COMPGI13: Advanced Topics in Machine Learning Assignment #2

Due on Wednesday, March 8, 2017

Thore Graepel, Koray Kavukcuoglu and Oriol Vinyals

Nitish Mutha: UCABMUT

Student No: 15113106

Github username: NitishMutha (https://github.com/NitishMutha)

March 8, 2017

Contents

LSTM and GRU

(a)

Yes, with LSTM (and respectively GRU) we can just store the current input for the next time step, using the forget gate f_t and input gate i_t (and respectively update gate). If the forget gate (respectively update gate) take in inputs from input x_t and hidden state h_{t-1} and if it outputs 0 through its sigmoidal activation function then the memory unit will just ignore the signal from the previous time step and just take the input state from the inpute gate and pass it as the output and the state for the next time step. [1]
i.e. For LSTM:

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \tag{1}$$

if $f_t = 0$ and $i_t = 1$ in eq 1 we get

$$C_t = \tilde{C}_t \tag{2}$$

Where $\tilde{C}_t$ is the output of the tanh layer from the input signal.
In case of GRU, the forget and input gate activations are combined into single update gate z_t which performs similar action as of LSTM. Consider the eq 3 of the GRU

$$h_t = (1 - z_t) \cdot h_{t-1} + z_t \cdot \tilde{h_t} \tag{3}$$

So if we put $z_t = 1$ in the eq 3 we get

$$h_t = \tilde{h_t} \tag{4}$$

(b)

Yes, in LSTM and GRU it is possible to store the previous state only and completely ignore the current state. For LSTM when the forget gate outputs 1 and input gate outputs 0 then all the state from the previous state is preserved and the input signal is ignored.
So if we put $f_t = 1$ and $i_t = 0$ in eq 1, we get

$$C_t = C_{t-1} \tag{5}$$

For GRU, if we put $z_t = 0$ in the eq 3 we get

$$h_t = h_{t-1} \tag{6}$$

(c)

GRU are the special case of LSTM, where GRU have only 2 - update z_t and reset r_t gate while LSTM has 3 - input i_t, forget f_t and output o_t gates. Where the GRU's update gate is a combination of LSTM's forget and input gate.
Consider LSTM gates equations as follows:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{7}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{8}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{9}$$

LSTM and GRU [(c)] continued on next page...

Now consider GRU gates as:

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) \tag{10}$$

Which also can be shown in terms of LSTM gates as follows

$$z_t = f_t + i_t \tag{11}$$

$$z_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) + \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{12}$$

We means the W_z is equivalent to combination of W_f and W_i

Also,

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t]) \tag{13}$$

which can also be written in terms of LSTM gates as:

$$r_t = o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{14}$$

Task 1:

(a)

Classification

In this task, the many-to-one Recurrent Neural Network models were used to classify the MNIST digits.
Types of models implemented for the classification task:

- LSTM (Long Short Term Memory)
  - 32 units
  - 64 units
  - 128 units
- GRU (Gated Recurrent Units)
  - 32 units
  - 64 units
  - 128 units

Model architecture:

The model takes inputs pixel by pixel of the flattened binarized image in a sequence. After the last input it generates the probability distribution over the 10 categories of digits. Implemented the LSTM and GRU models using TensorFlow library using LSTMCell, GRUCell and a dynamic_rnn wrapper. Added MultiRNNCell wrappers for the 3 layer models.

Model parameters:

- Time steps: 784
- Batch size: 256
- Epochs: 50
- Learning rate: 0.001
- Inputs dimensions: [Batch x 784 x 1]
- Affine layer on RNN output with hidden units: 100
- Output dimensions: [Batch x 10]

Training and Optimisation:

Initially, I build a vanilla version of the LSTM without using any normalisation or regularisation techniques (code: task1_vanilla.py). This model had a very low training and test accuracy ranging from 10% to 40% in 50 epochs with Test Accuracy: 0.120593 and Test Losses: 2.29288. As well as the accuracy was not stable and was changing each iteration. Added the graphs of LSTM vanilla version at the end of this section.

Then in order to improve the performance of the model, I train and optimise the models by applying few of the regularisation techniques as follows:

- Dropout: Added a dropout wrapper to the dynamic RNN with input_keep_prob = 1.0, output_keep_prob = 0.9
- Gradient Clipping: Added gradient clipping with threshold of ±5.0 in order prevent exploding gradients problem.
- Batch Normalisation: Added batch normalisation on the hidden layer of the output layer before relu activation in order to improve the convergence. (Note: Did not apply batch normalisation on the RNN cell) [2] [3]

Trained the model using Adam optimizer. Evaluated the model for various learning rates from the range of 0.01, 0.001, 0.003, 0.005, 0.0001 for 30 epochs and found that 0.001 was the optimal learning rate for the configuration of the model. And for the final version of the model used the same with 50 epochs of mini-batch training.

Results:

Model: LSTM	**(1 layer, 32 units)**	**(1 layer, 64 units)**	**(1 layer, 128 units)**	**(3 layer, 32 units)**
Testing Loss	0.938066	0.587523	0.355695	0.371238
Training Loss	1.0274	0.685184	0.335025	0.397177
Testing Accuracy	0.657552	0.789864	0.895533	0.885917
Training Accuracy	0.667969	0.746094	0.898438	0.871094

Table 1: Cross entropy and classification accuracy for LSTM model

Model: GRU	**(1 layer, 32 units)**	**(1 layer, 64 units)**	**(1 layer, 128 units)**	**(3 layer, 32 units)**
Testing Loss	0.373333	0.175594	0.086347	0.0762234
Training Loss	0.343612	0.148358	0.0406595	0.0611289
Testing Accuracy	0.882011	0.947015	0.975461	0.977965
Training Accuracy	0.890622	0.960938	0.976562	0.980469

Table 2: Cross entropy and classification accuracy for GRU model

How does this compare with the results you obtained in the first assignment, when training a model that "sees" the entire image at once?
The Multi Layer Perceptron and Convolutional Neural Networks performed much better and faster where they operated on whole image at once as compared to the Recurrent Neural Network which operated on one pixel at a time. MPL and CNN converges much faster and has better accuracy with a simpler network as compared to RNN. Where as we need to add regularisation and dropouts to the LSTM / GRU to make them stable and converge in considerable amount of time.
For the objective of digit classification convolutional neural network works better, while if the objective is to have some contextual information in the image or if any task in which past and future pixel are important to perform any inference then we might use RNN - LSTM which has a capacity to hold into memory the past states and perform contextual inference or prediction.

Graphs

- LSTM - 1 layer, 32 units

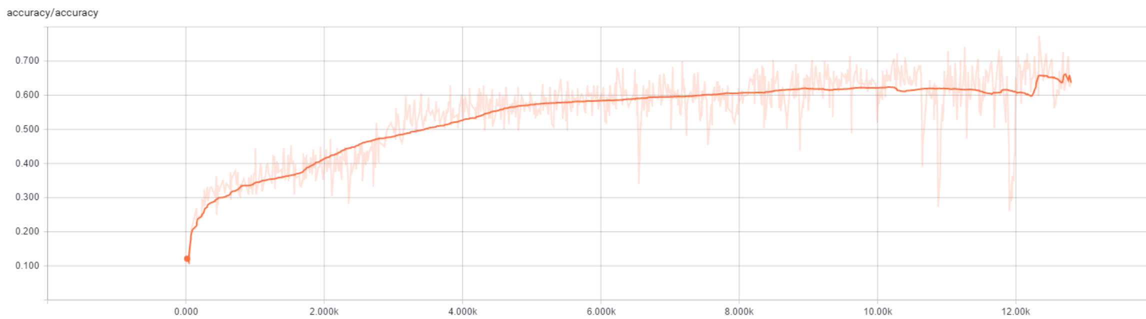


Figure 1: Classification accuracy on training set.

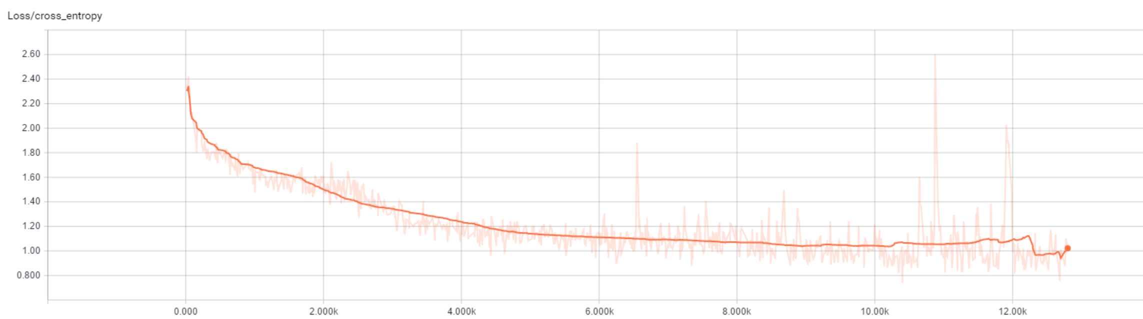


Figure 2: Cross entropy loss on training set.

- LSTM - 1 layer, 64 units

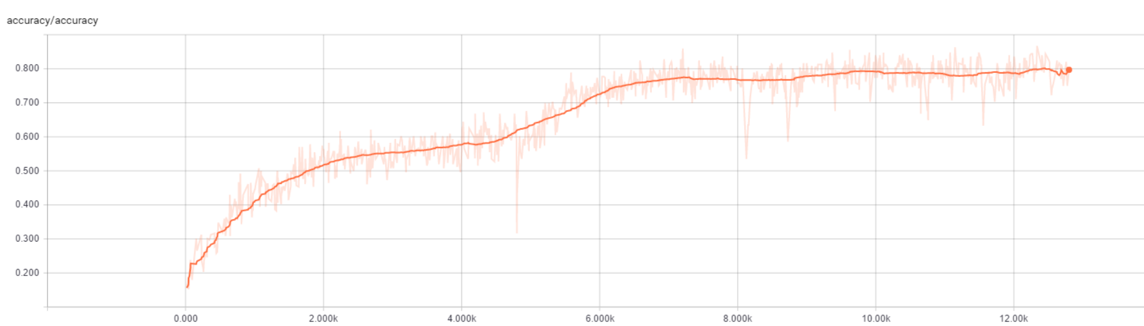


Figure 3: Classification accuracy on training set.

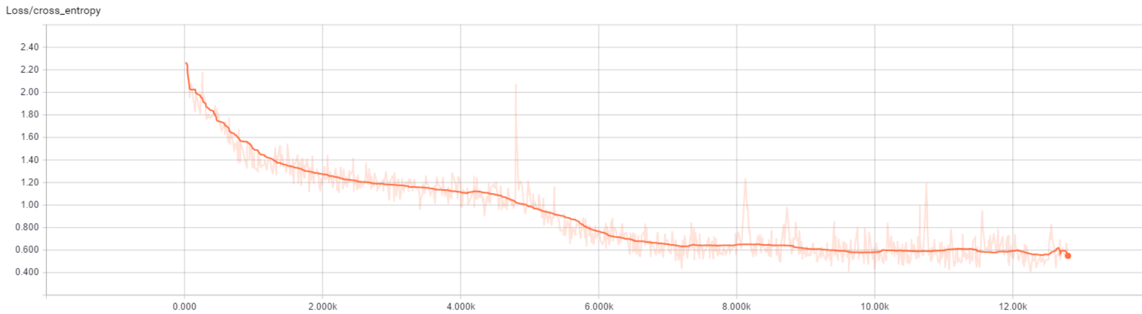


Figure 4: Cross entropy loss on training set.

- LSTM - 1 layer, 128 units

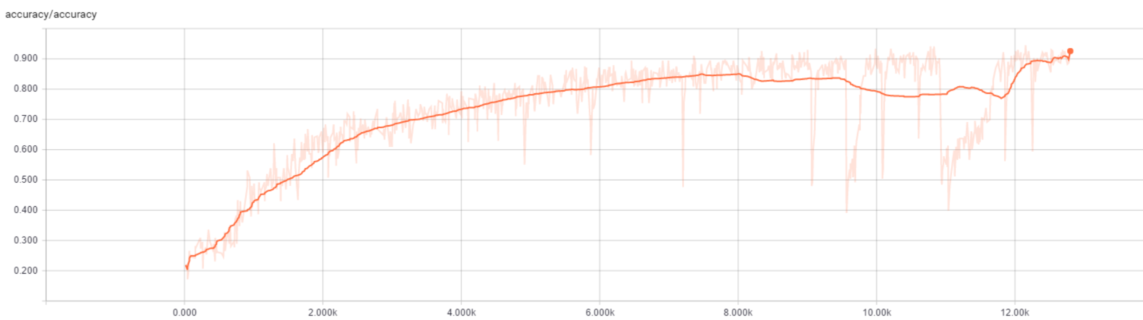


Figure 5: Classification accuracy on training set.

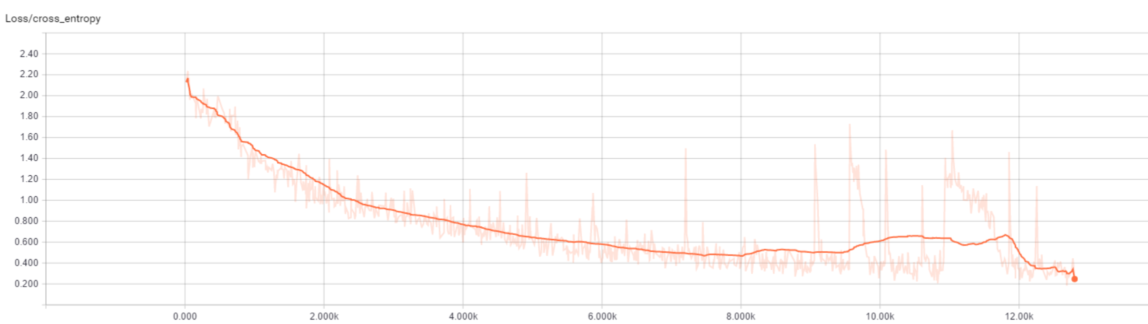


Figure 6: Cross entropy loss on training set.

