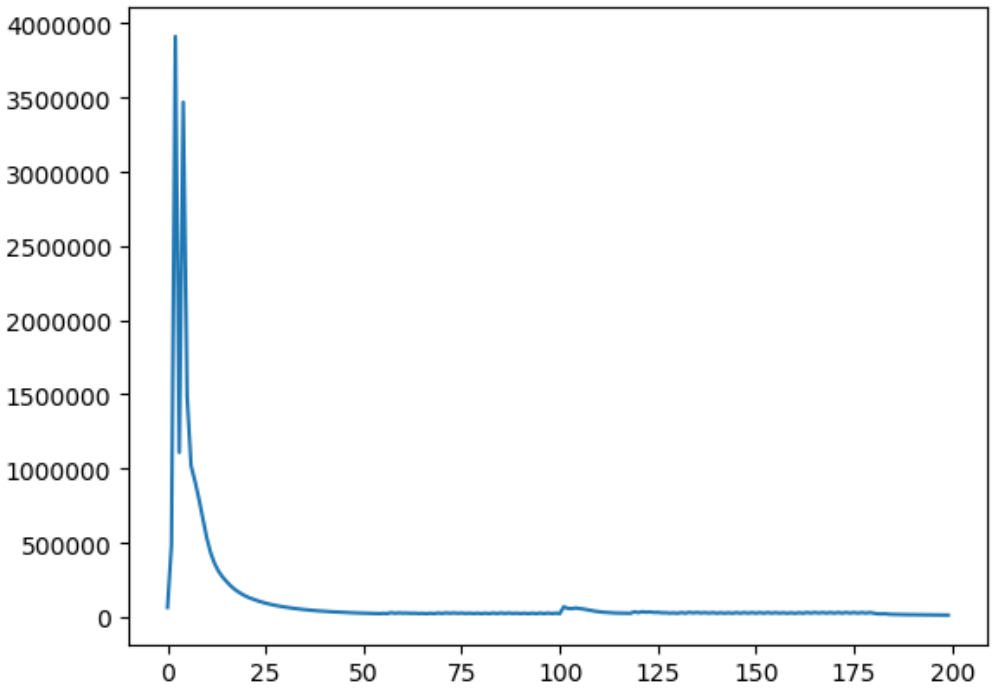
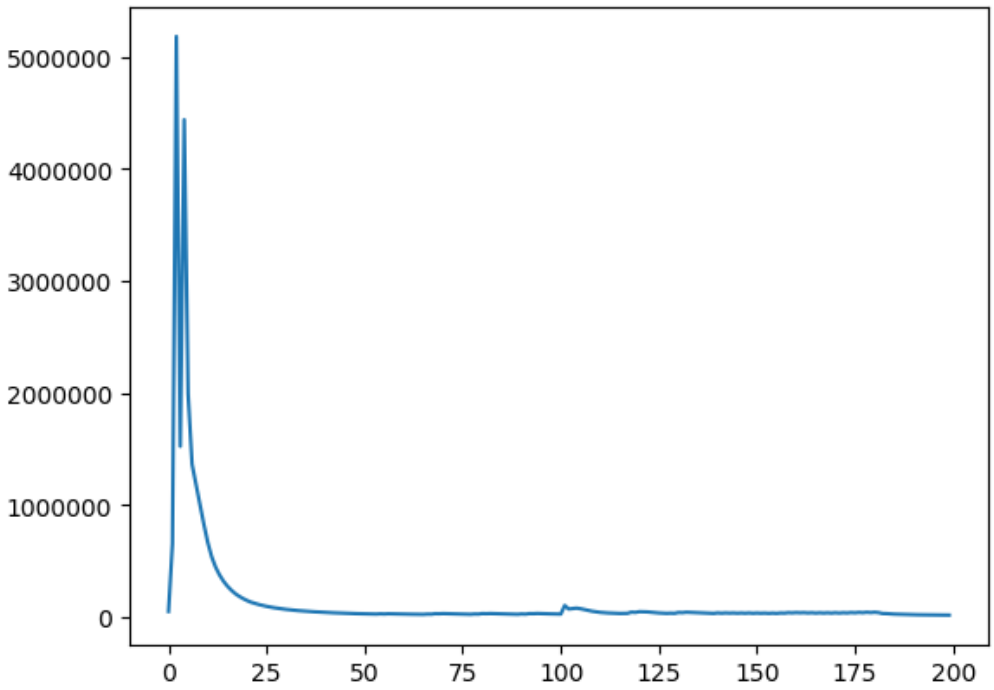
1 . Transfer Learning

