```
epoch 1, step: 0, tr_loss: 1372.55, vl_loss: 1114.05, tr_err: 6.36, vl_err: 9.04
epoch 2, step: 0, tr_loss: 1243.30, vl_loss: 1054.03, tr_err: 9.04, vl_err: 8.29
epoch 3, step: 0, tr_loss: 1160.20, vl_loss: 972.34, tr_err: 8.29, vl_err: 6.38
epoch 4, step: 0, tr_loss: 1072.39, vl_loss: 882.00, tr_err: 6.38, vl_err: 4.26
epoch 5, step: 0, tr_loss: 980.11, vl_loss: 797.85, tr_err: 4.26, vl_err: 2.75
epoch 6, step: 0, tr_loss: 893.64, vl_loss: 724.57, tr_err: 2.75, vl_err: 1.95
epoch 7, step: 0, tr_loss: 817.14, vl_loss: 663.52, tr_err: 1.95, vl_err: 1.54
epoch 8, step: 0, tr_loss: 752.15, vl_loss: 613.11, tr_err: 1.54, vl_err: 1.18
epoch 9, step: 0, tr_loss: 697.49, vl_loss: 570.86, tr_err: 1.18, vl_err: 0.99
epoch 10, step: 0, tr_loss: 651.03, vl_loss: 534.16, tr_err: 0.99, vl_err: 0.96
time_elapsed for 200 steps: 0.957,
('Sample mini-batch results: \npredicted string: ', array(['N', ' ', 'N', ' ', 'N', 'E', 'N', 'I', "'", 'H',
'U', 'N', ' ',
       'E', 'O', 'Y', 'O', 'L', 'P', 'S', 'O', 'T', 'A', 'G', 'N', ' ',
       'U', 'S', 'M', 'H', 'I', "'", 'I', 'N', 'R', 'B', 'A', ' ', 'N',
       ' ', ' ', ' ', ' ', ' ', ' ', ' ', ' ', ' ', ' ', ' ', ' ', ' ',
       ' ', ' ', ' ', ' ', ' ', ' ', ' ', ' ', ' ', ' ', ' ', ' '],
      dtype='|S1'))
 ('actual string1: ', array(['I', "'", 'M', ' ', 'L', 'O', 'O', 'K', 'I', 'N', 'G', ' ', 'F',
       'O', 'R', ' ', 'A', ' ', 'M', 'E', 'D', 'I', 'T', 'E', 'R', 'R',
       'A', 'N', 'E', 'A', 'N', ' ', 'R', 'E', 'S', 'T', 'A', 'U', 'R',
       'A', 'N', 'T'],
      dtype='|S1'))

epoch 11, step: 0, tr_loss: 610.37, vl_loss: 502.40, tr_err: 0.96, vl_err: 0.92
epoch 12, step: 0, tr_loss: 574.91, vl_loss: 474.43, tr_err: 0.92, vl_err: 0.86
epoch 13, step: 0, tr_loss: 543.52, vl_loss: 449.98, tr_err: 0.86, vl_err: 0.81
epoch 14, step: 0, tr_loss: 515.88, vl_loss: 428.08, tr_err: 0.81, vl_err: 0.75
epoch 15, step: 0, tr_loss: 491.04, vl_loss: 408.16, tr_err: 0.75, vl_err: 0.74
epoch 16, step: 0, tr_loss: 468.43, vl_loss: 390.10, tr_err: 0.74, vl_err: 0.71
epoch 17, step: 0, tr_loss: 447.90, vl_loss: 373.51, tr_err: 0.71, vl_err: 0.70
epoch 18, step: 0, tr_loss: 429.01, vl_loss: 358.15, tr_err: 0.70, vl_err: 0.68
epoch 19, step: 0, tr_loss: 411.54, vl_loss: 344.06, tr_err: 0.68, vl_err: 0.65
epoch 20, step: 0, tr_loss: 395.48, vl_loss: 331.02, tr_err: 0.65, vl_err: 0.66
time_elapsed for 200 steps: 0.950,
('Sample mini-batch results: \npredicted string: ', array(['N', ' ', 'N', ' ', 'N', ' ', 'N', 'E', 'M', ' ',
'E', 'A', 'E',
       'S', ' ', 'P', 'H', 'O', 'N', 'I', ' ', 'N', 'U', 'M', 'E', 'N'],
      dtype='|S1'))