- LSTM - 3 layer, 32 units

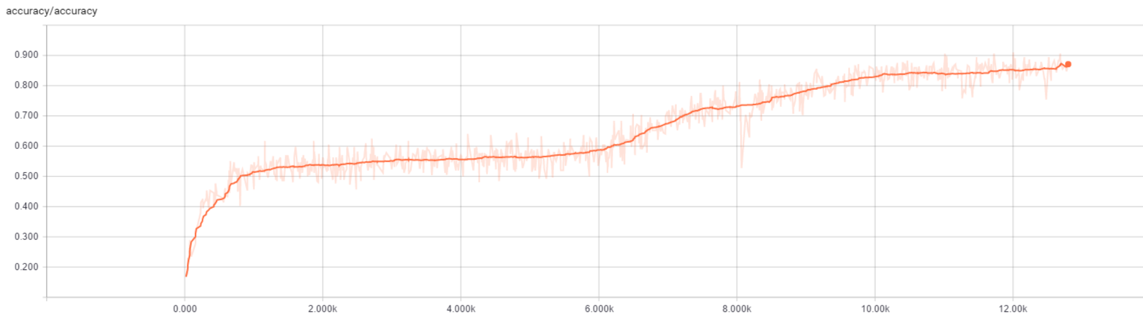


Figure 7: Classification accuracy on training set.

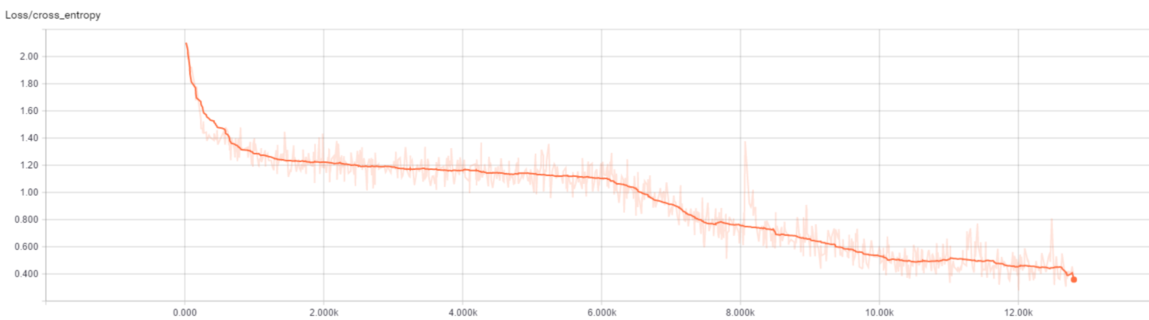


Figure 8: Cross entropy loss on training set.

- GRU - 1 layer, 32 units

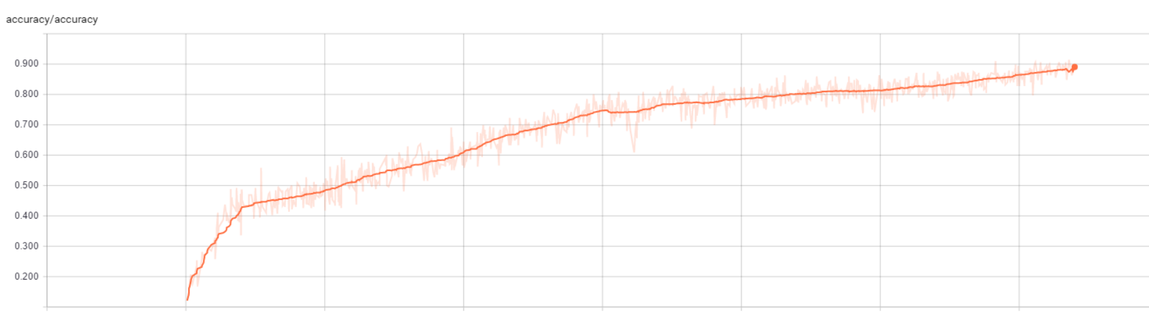


Figure 9: Classification accuracy on training set.

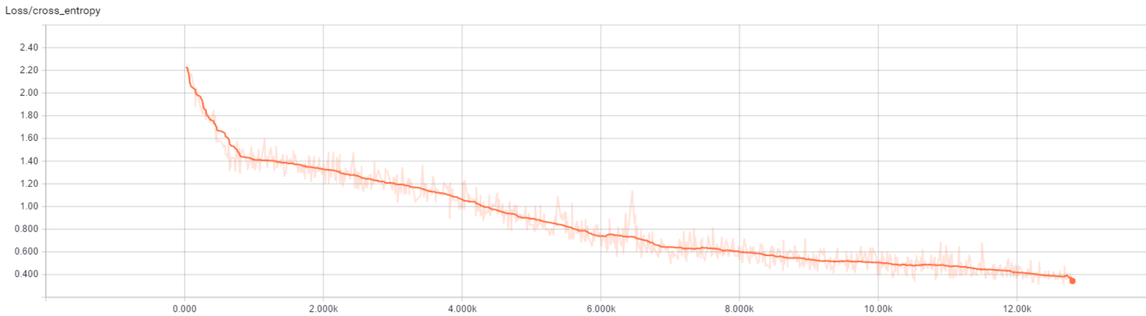


Figure 10: Cross entropy loss on training set.

- GRU - 1 layer, 64 units

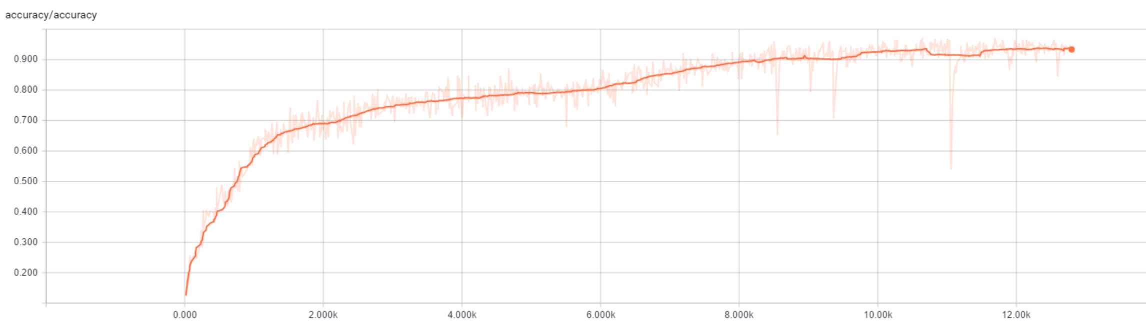


Figure 11: Classification accuracy on training set.

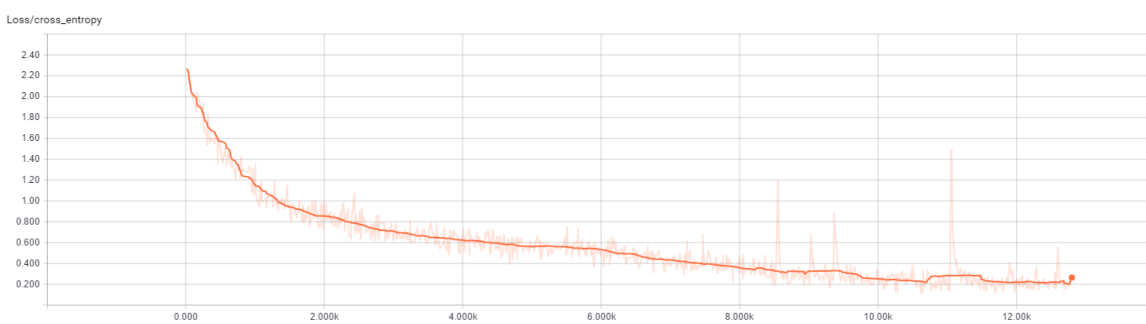


Figure 12: Cross entropy loss on training set.

- GRU - 1 layer, 128 units

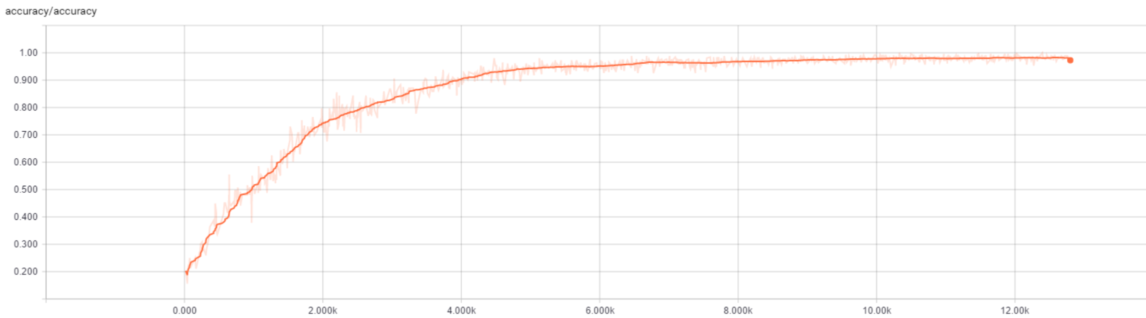


Figure 13: Classification accuracy on training set.

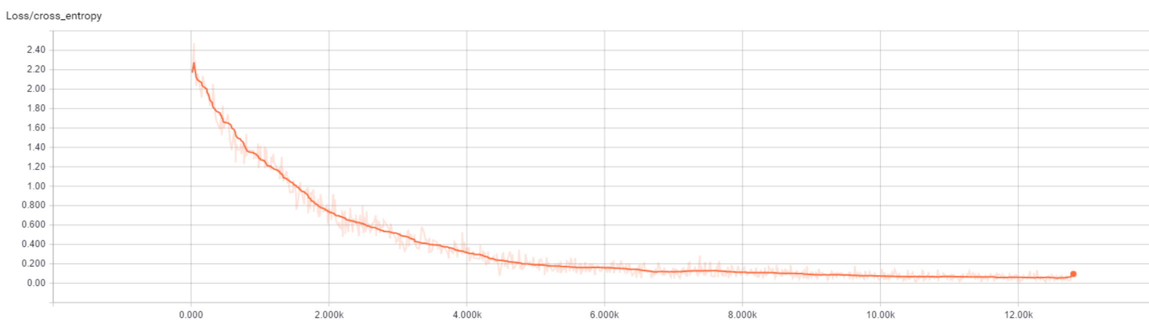


Figure 14: Cross entropy loss on training set.

- GRU - 3 layer, 32 units

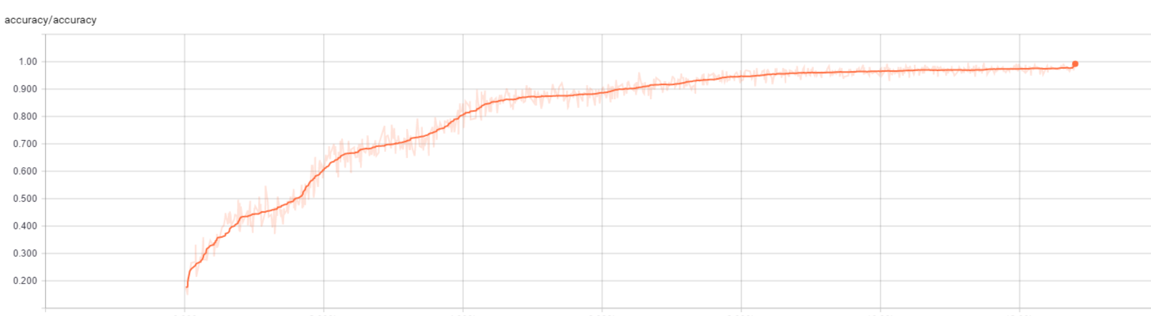


Figure 15: Classification accuracy on training set.

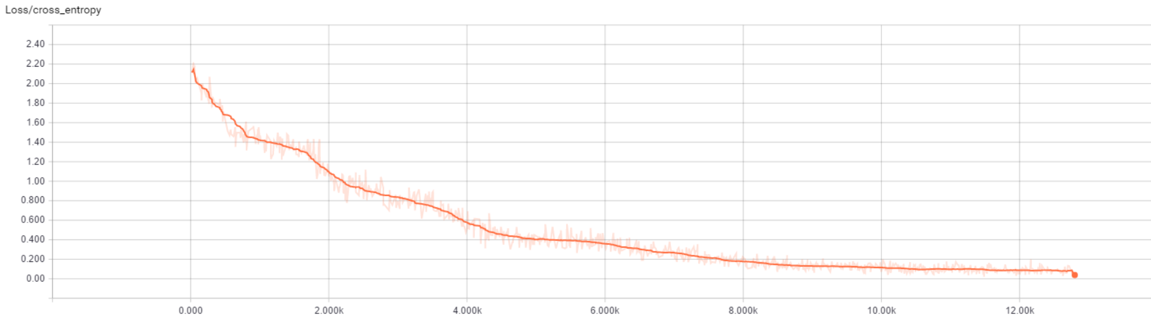


Figure 16: Cross entropy loss on training set.