Validation Accuracy1: 0.947712
Validation Accuracy2: 0.960784

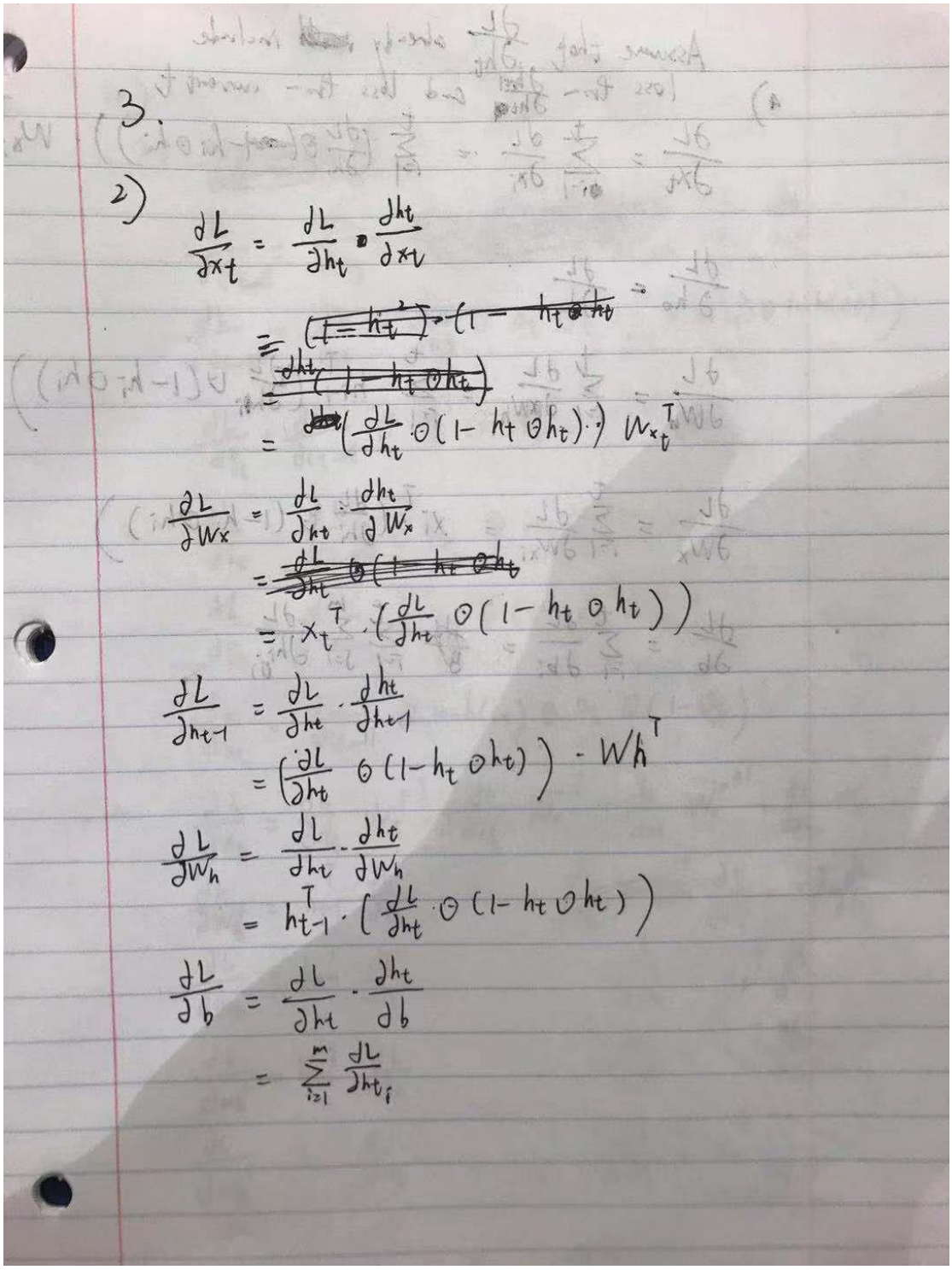
2. Style Transfer




4000000
3500000
3000000
2500000
2000000
1500000
1000000
500000
0
0
25
50
75
100
125
150
175
200