 ('actual string1: ', array(['I', ' ', 'W', 'O', 'U', 'L', 'D', ' ', 'L', 'I', 'K', 'E', ' ',
       'T', 'O', ' ', 'F', 'I', 'N', 'D', ' ', 'A', 'N', ' ', 'I', 'N',
       'D', 'I', 'A', 'N', ' ', 'R', 'E', 'S', 'T', 'A', 'U', 'R', 'A',
       'N', 'T', ' ', 'A', 'N', 'D', ' ', 'I', 'T', ' ', 'S', 'H', 'O',
       'U', 'L', 'D', ' ', 'B', 'E', ' ', 'I', 'N', ' ', 'T', 'H', 'E',
       ' ', 'G', 'I', 'R', 'T', 'O', 'N', ' ', 'A', 'R', 'E', 'A'],
      dtype='|S1'))
```

```
epoch 21, step: 0, tr_loss: 380.62, vl_loss: 319.02, tr_err: 0.66, vl_err: 0.59
epoch 22, step: 0, tr_loss: 366.90, vl_loss: 307.72, tr_err: 0.59, vl_err: 0.58
epoch 23, step: 0, tr_loss: 354.02, vl_loss: 297.23, tr_err: 0.58, vl_err: 0.57
epoch 24, step: 0, tr_loss: 342.04, vl_loss: 287.44, tr_err: 0.57, vl_err: 0.54
epoch 25, step: 0, tr_loss: 330.85, vl_loss: 278.26, tr_err: 0.54, vl_err: 0.48
epoch 26, step: 0, tr_loss: 320.35, vl_loss: 269.61, tr_err: 0.48, vl_err: 0.44
epoch 27, step: 0, tr_loss: 310.46, vl_loss: 261.48, tr_err: 0.44, vl_err: 0.42
epoch 28, step: 0, tr_loss: 301.16, vl_loss: 253.82, tr_err: 0.42, vl_err: 0.46
epoch 29, step: 0, tr_loss: 292.39, vl_loss: 246.60, tr_err: 0.46, vl_err: 0.40
epoch 30, step: 0, tr_loss: 284.13, vl_loss: 239.78, tr_err: 0.40, vl_err: 0.39
time_elapsed for 200 steps: 0.954,
('Sample mini-batch results: \npredicted string: ', array(['I', ' ', 'N', ' ', 'N', ' ', 'N', 'E', 'D', ' ', 'H',
'E', ' ',
       'A', 'D', 'R', 'E', 'S', ' ', 'P', 'H', 'O', 'N', 'E', ' ', 'N',
       'U', 'M', 'B', 'E', 'R', ' ', 'A', 'N'],
      dtype='|S1'))
 ('actual string1: ', array(['I', ' ', 'W', 'O', 'U', 'L', 'D', ' ', 'L', 'I', 'K', 'E', ' ',
       'T', 'O', ' ', 'F', 'I', 'N', 'D', ' ', 'A', 'N', ' ', 'I', 'N',
       'D', 'I', 'A', 'N', ' ', 'R', 'E', 'S', 'T', 'A', 'U', 'R', 'A',
       'N', 'T', ' ', 'A', 'N', 'D', ' ', 'I', 'T', ' ', 'S', 'H', 'O',
       'U', 'L', 'D', ' ', 'B', 'E', ' ', 'I', 'N', ' ', 'T', 'H', 'E',
       ' ', 'G', 'I', 'R', 'T', 'O', 'N', ' ', 'A', 'R', 'E', 'A'],
      dtype='|S1'))

epoch 31, step: 0, tr_loss: 276.32, vl_loss: 233.32, tr_err: 0.39, vl_err: 0.39
epoch 32, step: 0, tr_loss: 268.92, vl_loss: 227.18, tr_err: 0.39, vl_err: 0.36
epoch 33, step: 0, tr_loss: 261.89, vl_loss: 221.38, tr_err: 0.36, vl_err: 0.35
epoch 34, step: 0, tr_loss: 255.24, vl_loss: 215.88, tr_err: 0.35, vl_err: 0.32
epoch 35, step: 0, tr_loss: 248.93, vl_loss: 210.66, tr_err: 0.32, vl_err: 0.33
epoch 36, step: 0, tr_loss: 242.93, vl_loss: 205.68, tr_err: 0.33, vl_err: 0.32
epoch 37, step: 0, tr_loss: 237.22, vl_loss: 200.93, tr_err: 0.32, vl_err: 0.32