- LSTM Vanilla, 1 layer, 32 units

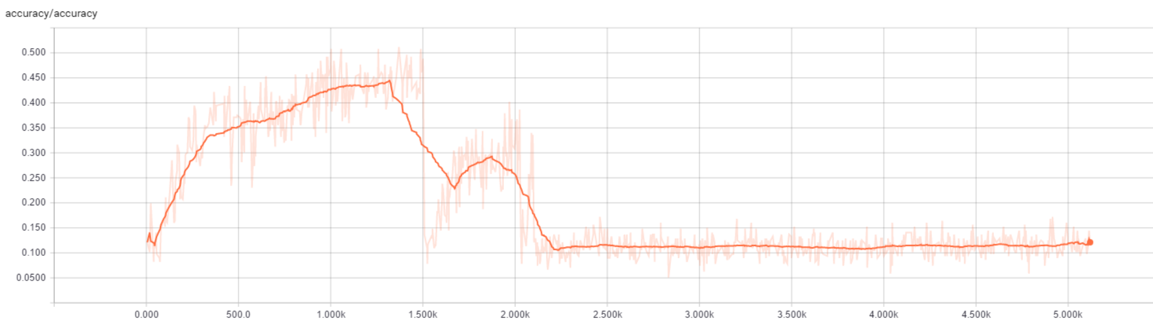


Figure 17: Classification accuracy on training set.

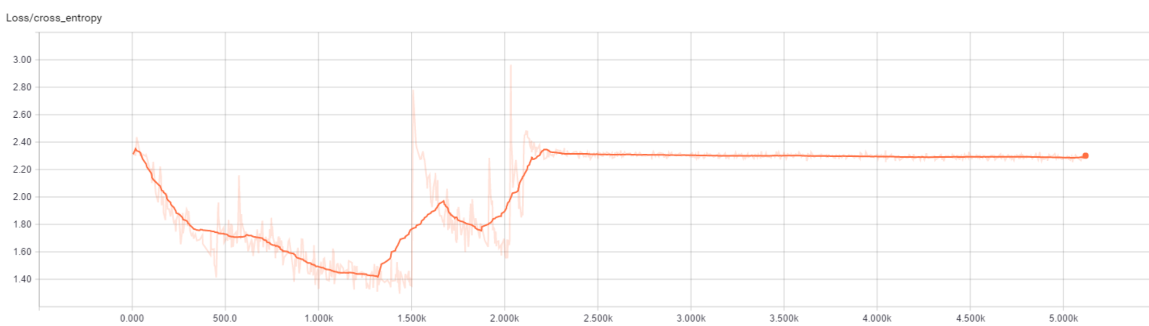


Figure 18: Cross entropy loss on training set.

Task 2:

(a)

Pixel Prediction

As the performance of the GRU proved to be better than the LSTM in the task one, I selected GRU to perform pixel prediction task.
The model used in the pixel prediction is a many-to-many recurrent model. Each time step the input pixel value and recurrent state will generate an input to predict the probability of the next pixel in the sequence.

Implemented the 32, 64, 128 units of GRU models along with stacked 3 layer 32 units GRU model and the resulting cross entropy are shown in the table 3

Model parameters:

- Time steps: 783
- Batch size: 256
- Epochs: 40
- Learning rate: 0.001
- Inputs dimensions: [Batch x 783 x 1]
- Affine layer with output dim on each unit: 1
- Output dimensions: [Batch x 783]

Regularisation:

- Gradient Clipping
- Dropout
- Batch normalization on output affine layer

(a) Results

Model: LSTM	**(1 layer, 32 units)**	**(1 layer, 64 units)**	**(1 layer, 128 units)**	**(3 layer, 32 units)**
Testing Loss	0.10241	0.0969224	0.0904972	0.0976084
Training Loss	0.0969323	0.0888697	0.084762	0.0890505

Table 3: Cross entropy - GRU pixel prediction

- GRU - 1 layer, 32 units

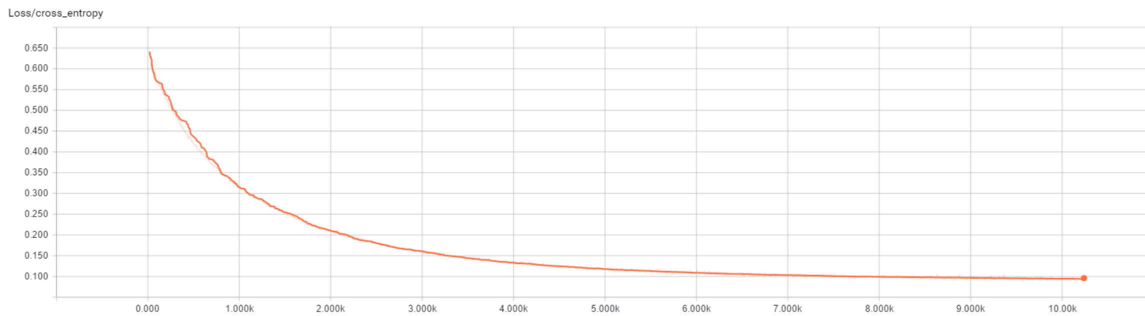


Figure 19: Cross Entropy for pixel prediction on 32 unit GRU

- GRU - 1 layer, 64 units

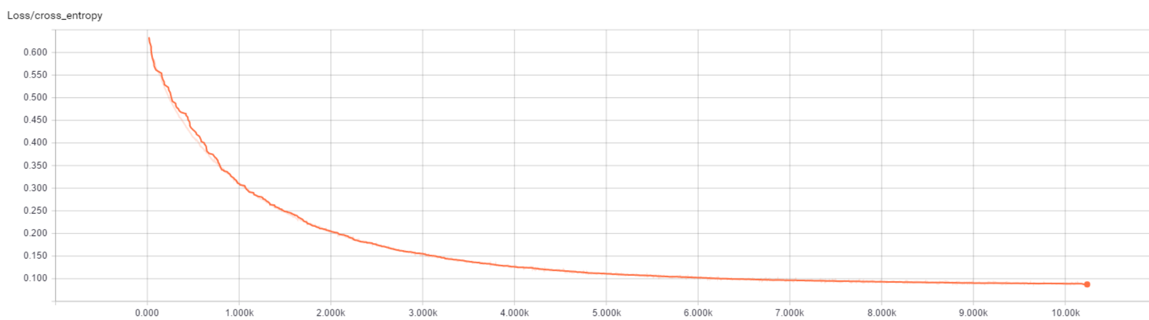


Figure 20: Cross Entropy for pixel prediction on 64 unit GRU

- GRU - 1 layer, 128 units

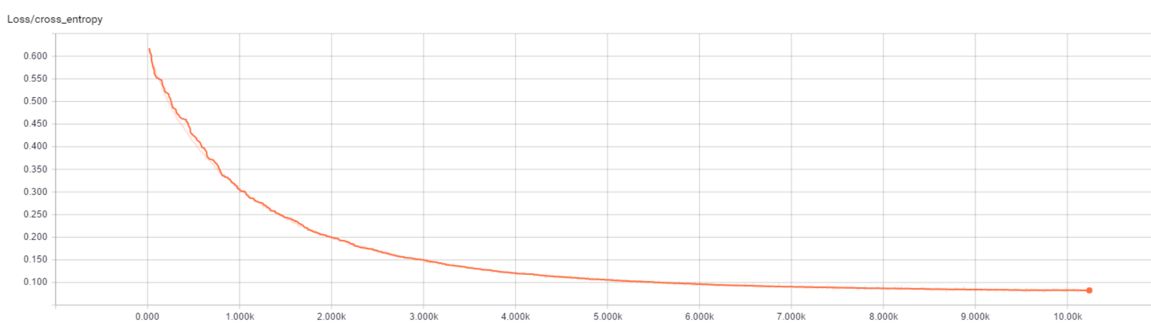


Figure 21: Cross Entropy for pixel prediction on 128 unit GRU

Task 2: [(a)] continued on next page...

- GRU - 3 layer, 32 units

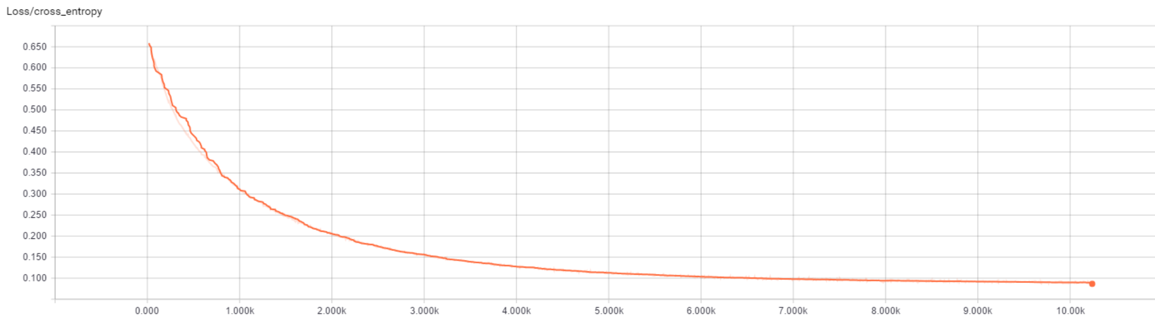


Figure 22: Cross Entropy for pixel prediction on 32 unit, 3 layer GRU

(b) Predict missing parts and compare with GT

Sampled 100 random images from the test set and removed the last 300 pixels from the image. And used the previously trained models to predict the missing 1 step, 10 steps, 28 steps and 300 (remaining) steps of pixels of the masked image.
Following is are the 3 examples of the predicted pixels and cross entropy comparison with the ground truth. The cross entropy's for the 10, 28 and 300 steps have been averages with 10 samples. Additionally you can find all the values of the cross entropy in the table in appendix section at the end of the report.
And out of the generated data set, picked 3 examples which demonstrates successful, failed and high variance between samples in pixel predictions.

Cross Entropy Comparison and Visualize completing the image:

1. Model - GRU (32 Units, 1 Layer)

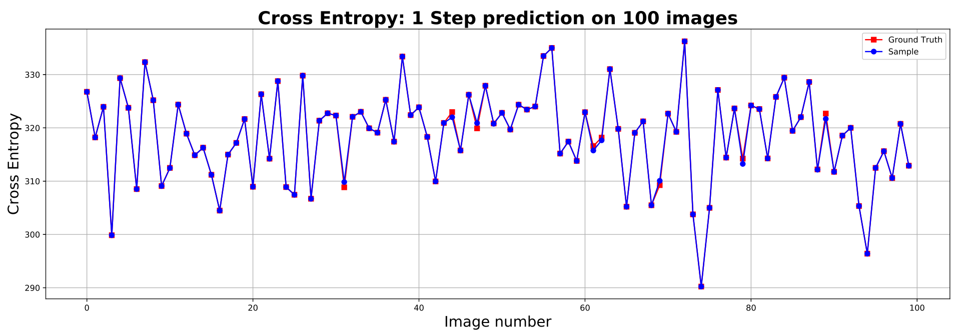


Figure 23: Cross Entropy 1 Step, 32 unit GRU

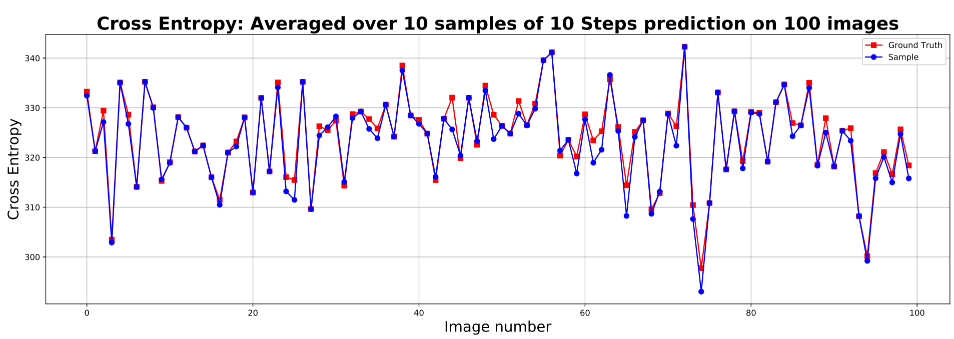


Figure 24: Cross Entropy 10 Steps, 32 unit GRU

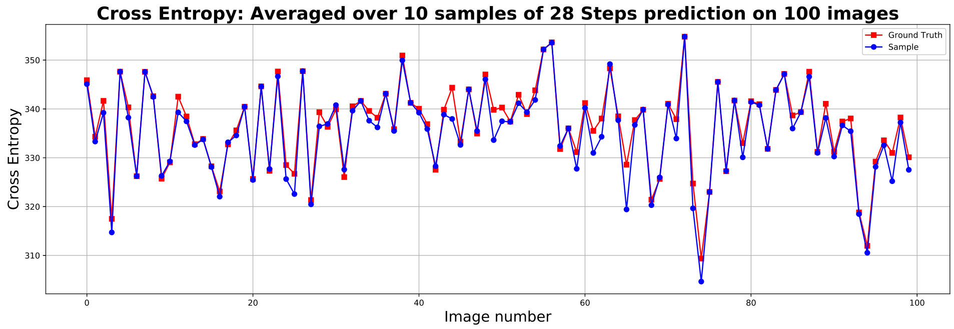


Figure 25: Cross Entropy 28 Steps, 32 unit GRU

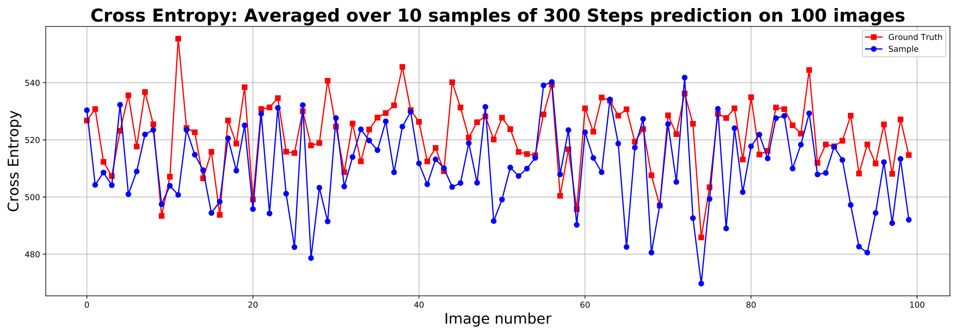


Figure 26: Cross Entropy 300 Steps, 32 unit GRU

Task 2: [(a)] continued on next page. . .