5000000
4000000
3000000
2000000
1000000
0
0
25
50
75
100
125
150
175
200

3. RNN


3.
2)
$\frac{\partial L}{\partial x_t} = \frac{\partial L}{\partial h_t} \cdot \frac{\partial h_t}{\partial x_t}$
$= \frac{\partial h_t}{\partial (1 - h_t \odot h_t)} = \left(\frac{\partial L}{\partial h_t} \odot (1 - h_t \odot h_t)\right) W_{x_t}^T$
$\frac{\partial L}{\partial W_x} = \frac{\partial L}{\partial h_t} \cdot \frac{\partial h_t}{\partial W_x}$
$= x_t^T \cdot \left(\frac{\partial L}{\partial h_t} \odot (1 - h_t \odot h_t)\right)$
$\frac{\partial L}{\partial h_{t-1}} = \frac{\partial L}{\partial h_t} \cdot \frac{\partial h_t}{\partial h_{t-1}}$
$= \left(\frac{\partial L}{\partial h_t} \odot (1 - h_t \odot h_t)\right) \cdot W h^T$
$\frac{\partial L}{\partial W_h} = \frac{\partial L}{\partial h_t} \cdot \frac{\partial h_t}{\partial W_h}$
$= h_{t-1}^T \cdot \left(\frac{\partial L}{\partial h_t} \odot (1 - h_t \odot h_t)\right)$
$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial h_t} \cdot \frac{\partial h_t}{\partial b}$
$= \sum_{i=1}^{m} \frac{\partial L}{\partial h_{t_i}}$

4) Assume that $\frac{\partial L}{\partial h_t}$ already include loss from $\frac{\partial L}{\partial h_{t+1}}$ and loss from current t

$$\frac{\partial L}{\partial x_t} = \sum_{i=1}^{t} \frac{\partial L}{\partial x_i} = \sum_{i=1}^{t} \left(\frac{\partial L}{\partial h_i} \odot (1 - h_i \odot h_i) \right) \cdot W_{x_i}^{T}$$