epoch 38, step: 0, tr_loss: 231.76, vl_loss: 196.39, tr_err: 0.32, vl_err: 0.32
epoch 39, step: 0, tr_loss: 226.55, vl_loss: 192.05, tr_err: 0.32, vl_err: 0.32
epoch 40, step: 0, tr_loss: 221.57, vl_loss: 187.90, tr_err: 0.32, vl_err: 0.30
time_elapsed for 200 steps: 0.986,
('Sample mini-batch results: \npredicted string: ', array(['I', "'", ' ', 'N', ' ', 'O', 'K', 'I', 'N', 'G', ' ',
'F', 'O',
       'R', ' ', 'A', ' ', 'C', 'H', 'E', 'A', 'P', ' ', 'P', 'U', 'B',
       ' ', 'I', 'N', ' ', 'N', ' ', ' ', ' ', ' ', ' '],
      dtype='|S1'))
 ('actual string1: ', array(['I', "'", 'M', ' ', 'L', 'O', 'O', 'K', 'I', 'N', 'G', ' ', 'F',
       'O', 'R', ' ', 'A', ' ', 'M', 'E', 'D', 'I', 'T', 'E', 'R', 'R',
       'A', 'N', 'E', 'A', 'N', ' ', 'R', 'E', 'S', 'T', 'A', 'U', 'R',
       'A', 'N', 'T'],
      dtype='|S1'))
```

```
epoch 41, step: 0, tr_loss: 216.79, vl_loss: 183.91, tr_err: 0.30, vl_err: 0.30
epoch 42, step: 0, tr_loss: 212.21, vl_loss: 180.09, tr_err: 0.30, vl_err: 0.29
epoch 43, step: 0, tr_loss: 207.82, vl_loss: 176.44, tr_err: 0.29, vl_err: 0.31
epoch 44, step: 0, tr_loss: 203.63, vl_loss: 172.91, tr_err: 0.31, vl_err: 0.28
epoch 45, step: 0, tr_loss: 199.57, vl_loss: 169.53, tr_err: 0.28, vl_err: 0.27
epoch 46, step: 0, tr_loss: 195.68, vl_loss: 166.27, tr_err: 0.27, vl_err: 0.25
epoch 47, step: 0, tr_loss: 191.93, vl_loss: 163.12, tr_err: 0.25, vl_err: 0.23
epoch 48, step: 0, tr_loss: 188.32, vl_loss: 160.11, tr_err: 0.23, vl_err: 0.25
epoch 49, step: 0, tr_loss: 184.85, vl_loss: 157.21, tr_err: 0.25, vl_err: 0.25
epoch 50, step: 0, tr_loss: 181.51, vl_loss: 154.41, tr_err: 0.25, vl_err: 0.25
time_elapsed for 200 steps: 0.992,
('Sample mini-batch results: \npredicted string: ', array(['I', ' ', 'N', ' ', 'N', 'E', 'D', ' ', 'T', 'H',
'E', ' ', 'A',
       'D', 'R', 'E', 'S', ' ', 'P', 'H', 'O', 'N', 'E', ' ', 'N', 'U',
       'M', 'B', 'E', 'R', ' ', 'A', 'N', 'R', 'N'],
      dtype='|S1'))
 ('actual string1: ', array(['I', ' ', 'W', 'O', 'U', 'L', 'D', ' ', 'L', 'I', 'K', 'E', ' ',
       'T', 'O', ' ', 'F', 'I', 'N', 'D', ' ', 'A', 'N', ' ', 'I', 'N',
       'D', 'I', 'A', 'N', ' ', 'R', 'E', 'S', 'T', 'A', 'U', 'R', 'A',
       'N', 'T', ' ', 'A', 'N', 'D', ' ', 'I', 'T', ' ', 'S', 'H', 'O',
       'U', 'L', 'D', ' ', 'B', 'E', ' ', 'I', 'N', ' ', 'T', 'H', 'E',
       ' ', 'G', 'I', 'R', 'T', 'O', 'N', ' ', 'A', 'R', 'E', 'A'],
      dtype='|S1'))

epoch 51, step: 0, tr_loss: 178.30, vl_loss: 151.71, tr_err: 0.25, vl_err: 0.26
epoch 52, step: 0, tr_loss: 175.18, vl_loss: 149.11, tr_err: 0.26, vl_err: 0.25