Successful:

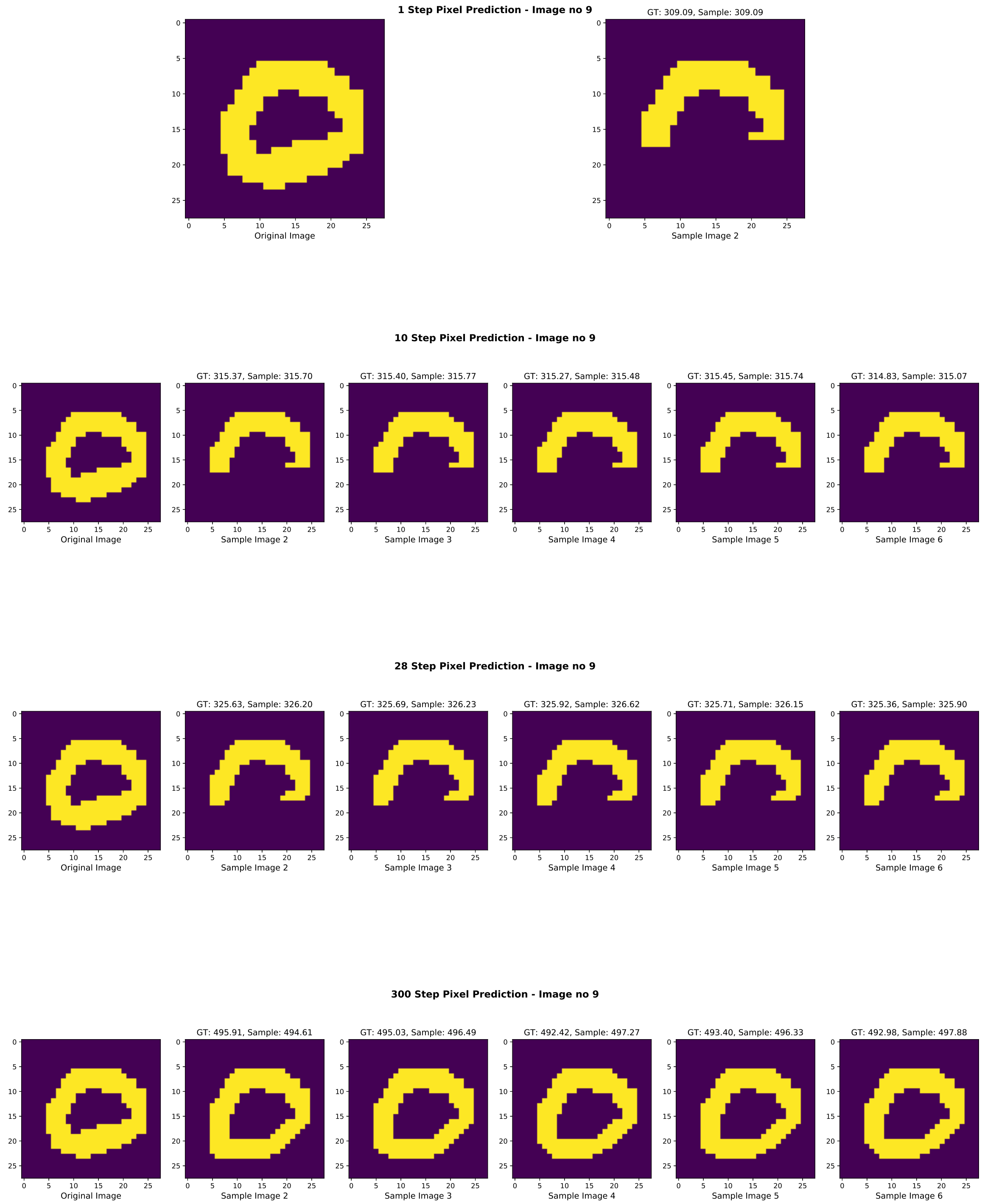


Figure 27: 1, 10, 28, 300 step pixel prediction, 32 layer, 1 layer GRU

Failure:

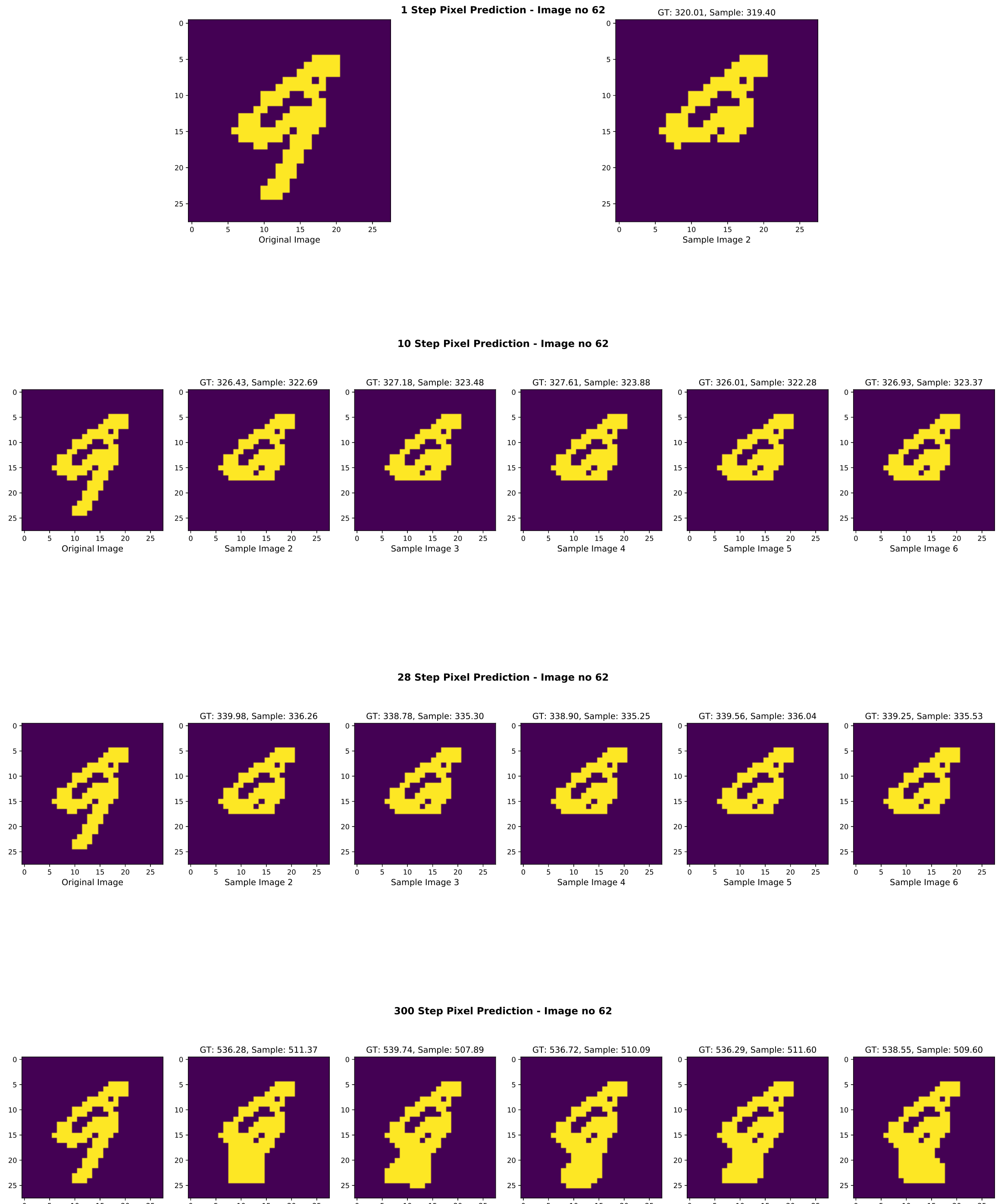


Figure 28: 1, 10, 28, 300 step pixel prediction, 32 unit, 1 layer GRU

Variance:

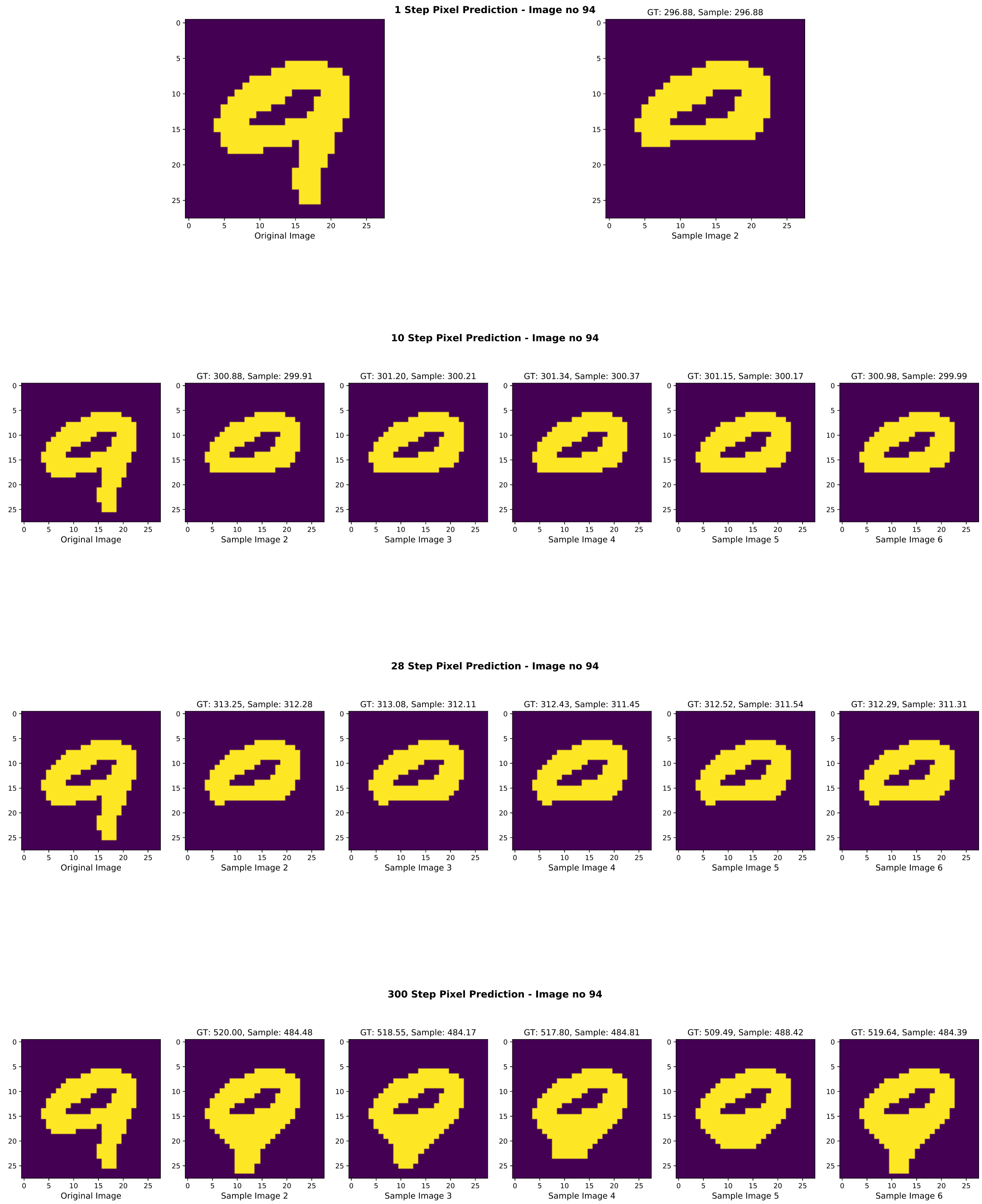


Figure 29: 1, 10, 28, 300 step pixel prediction, 32 unit, 1 layer GRU

2. Model - GRU (64 Units, 1 Layer)

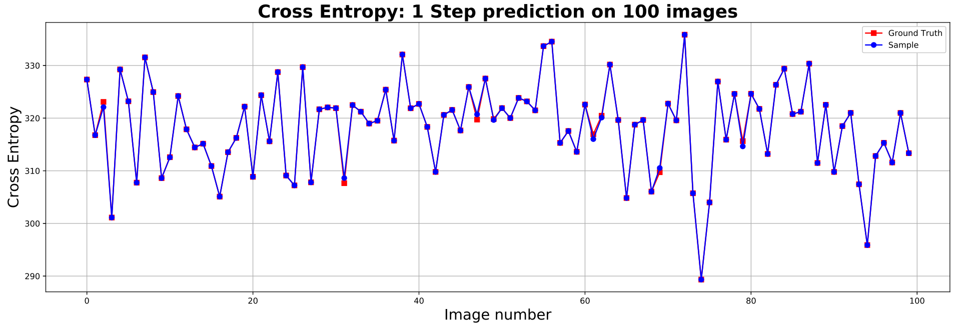


Figure 30: Cross Entropy 1 Step, 64 unit GRU

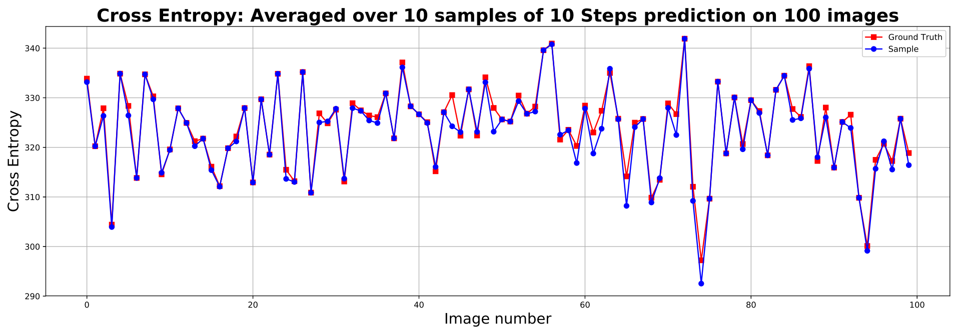


Figure 31: Cross Entropy 10 Steps, 64 unit GRU

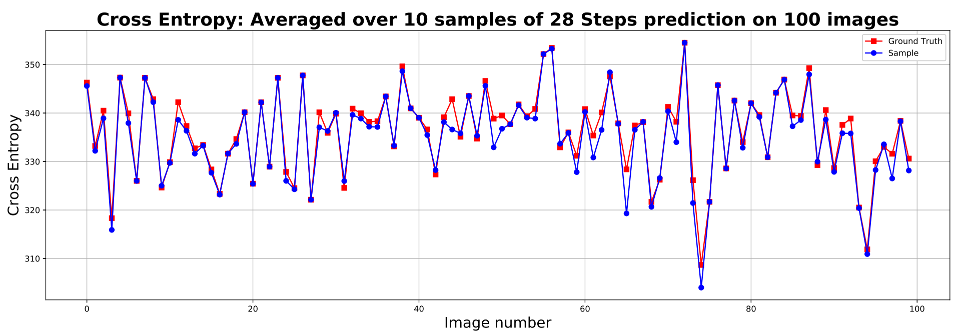


Figure 32: Cross Entropy 28 Steps, 64 unit GRU

Task 2: [(a)] continued on next page...