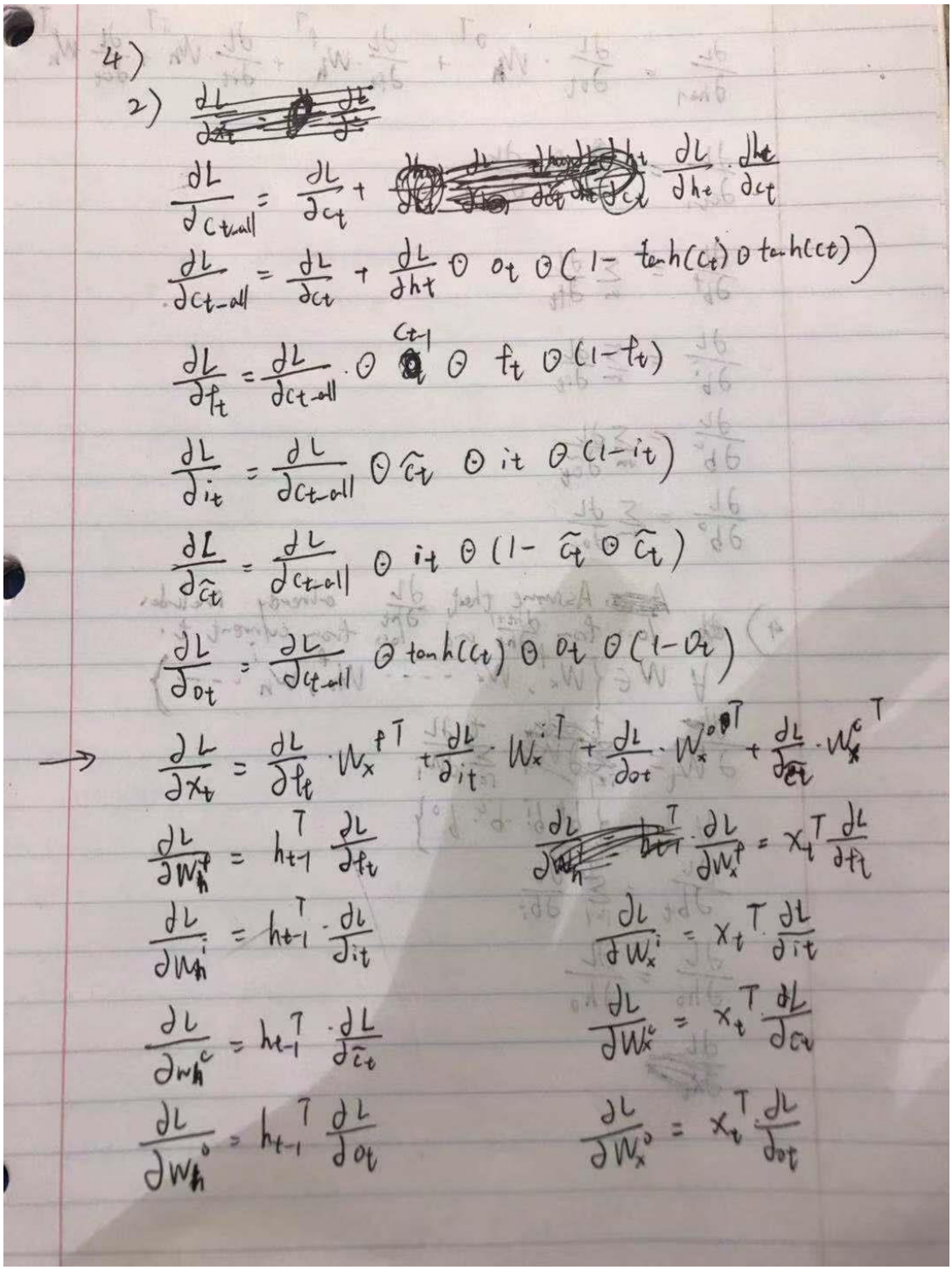
$$\frac{\partial L}{\partial h_0} = \frac{\partial L}{\partial h_0}$$

$$\frac{\partial L}{\partial W_h} = \sum_{i=1}^{t} \frac{\partial L}{\partial W_{h_i}} = \sum_{i=1}^{t} h_{i-1}^{T} \left(\frac{\partial L}{\partial h_i} \odot (1 - h_i \odot h_i) \right)$$

$$\frac{\partial L}{\partial W_x} = \sum_{i=1}^{t} \frac{\partial L}{\partial W_{x_i}} = x_i^{T} \cdot \left(\frac{\partial L}{\partial h_i} \odot (1 - h_i \odot h_i) \right)$$

$$\frac{\partial L}{\partial b} = \sum_{i=1}^{t} \frac{\partial L}{\partial b_i} = \sum_{i=1}^{t} \sum_{j=1}^{m} \frac{\partial L}{\partial h_{ij}}$$

4. LSTM



4)

2) $\frac{\partial L}{\partial c_{t-all}} = \frac{\partial L}{\partial c_t} + \frac{\partial L}{\partial h_t} \cdot \frac{\partial h_t}{\partial c_t}$

$\frac{\partial L}{\partial c_{t-all}} = \frac{\partial L}{\partial c_t} + \frac{\partial L}{\partial h_t} \odot o_t \odot (1 - \tanh(c_t) \odot \tanh(c_t))$

$\frac{\partial L}{\partial f_t} = \frac{\partial L}{\partial c_{t-all}} \odot c_{t-1} \odot f_t \odot (1 - f_t)$

$\frac{\partial L}{\partial i_t} = \frac{\partial L}{\partial c_{t-all}} \odot \hat{c}_t \odot i_t \odot (1 - i_t)$

$\frac{\partial L}{\partial \hat{c}_t} = \frac{\partial L}{\partial c_{t-all}} \odot i_t \odot (1 - \hat{c}_t \odot \hat{c}_t)$

$\frac{\partial L}{\partial o_t} = \frac{\partial L}{\partial c_{t-all}} \odot \tanh(c_t) \odot o_t \odot (1 - o_t)$