epoch 53, step: 0, tr_loss: 172.19, vl_loss: 146.59, tr_err: 0.25, vl_err: 0.25
epoch 54, step: 0, tr_loss: 169.30, vl_loss: 144.16, tr_err: 0.25, vl_err: 0.24
epoch 55, step: 0, tr_loss: 166.49, vl_loss: 141.80, tr_err: 0.24, vl_err: 0.25
epoch 56, step: 0, tr_loss: 163.78, vl_loss: 139.52, tr_err: 0.25, vl_err: 0.25
epoch 57, step: 0, tr_loss: 161.15, vl_loss: 137.30, tr_err: 0.25, vl_err: 0.22
epoch 58, step: 0, tr_loss: 158.60, vl_loss: 135.16, tr_err: 0.22, vl_err: 0.25
epoch 59, step: 0, tr_loss: 156.13, vl_loss: 133.08, tr_err: 0.25, vl_err: 0.25
epoch 60, step: 0, tr_loss: 153.74, vl_loss: 131.07, tr_err: 0.25, vl_err: 0.24
time_elapsed for 200 steps: 0.973,
('Sample mini-batch results: \npredicted string: ', array(['T', 'H', 'A', 'N', 'K', ' ', 'Y', 'O', 'U', ' ',
'G', 'O', 'D',
       'B', 'Y', 'E', ' ', ' ', ' ', ' ', ' ', ' ', ' ', ' ', ' ', ' ',
       ' ', ' ', ' ', ' ', ' ', ' ', ' ', ' ', ' ', ' '],
      dtype='|S1'))
 ('actual string1: ', array(['I', ' ', 'A', 'M', ' ', 'L', 'O', 'O', 'K', 'I', 'N', 'G', ' ',
       'F', 'O', 'R', ' ', 'A', ' ', 'P', 'U', 'B', ' ', 'I', 'N', ' ',
       'T', 'H', 'E', ' ', 'R', 'I', 'V', 'E', 'R', 'S', 'I', 'D', 'E',
       ' ', 'A', 'R', 'E', 'A'],
      dtype='|S1'))

epoch 61, step: 0, tr_loss: 151.42, vl_loss: 129.12, tr_err: 0.24, vl_err: 0.26
```

```
epoch 62, step: 0, tr_loss: 149.17, vl_loss: 127.23, tr_err: 0.26, vl_err: 0.26
epoch 63, step: 0, tr_loss: 146.99, vl_loss: 125.39, tr_err: 0.26, vl_err: 0.22
epoch 64, step: 0, tr_loss: 144.88, vl_loss: 123.61, tr_err: 0.22, vl_err: 0.23
epoch 65, step: 0, tr_loss: 142.82, vl_loss: 121.87, tr_err: 0.23, vl_err: 0.24
epoch 66, step: 0, tr_loss: 140.82, vl_loss: 120.19, tr_err: 0.24, vl_err: 0.28
epoch 67, step: 0, tr_loss: 138.89, vl_loss: 118.56, tr_err: 0.28, vl_err: 0.27
epoch 68, step: 0, tr_loss: 137.01, vl_loss: 116.98, tr_err: 0.27, vl_err: 0.27
epoch 69, step: 0, tr_loss: 135.18, vl_loss: 115.44, tr_err: 0.27, vl_err: 0.25
epoch 70, step: 0, tr_loss: 133.40, vl_loss: 113.94, tr_err: 0.25, vl_err: 0.24
time_elapsed for 200 steps: 0.970,
('Sample mini-batch results: \npredicted string: ', array(['I', ' ', 'N', 'E', 'D', ' ', 'T', 'H', 'E', ' ', 'A',
'D', 'R',
       'E', 'S', ' ', 'P', 'H', 'O', 'N', 'E', ' ', 'N', 'U', 'M', 'B',
       'E', 'R', ' ', 'A', 'N', 'E', 'N', 'P', 'N', 'I', 'C', 'N'],
      dtype='|S1'))
 ('actual string1: ', array(['I', ' ', 'W', 'O', 'U', 'L', 'D', ' ', 'L', 'I', 'K', 'E', ' ',
       'T', 'O', ' ', 'F', 'I', 'N', 'D', ' ', 'A', 'N', ' ', 'I', 'N',
       'D', 'I', 'A', 'N', ' ', 'R', 'E', 'S', 'T', 'A', 'U', 'R', 'A',