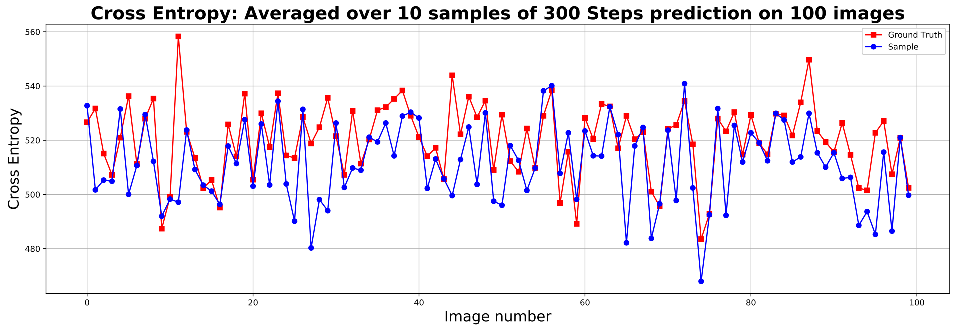
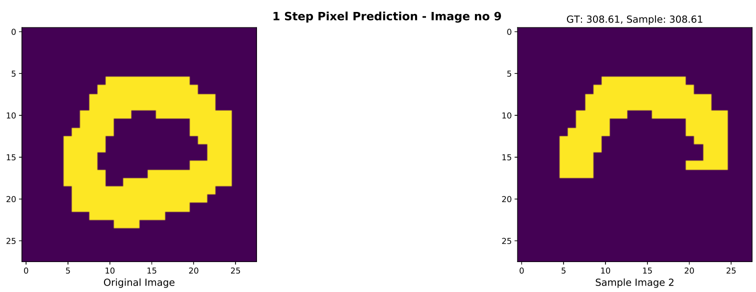
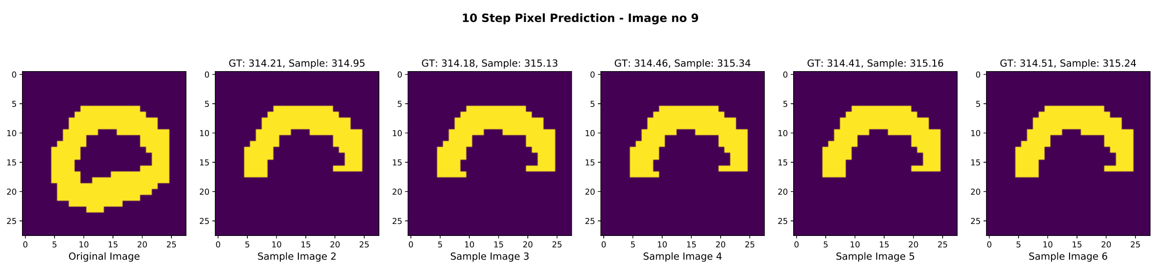
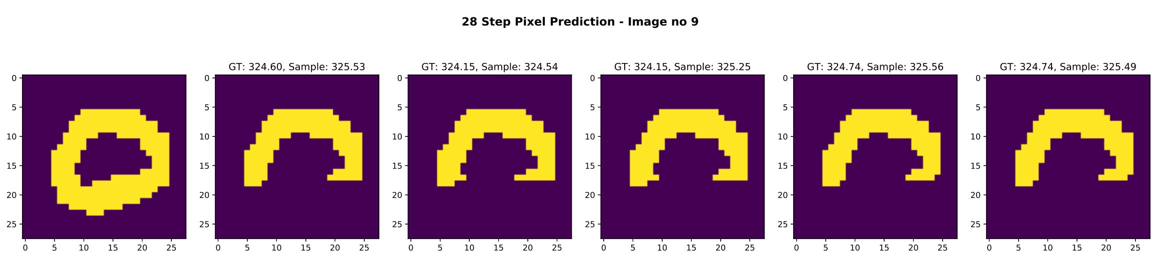


Figure 33: Cross Entropy 300 Steps, 64 unit GRU

Successful:







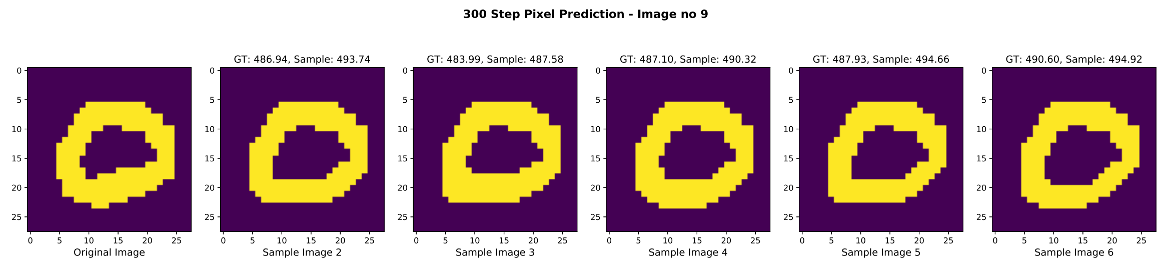
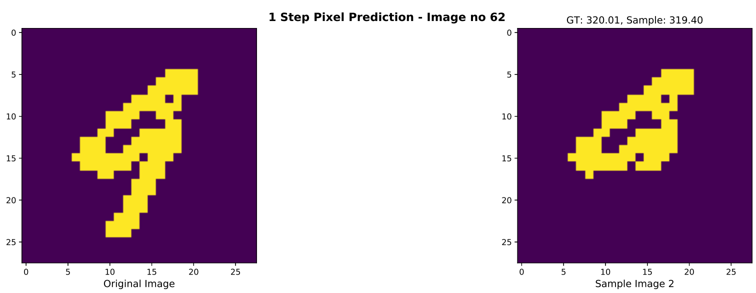
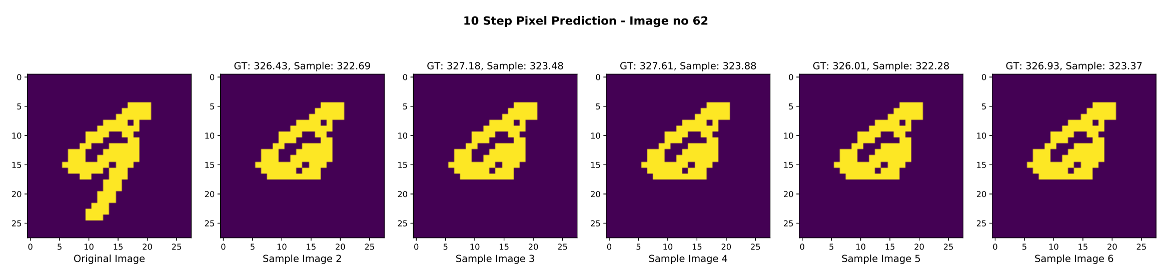
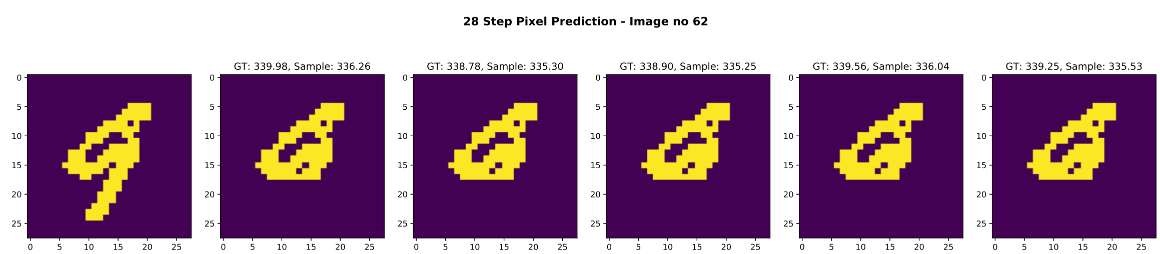


Figure 34: 1, 10, 28, 300 step pixel prediction, 64 layer, 1 layer GRU

Failure:







Task 2: [(a)] continued on next page. . .

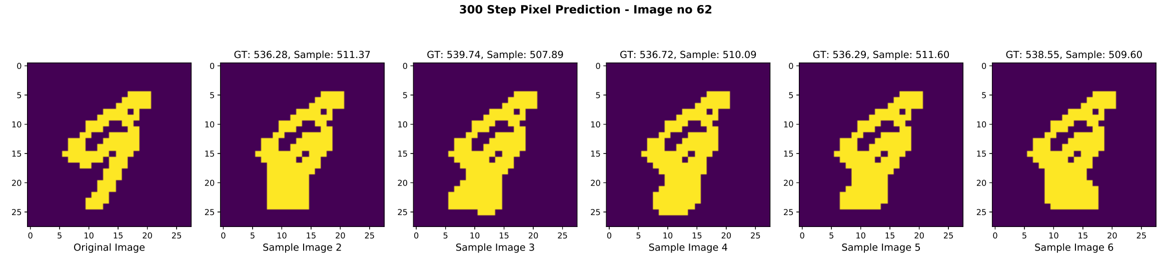
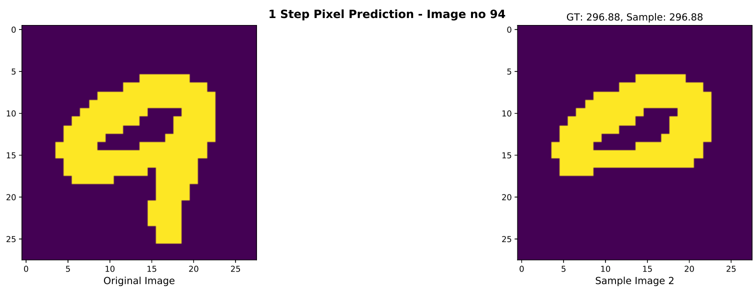
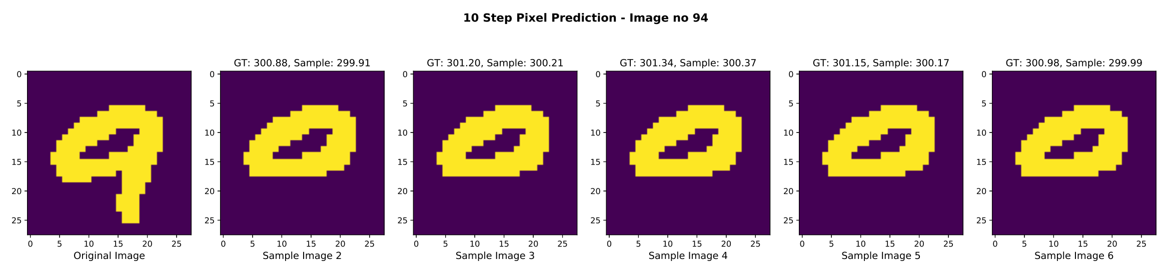
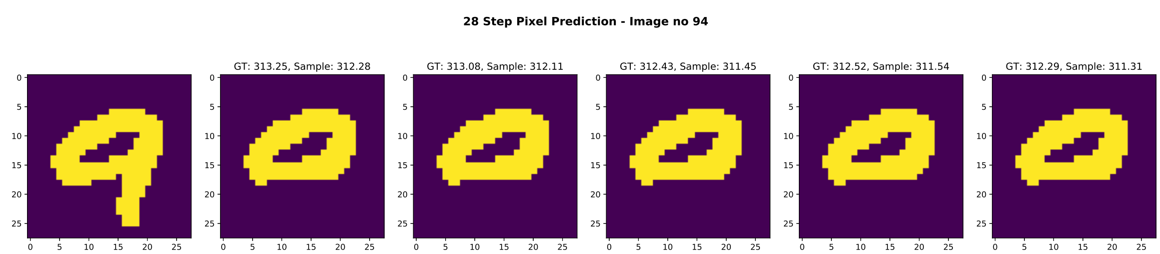


Figure 35: 1, 10, 28, 300 step pixel prediction, 64 unit, 1 layer GRU

Variance:







Task 2: [(a)] continued on next page. . .