→ $\frac{\partial L}{\partial x_t} = \frac{\partial L}{\partial f_t} \cdot W_x^{f\,T} + \frac{\partial L}{\partial i_t} \cdot W_x^{i\,T} + \frac{\partial L}{\partial o_t} \cdot W_x^{o\,T} + \frac{\partial L}{\partial \hat{c}_t} \cdot W_x^{c\,T}$

$\frac{\partial L}{\partial W_h^f} = h_{t-1}^T \frac{\partial L}{\partial f_t}$ $\qquad$ $\frac{\partial L}{\partial W_x^f} = x_t^T \frac{\partial L}{\partial f_t}$

$\frac{\partial L}{\partial W_h^i} = h_{t-1}^T \cdot \frac{\partial L}{\partial i_t}$ $\qquad$ $\frac{\partial L}{\partial W_x^i} = x_t^T \frac{\partial L}{\partial i_t}$

$\frac{\partial L}{\partial W_h^c} = h_{t-1}^T \cdot \frac{\partial L}{\partial \hat{c}_t}$ $\qquad$ $\frac{\partial L}{\partial W_x^c} = x_t^T \frac{\partial L}{\partial c_t}$

$\frac{\partial L}{\partial W_h^o} = h_{t-1}^T \frac{\partial L}{\partial o_t}$ $\qquad$ $\frac{\partial L}{\partial W_x^o} = x_t^T \cdot \frac{\partial L}{\partial o_t}$

$$\frac{\partial L}{\partial h_{t-1}} = \frac{\partial L}{\partial o_t} \cdot W_h^{o\top} + \frac{\partial L}{\partial f_t} \cdot W_h^{f\top} + \frac{\partial L}{\partial i_t} \cdot W_h^{i\top} + \frac{\partial L}{\partial \tilde{c}_t} W_h^{c\top}$$

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

$$\frac{\partial L}{\partial b^f} = \sum_m \frac{\partial L}{\partial f_t}$$

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

$$\frac{\partial L}{\partial b^c} = \sum_m \frac{\partial L}{\partial \tilde{c}_t}$$

$$\frac{\partial L}{\partial b^o} = \sum_m \frac{\partial L}{\partial o_t}$$

4) Assume that $\frac{\partial L}{\partial h_t}$ already includes loss from $\frac{\partial h_{t+1}}{\partial h_t}$ and loss from current t.

$$\forall W \in \{W_x^f, W_x^i, \dots, W_h^f, W_h^i, \dots\}$$

$$\frac{\partial L}{\partial W_t} = \sum_{i=1}^{t} \frac{\partial L}{\partial W_i}$$

$$\forall b \in \{b^f, b^i, b^c, b^o\}$$

$$\frac{\partial L}{\partial b_t} = \sum_{i=1}^{t} \frac{\partial L}{\partial b_i}$$