       'N', 'T', ' ', 'A', 'N', 'D', ' ', 'I', 'T', ' ', 'S', 'H', 'O',
       'U', 'L', 'D', ' ', 'B', 'E', ' ', 'I', 'N', ' ', 'T', 'H', 'E',
       ' ', 'G', 'I', 'R', 'T', 'O', 'N', ' ', 'A', 'R', 'E', 'A'],
      dtype='|S1'))

epoch 71, step: 0, tr_loss: 131.67, vl_loss: 112.49, tr_err: 0.24, vl_err: 0.25
epoch 72, step: 0, tr_loss: 129.99, vl_loss: 111.07, tr_err: 0.25, vl_err: 0.27
epoch 73, step: 0, tr_loss: 128.35, vl_loss: 109.70, tr_err: 0.27, vl_err: 0.29
epoch 74, step: 0, tr_loss: 126.76, vl_loss: 108.35, tr_err: 0.29, vl_err: 0.27
epoch 75, step: 0, tr_loss: 125.21, vl_loss: 107.05, tr_err: 0.27, vl_err: 0.23
epoch 76, step: 0, tr_loss: 123.70, vl_loss: 105.77, tr_err: 0.23, vl_err: 0.24
epoch 77, step: 0, tr_loss: 122.22, vl_loss: 104.53, tr_err: 0.24, vl_err: 0.27
epoch 78, step: 0, tr_loss: 120.79, vl_loss: 103.32, tr_err: 0.27, vl_err: 0.25
epoch 79, step: 0, tr_loss: 119.39, vl_loss: 102.14, tr_err: 0.25, vl_err: 0.24
epoch 80, step: 0, tr_loss: 118.02, vl_loss: 100.99, tr_err: 0.24, vl_err: 0.27
time_elapsed for 200 steps: 1.055,
('Sample mini-batch results: \npredicted string: ', array(['I', '"', 'M', ' ', 'L', 'O', 'K', 'I', 'N', 'G', '
', 'F', 'O',
       'R', ' ', 'A', ' ', 'C', 'H', 'E', 'A', 'P', ' ', 'P', 'U', 'B',
       ' ', 'I', 'N', ' ', 'N', 'E', 'N', 'S', 'N', ' ', ' '],
      dtype='|S1'))
 ('actual string1: ', array(['I', '"', 'M', ' ', 'L', 'O', 'O', 'K', 'I', 'N', 'G', ' ', 'F',
       'O', 'R', ' ', 'A', ' ', 'M', 'E', 'D', 'I', 'T', 'E', 'R', 'R',
       'A', 'N', 'E', 'A', 'N', ' ', 'R', 'E', 'S', 'T', 'A', 'U', 'R',
       'A', 'N', 'T'],
      dtype='|S1'))

epoch 81, step: 0, tr_loss: 116.69, vl_loss: 99.87, tr_err: 0.27, vl_err: 0.28
epoch 82, step: 0, tr_loss: 115.39, vl_loss: 98.77, tr_err: 0.28, vl_err: 0.24
```

```
epoch 83, step: 0, tr_loss: 114.12, vl_loss: 97.70, tr_err: 0.24, vl_err: 0.27
epoch 84, step: 0, tr_loss: 112.88, vl_loss: 96.66, tr_err: 0.27, vl_err: 0.29
epoch 85, step: 0, tr_loss: 111.67, vl_loss: 95.64, tr_err: 0.29, vl_err: 0.30
epoch 86, step: 0, tr_loss: 110.48, vl_loss: 94.64, tr_err: 0.30, vl_err: 0.25
epoch 87, step: 0, tr_loss: 109.33, vl_loss: 93.67, tr_err: 0.25, vl_err: 0.24
epoch 88, step: 0, tr_loss: 108.20, vl_loss: 92.72, tr_err: 0.24, vl_err: 0.27
epoch 89, step: 0, tr_loss: 107.10, vl_loss: 91.79, tr_err: 0.27, vl_err: 0.27
epoch 90, step: 0, tr_loss: 106.02, vl_loss: 90.88, tr_err: 0.27, vl_err: 0.28
time_elapsed for 200 steps: 1.022,
('Sample mini-batch results: \npredicted string: ', array(['T', 'H', 'A', 'N', 'K', ' ', 'Y', 'O', 'U', ' ',