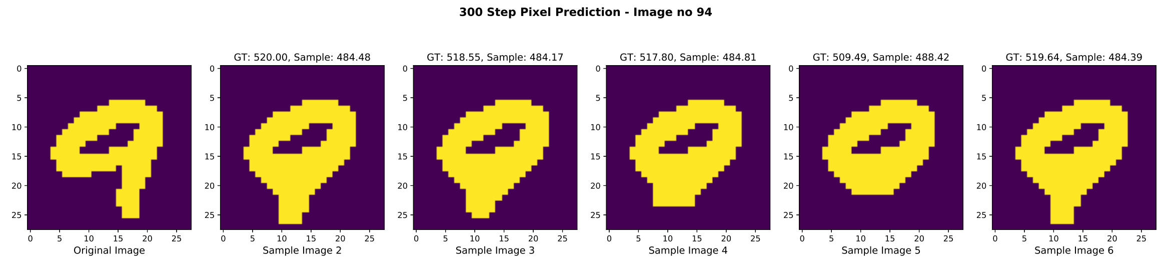


Figure 36: 1, 10, 28, 300 step pixel prediction, 64 unit, 1 layer GRU

3. Model - GRU (128 Units, 1 Layer)

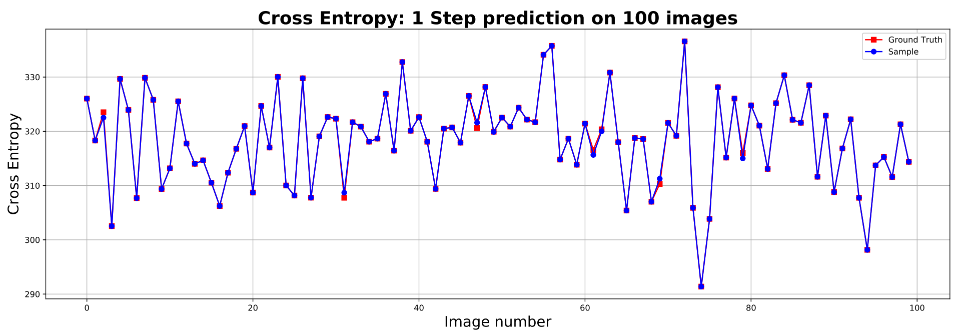


Figure 37: Cross Entropy 1 Step, 128 unit GRU

=

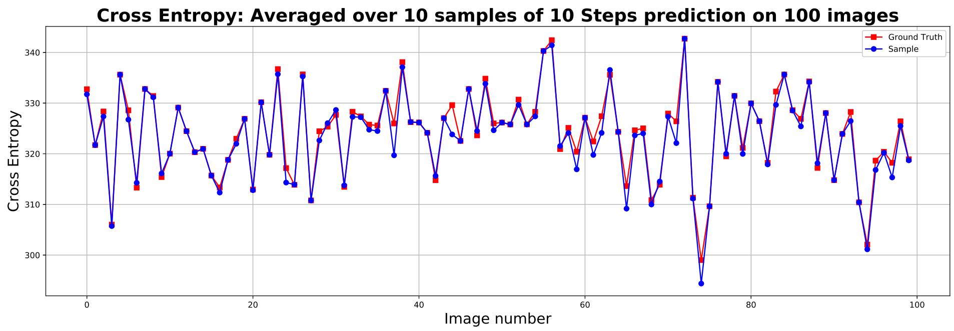


Figure 38: Cross Entropy 10 Steps, 128 unit GRU

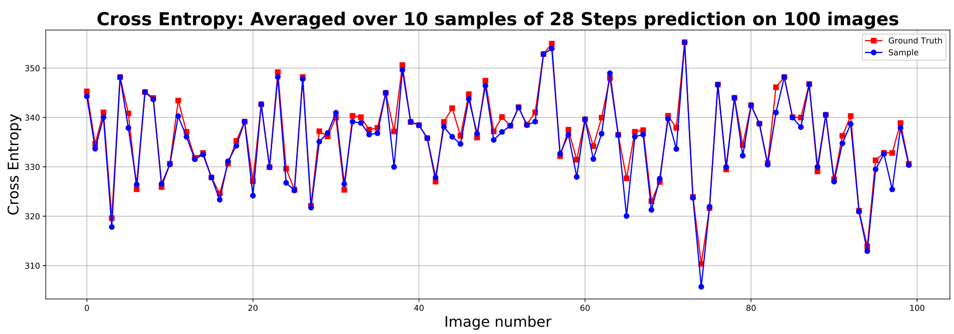


Figure 39: Cross Entropy 28 Steps, 128 unit GRU

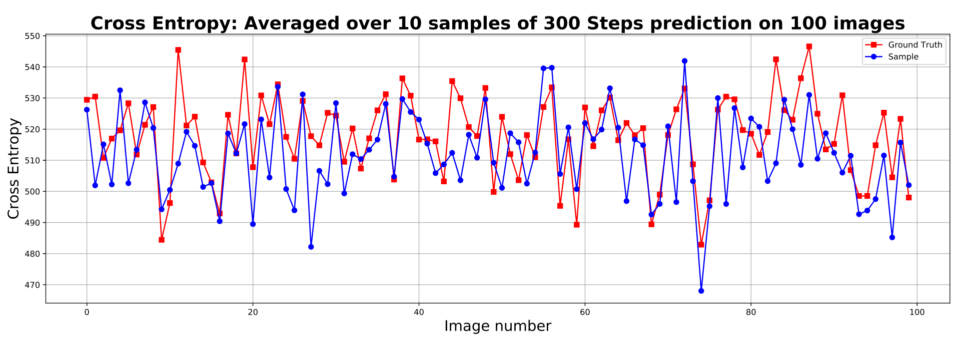


Figure 40: Cross Entropy 300 Steps, 128 unit GRU

Successful:

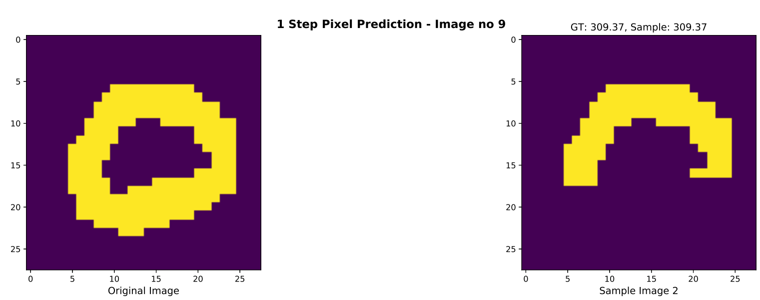


Figure 41: 1, 10, 28, 300 step pixel prediction, 128 layer, 1 layer GRU

Task 2: [(a)] continued on next page…

Failure:

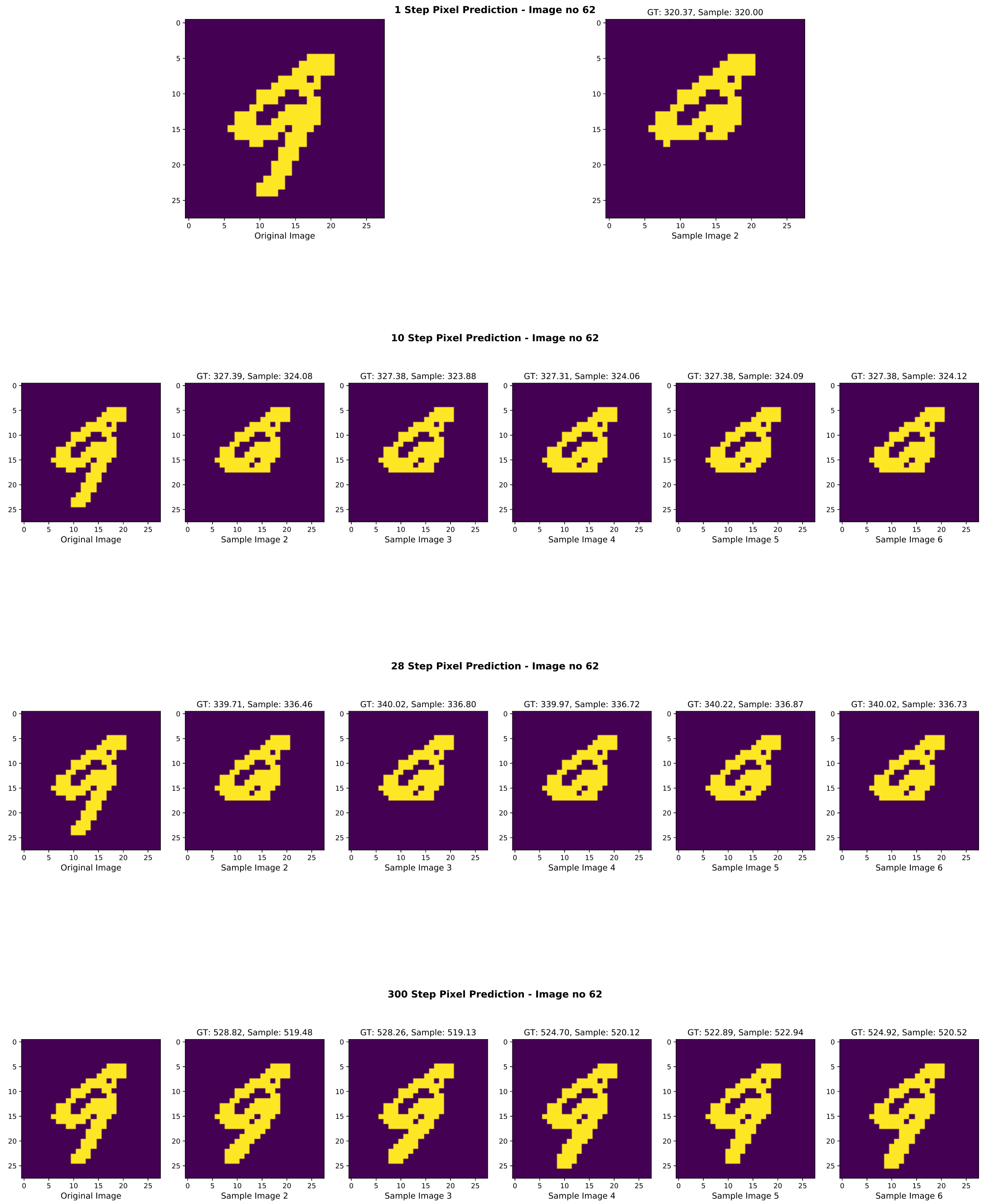


Figure 42: 1, 10, 28, 300 step pixel prediction, 128 unit, 1 layer GRU

Task 2: [(a)] continued on next page...

Variance:

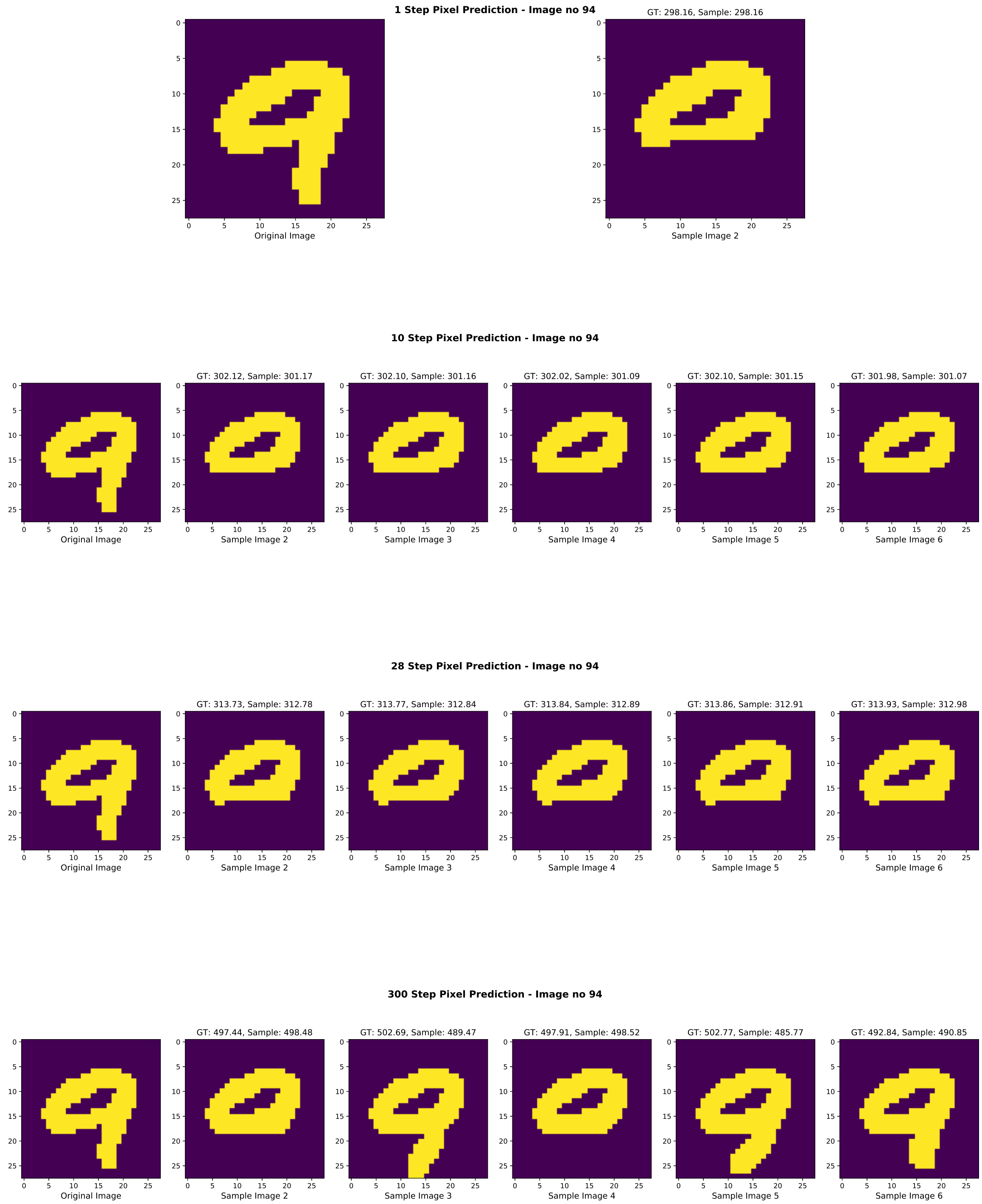


Figure 43: 1, 10, 28, 300 step pixel prediction, 128 unit, 1 layer GRU

Task 2: [(a)] continued on next page...