$$\frac{\partial L}{\partial h_0} = \frac{\partial L}{\partial h_0}$$

$$\frac{\partial L}{\partial x_t}$$

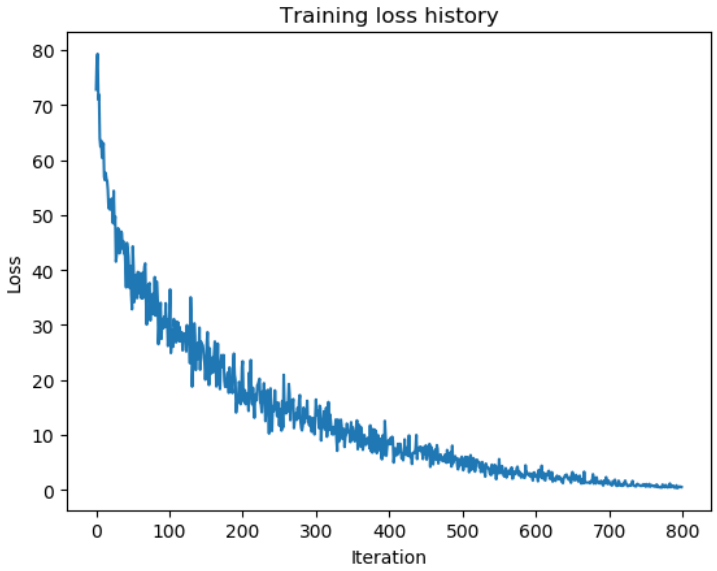
5. Application to Image Captioning

RNN

Average BLEU score for train: 0.230041

Average BLEU score for val: 0.210692

On train size 1000

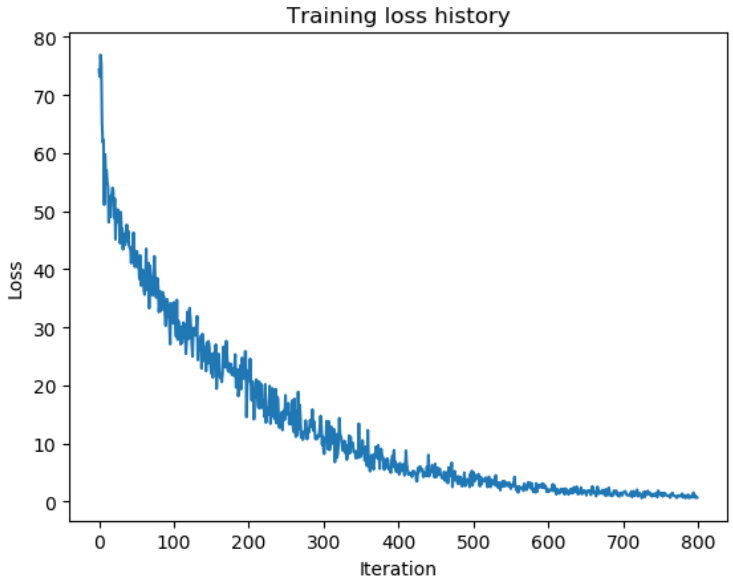


LSTM

Average BLEU score for train: 0.212054

Average BLEU score for val: 0.214747

On train size 1000



train
a man with his back <UNK> walking down the sidewalk <END>
GT:<START> a man with his back <UNK> walking down the sidewalk <END

train
a big parked train as people walk by it <END>
GT:<START> a big parked train as people walk by it <END>

train
a close up of some pizza on a pan <END>
GT:<START> a close up of some pizza on a pan <END>

train
an open outdoor market has lots of bananas <END>
GT:<START> an open outdoor market has lots of bananas <END>

train
the man in the hat is holding up a sign <END>
GT:<START> the man in the hat is holding up a sign <END>

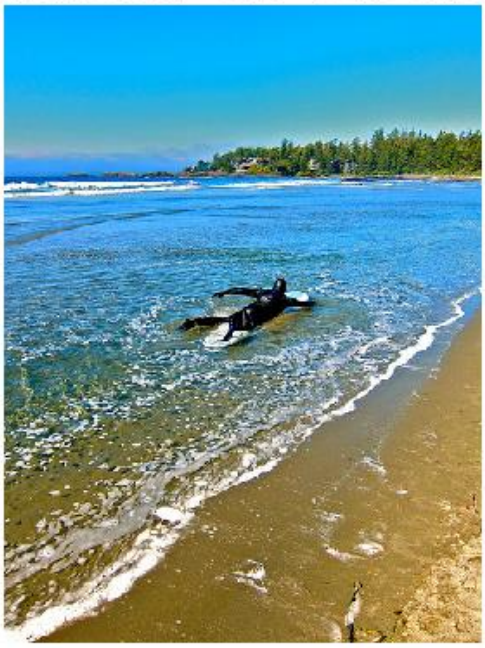
val
desk and a picture of a laptop with <UNK> <UNK> <UNK> <END>
GT:<START> a white plate of food on a table <END>

val
two giraffes standing in front of a building <END>
GT:<START> zebras eat from a <UNK> set up in their zoo enclosure <END>

val
a very <UNK> bus on a <UNK> <UNK> <END>
GT:<START> lots of boats <UNK> in a <UNK> <END>

val
an young woman in a shirt playing a video game with his neck <END>
GT:<START> a baseball player throwing a ball at a person <END>

val
a couple of people walking across a side of a water <END>
GT:<START> a man in a wet suit <UNK> at the edge of the water <END>



6. Application to text classification
    1. 94.58%
    2. 92.23%

Unlabeled text sentiment analysis result stored in code/data/result.txt