'G', 'O', 'D',
       'B', 'Y', 'E', ' ', ' ', ' ', ' ', ' ', ' ', ' ', ' ', ' ', ' ',
       ' ', ' ', ' ', ' ', ' ', ' ', ' ', ' ', ' ', ' ', ' ', ' ', ' ',
       ' ', ' ', ' '],
      dtype='|S1'))
 ('actual string1: ', array(['I', ' ', 'A', 'M', ' ', 'L', 'O', 'O', 'K', 'I', 'N', 'G', ' ',
       'F', 'O', 'R', ' ', 'A', ' ', 'P', 'U', 'B', ' ', 'I', 'N', ' ',
       'T', 'H', 'E', ' ', 'R', 'I', 'V', 'E', 'R', 'S', 'I', 'D', 'E',
       ' ', 'A', 'R', 'E', 'A'],
      dtype='|S1'))

epoch 91, step: 0, tr_loss: 104.96, vl_loss: 89.98, tr_err: 0.28, vl_err: 0.25
epoch 92, step: 0, tr_loss: 103.93, vl_loss: 89.11, tr_err: 0.25, vl_err: 0.24
epoch 93, step: 0, tr_loss: 102.91, vl_loss: 88.26, tr_err: 0.24, vl_err: 0.22
epoch 94, step: 0, tr_loss: 101.92, vl_loss: 87.42, tr_err: 0.22, vl_err: 0.24
epoch 95, step: 0, tr_loss: 100.95, vl_loss: 86.61, tr_err: 0.24, vl_err: 0.24
epoch 96, step: 0, tr_loss: 100.00, vl_loss: 85.80, tr_err: 0.24, vl_err: 0.24
epoch 97, step: 0, tr_loss: 99.07, vl_loss: 85.02, tr_err: 0.24, vl_err: 0.24
epoch 98, step: 0, tr_loss: 98.16, vl_loss: 84.25, tr_err: 0.24, vl_err: 0.24
epoch 99, step: 0, tr_loss: 97.26, vl_loss: 83.50, tr_err: 0.24, vl_err: 0.28
epoch 100, step: 0, tr_loss: 96.39, vl_loss: 82.76, tr_err: 0.28, vl_err: 0.30
time_elapsed for 200 steps: 0.981,
('Sample mini-batch results: \npredicted string: ', array(['I', ' ', 'N', 'E', 'D', ' ', 'T', 'H', 'E', ' ', 'A',
'D', 'R',
       'E', 'S', ' ', 'P', 'H', 'O', 'N', 'E', ' ', 'N', 'U', 'M', 'B',
       'E', 'R', ' ', 'A', 'N', 'E', 'N', 'D', 'N', 'T', 'N', 'E', 'N',
       'R', 'I', 'N'],
      dtype='|S1'))
 ('actual string1: ', array(['I', ' ', 'W', 'O', 'U', 'L', 'D', ' ', 'L', 'I', 'K', 'E', ' ',
       'T', 'O', ' ', 'F', 'I', 'N', 'D', ' ', 'A', 'N', ' ', 'I', 'N',
       'D', 'I', 'A', 'N', ' ', 'R', 'E', 'S', 'T', 'A', 'U', 'R', 'A',
       'N', 'T', ' ', 'A', 'N', 'D', ' ', 'I', 'T', ' ', 'S', 'H', 'O',
       'U', 'L', 'D', ' ', 'B', 'E', ' ', 'I', 'N', ' ', 'T', 'H', 'E',
       ' ', 'G', 'I', 'R', 'T', 'O', 'N', ' ', 'A', 'R', 'E', 'A'],
      dtype='|S1'))

epoch 101, step: 0, tr_loss: 95.53, vl_loss: 82.03, tr_err: 0.30, vl_err: 0.25
```

```
epoch 102, step: 0, tr_loss: 94.68, vl_loss: 81.32, tr_err: 0.25, vl_err: 0.24
epoch 103, step: 0, tr_loss: 93.86, vl_loss: 80.62, tr_err: 0.24, vl_err: 0.25
epoch 104, step: 0, tr_loss: 93.05, vl_loss: 79.94, tr_err: 0.25, vl_err: 0.25
epoch 105, step: 0, tr_loss: 92.25, vl_loss: 79.27, tr_err: 0.25, vl_err: 0.25
epoch 106, step: 0, tr_loss: 91.47, vl_loss: 78.61, tr_err: 0.25, vl_err: 0.24