4. Model - GRU (32 Units, 3 Layer)

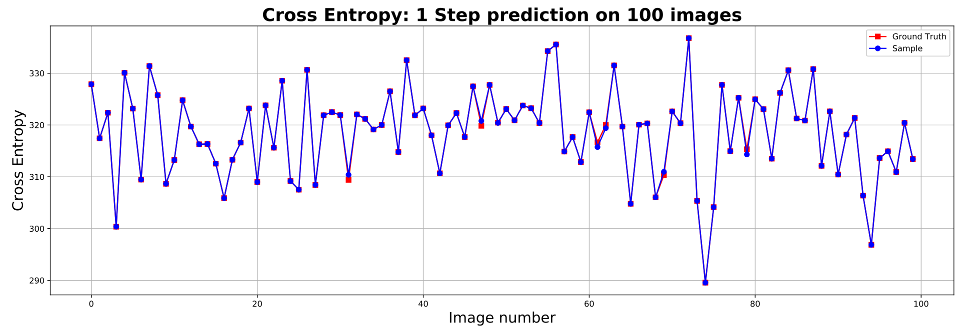


Figure 44: Cross Entropy 1 Step, 32 unit, 3 layer GRU

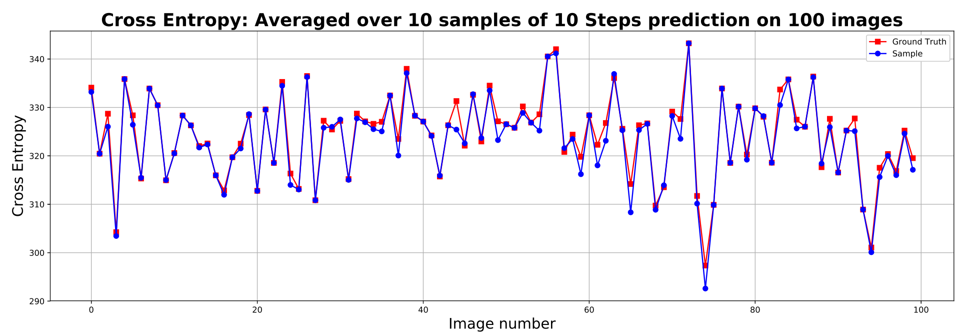


Figure 45: Cross Entropy 10 Steps, 32 unit, 3 layer GRU

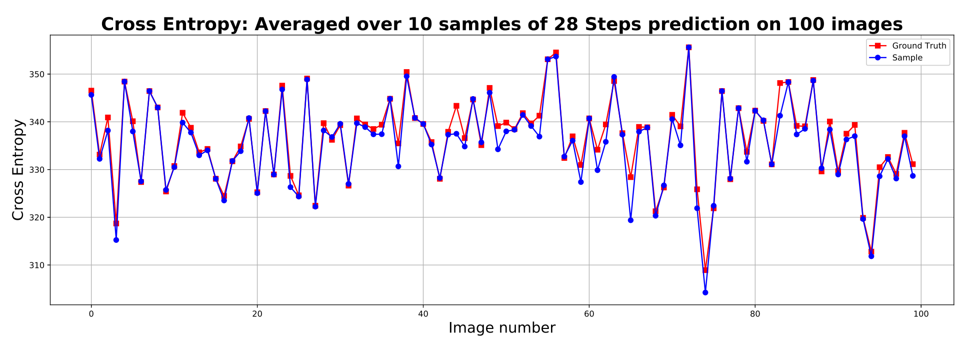


Figure 46: Cross Entropy 28 Steps, 32 unit, 3 layer GRU

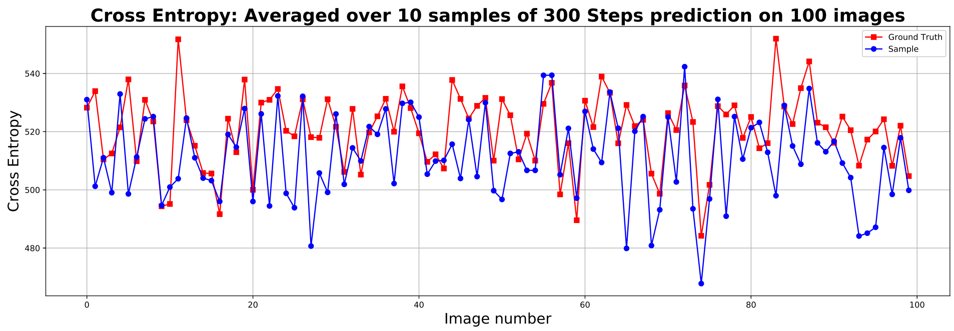
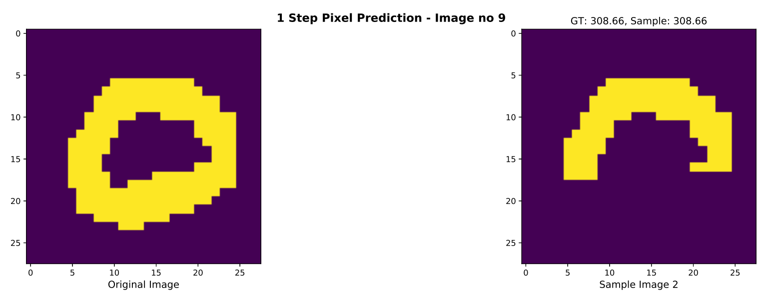
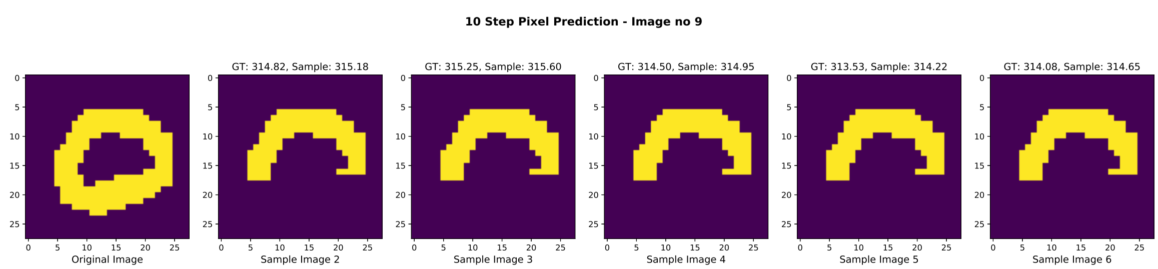
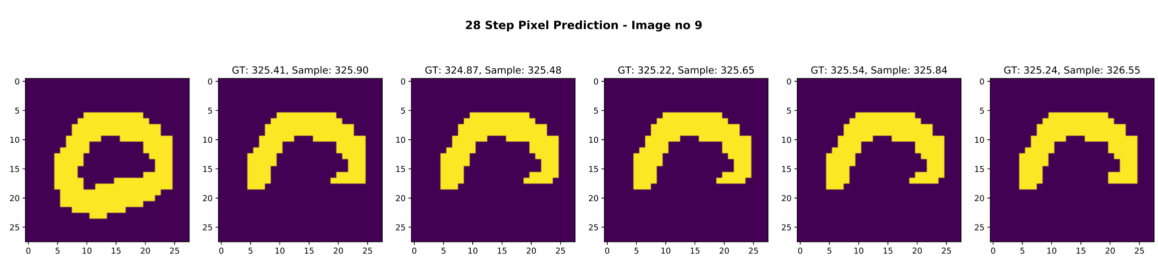


Figure 47: Cross Entropy 300 Steps, 32 unit, 3 layer GRU

Successful







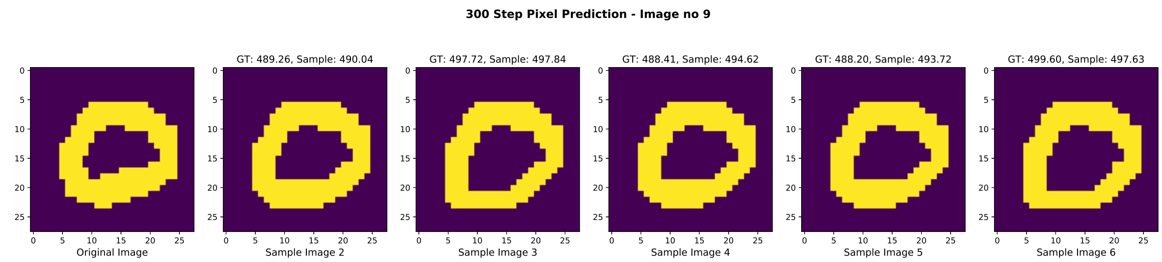
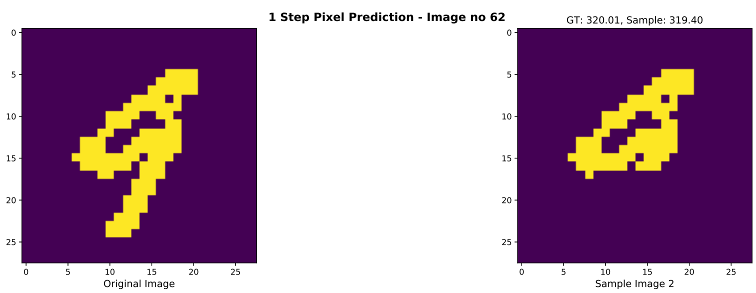
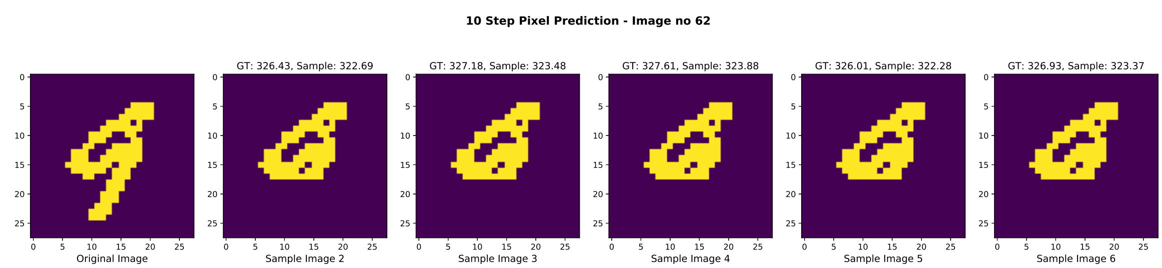
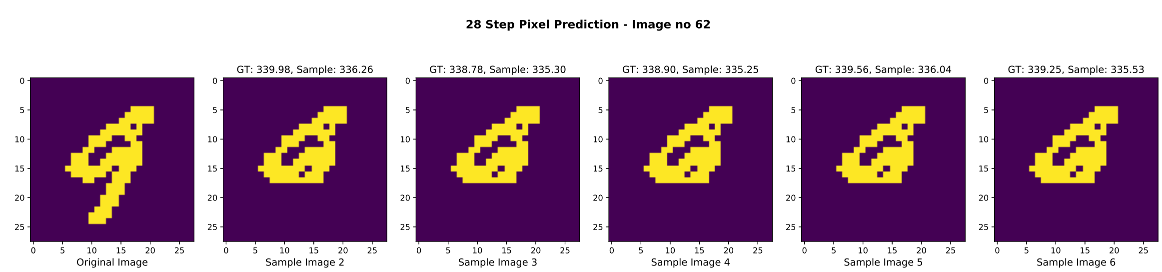


Figure 48: 1, 10, 28, 300 step pixel prediction, 32 unit, 3 layer GRU

Failure







Task 2: [(a)] continued on next page...

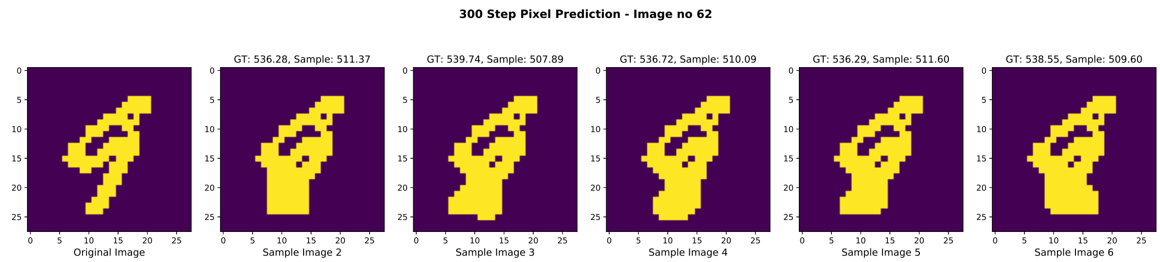
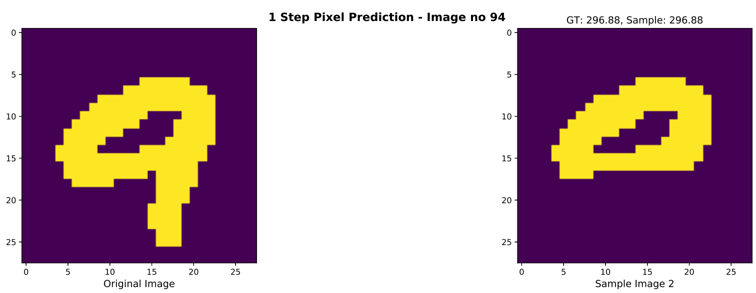
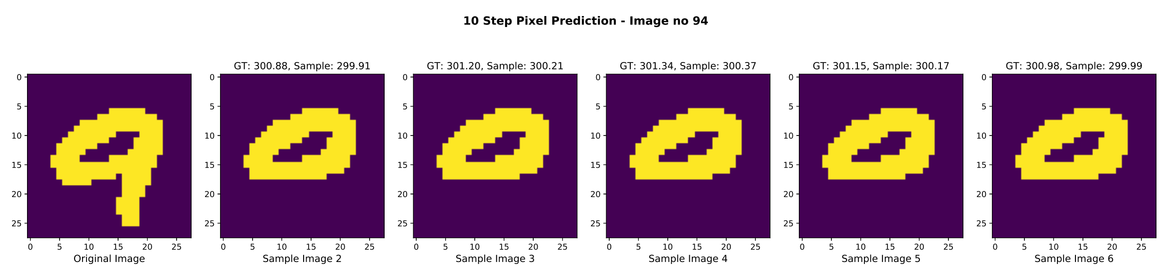
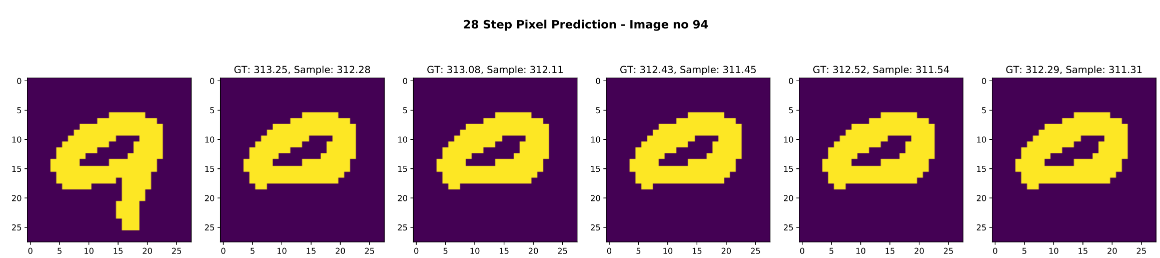


Figure 49: 1, 10, 28, 300 step pixel prediction, 32 unit, 3 layer GRU

Variance







Task 2: [(a)] continued on next page. . .