epoch 107, step: 0, tr_loss: 90.70, vl_loss: 77.96, tr_err: 0.24, vl_err: 0.24
epoch 108, step: 0, tr_loss: 89.95, vl_loss: 77.33, tr_err: 0.24, vl_err: 0.26
epoch 109, step: 0, tr_loss: 89.21, vl_loss: 76.71, tr_err: 0.26, vl_err: 0.26
epoch 110, step: 0, tr_loss: 88.49, vl_loss: 76.10, tr_err: 0.26, vl_err: 0.27
time_elapsed for 200 steps: 0.961,
('Sample mini-batch results: \npredicted string: ', array(['I', """, 'M', ' ', 'L', 'O', 'K', 'I', 'N', 'G', '
', 'F', 'O',
       'R', ' ', 'A', ' ', 'C', 'H', 'E', 'A', 'P', ' ', 'P', 'U', 'B',
       ' ', 'I', 'N', ' ', 'N', 'E', 'N', 'S', 'N', ' ', ' ', ' ', ' ',
       ' ', ' ', ' '],
      dtype='|S1'))
 ('actual string1: ', array(['I', """, 'M', ' ', 'L', 'O', 'O', 'K', 'I', 'N', 'G', ' ', 'F',
       'O', 'R', ' ', 'A', ' ', 'M', 'E', 'D', 'I', 'T', 'E', 'R', 'R',
       'A', 'N', 'E', 'A', 'N', ' ', 'R', 'E', 'S', 'T', 'A', 'U', 'R',
       'A', 'N', 'T'],
      dtype='|S1'))

epoch 111, step: 0, tr_loss: 87.78, vl_loss: 75.50, tr_err: 0.27, vl_err: 0.27
epoch 112, step: 0, tr_loss: 87.08, vl_loss: 74.91, tr_err: 0.27, vl_err: 0.24
epoch 113, step: 0, tr_loss: 86.39, vl_loss: 74.33, tr_err: 0.24, vl_err: 0.31
epoch 114, step: 0, tr_loss: 85.72, vl_loss: 73.76, tr_err: 0.31, vl_err: 0.28
epoch 115, step: 0, tr_loss: 85.06, vl_loss: 73.20, tr_err: 0.28, vl_err: 0.22
epoch 116, step: 0, tr_loss: 84.40, vl_loss: 72.65, tr_err: 0.22, vl_err: 0.26
epoch 117, step: 0, tr_loss: 83.76, vl_loss: 72.11, tr_err: 0.26, vl_err: 0.25
epoch 118, step: 0, tr_loss: 83.13, vl_loss: 71.58, tr_err: 0.25, vl_err: 0.25
epoch 119, step: 0, tr_loss: 82.51, vl_loss: 71.06, tr_err: 0.25, vl_err: 0.25
epoch 120, step: 0, tr_loss: 81.91, vl_loss: 70.55, tr_err: 0.25, vl_err: 0.27
time_elapsed for 200 steps: 0.981,
('Sample mini-batch results: \npredicted string: ', array(['T', 'H', 'A', 'N', 'K', ' ', 'Y', 'O', 'U', ' ',
'G', 'O', 'D',
       'B', 'Y', 'E', ' ', ' ', ' ', ' ', ' ', ' ', ' ', ' ', ' ', ' ',
       ' ', ' ', ' ', ' ', ' ', ' ', ' ', ' ', ' ', ' ', ' ', ' '],
      dtype='|S1'))
 ('actual string1: ', array(['I', ' ', 'A', 'M', ' ', 'L', 'O', 'O', 'K', 'I', 'N', 'G', ' ',
       'F', 'O', 'R', ' ', 'A', ' ', 'P', 'U', 'B', ' ', 'I', 'N', ' ',
       'T', 'H', 'E', ' ', 'R', 'I', 'V', 'E', 'R', 'S', 'I', 'D', 'E',
       ' ', 'A', 'R', 'E', 'A'],
      dtype='|S1'))

epoch 121, step: 0, tr_loss: 81.31, vl_loss: 70.04, tr_err: 0.27, vl_err: 0.26
epoch 122, step: 0, tr_loss: 80.72, vl_loss: 69.55, tr_err: 0.26, vl_err: 0.26
epoch 123, step: 0, tr_loss: 80.14, vl_loss: 69.06, tr_err: 0.26, vl_err: 0.26
```