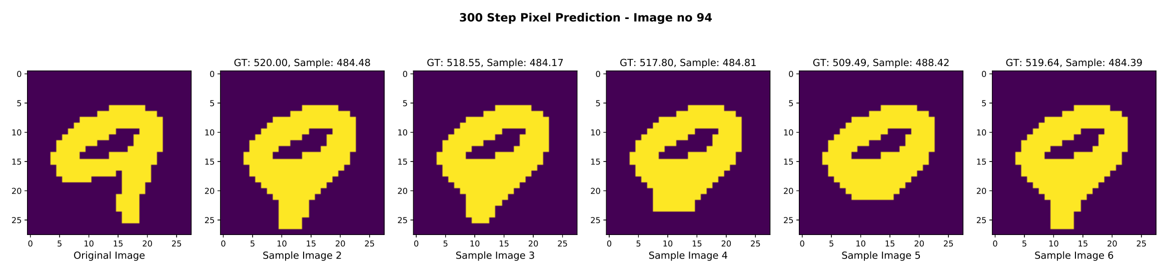


Figure 50: 1, 10, 28, 300 step pixel prediction, 32 unit, 3 layer GRU

Task 3: In-painting

(a)

One-pixel missing

Using the pixel prediction technique from Task 2b, we can formulate an equation to find the missing pixel in an image. In this task we can use the pixel information before and after the missing pixel take advantage of this to find most probable pixel.
Lets have some mathematical proof for this. Assume:

- **x** be the binarized input image vector of the shape 1 x 784.
- **m** be the missing pixel index
- N be the total length of the image vector **x**

As per the problem we have past pixel information as $x_{1:m-1}$ and future pixels as $x_{m+1:N}$. By Bayes rule we can say,

$$P(x_m|x_{1:m-1}, x_{m+1:N}) = \frac{P(x_m, x_{1:m-1}, x_{m+1,N})}{P(x_{1:m-1}, x_{m+1,N})} \tag{15}$$

$$P(x_m, x_{1:m-1}, x_{m+1:N}) = P(x_{m+1:N}|x_m, x_{1:m-1})P(x_m, x_{1:m-1}) \tag{16}$$

$$P(x_m, x_{1:m-1}, x_{m+1:N}) = P(x_{m+1:N}|x_m, x_{1:m-1})P(x_m|x_{1:m-1})P(x_{1:m-1}) \tag{17}$$

Which shows product of past, present and future pixel sequence probabilities.
So we can rewrite eq 15 as

$$P(x_m|x_{1:m-1}, x_{m+1:N}) = \frac{P(x_{m+1:N}|x_m, x_{1:m-1})P(x_m|x_{1:m-1})P(x_{1:m-1})}{P(x_{m+1:N}|x_{1,m-1})P(x_{1:m-1})} \tag{18}$$

Which is same as

$$P(x_m|x_{1:m-1}, x_{m+1:N}) = \frac{P(x_{m+1:N}|x_m, x_{1:m-1})P(x_m|x_{1:m-1})P(x_{1:m-1})}{\sum_{x_m} P(x_{m+1:N}|x_m, x_{1:m-1})P(x_m|x_{1:m-1})P(x_{1:m-1})} \tag{19}$$

Now using the above equation we can write the cross entropy equation which we want to minimize.

$$x_m = argmin_{x_m} \sum_t x_t \cdot log(P(x_t|x_{1:m}, x_m)) \tag{20}$$

Using the eq 20 we can see that the selection of x_m would affect the future pixels probability. Hence with the selection of m pixel we shall see which choice of pixel minimises the cross entropy, and that would be the most probable missing pixel.

Window of 2x2 pixels missing

Using the similar principle as used in one missing pixel we can extend the formulation for 2x2 missing pixel. So for pixels $x_m, x_{m+1}, x_n, x_{n+1}$ we can write the equation as follows:

$$(x_m, x_{m+1}, x_n, x_{n+1}) = argmin_{x_m, x_{m+1}, x_n, x_{n+1}} \sum_t x_t \cdot log(P(x_t|x_m, x_{m+1}, x_n, x_{n+1})) \tag{21}$$

where m and n are the missing pixel index for the window in the vector **x**

(b)

One-pixel missing visualization (128 units, 1 layer GRU)

In order to perform the in painting of the missing pixels we first take all possible versions of the given image with one missing pixel. In this case we have 2 cases, once with fusing missing pixel as 0 and other with fusing missing pixel with 1. Now we take both of these images and pass it through the 128 unit, 1 layer GRU model. Then we take cross entropy of both of these images (considering past and future information) and compare which ever has low cross entropy, that is the most probable in painted version of the image.

It turns out to be that the most probable image has almost the same cross entropy with that of the ground truth, with precision up to decimal places.
In-painting accuracy: 0.9998

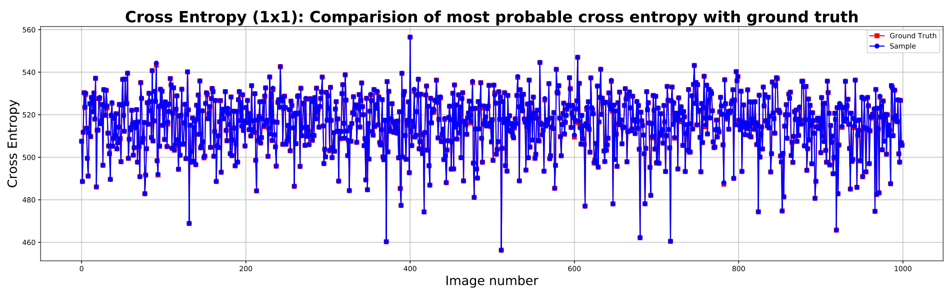


Figure 51: 1x1 cross entropy comparison

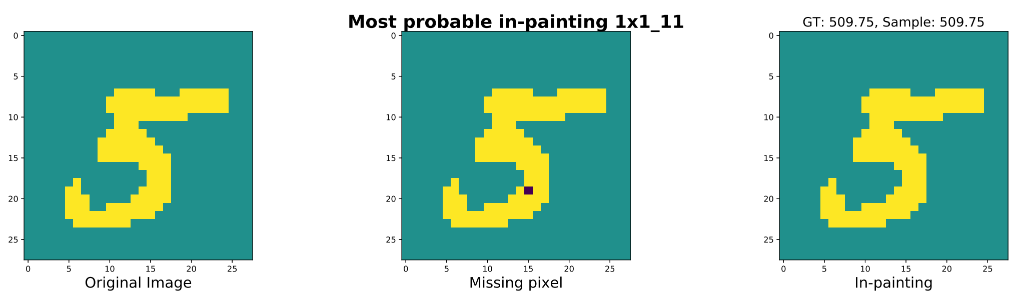


Figure 52: 1 Missing pixel prediction visualization

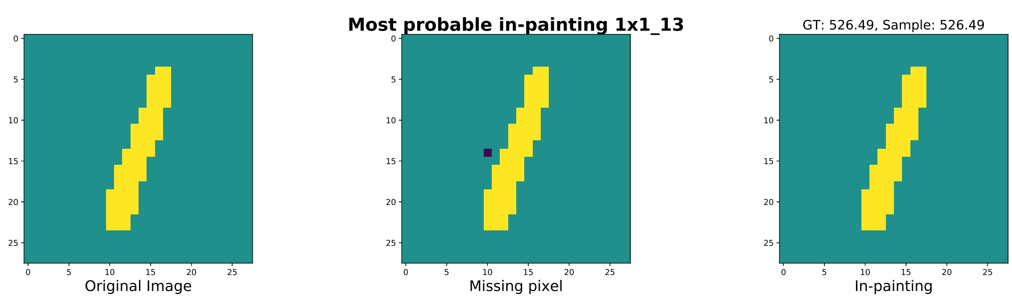


Figure 53: 1 Missing pixel prediction visualization

Window of 2x2 pixels missing visualization (128 units, 1 layer GRU)

For the 2x2 window, we do similar process for predicting the most probable in painting, but in this case we have 4 unknown pixels to predict. Which makes 2^4 combinations of the 2x2 pixel window. So we take all such possible version of the image and pass it through the model to get the 16 cross entropy's of these probable images. And again, whichever has the lowest cross entropy we choose that to be the most probable image.

It is also observed that the cross entropy is almost same as the cross entropy of the ground truth image. Which proves that the image we selected is the most probable one.
In-painting accuracy: 0.9988

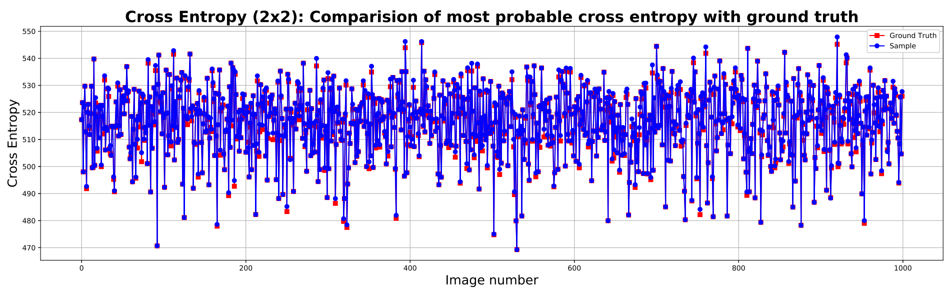


Figure 54: 1x1 cross entropy comparison

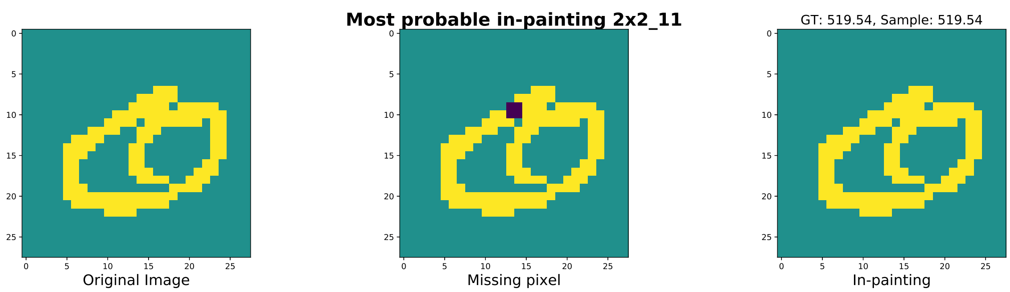


Figure 55: 1 Missing pixel prediction visualization

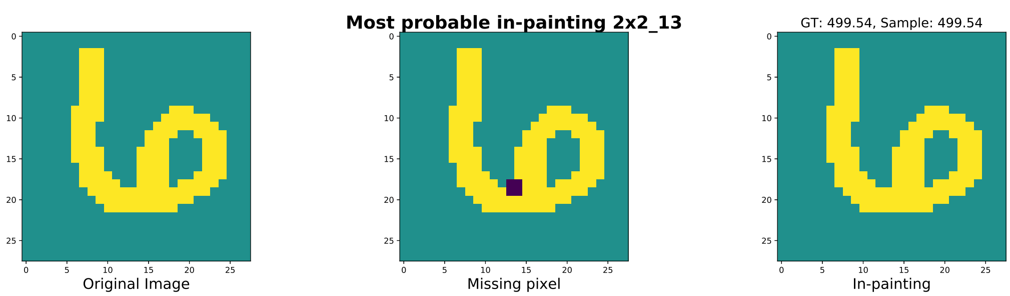


Figure 56: 1 Missing pixel prediction visualization

Appendices

Task 2 a (cross entropy)

1. Cross entropy table (GRU 32 unit, 1 layer)

Following are the values of the cross entropy of ground truth for 1, 10, 28, 300 step (xentGT1, xentGT10, xentGT28, xentGT300 respectively) and sample for 1, 10, 28, 300 steps (xentSAM1, xentSAM10, xentSAM28, xentSAM300 respectively)

Table 4: Cross entropy table (GRU 32 unit, 1 layer)

xentGT1	xentGT10	xentGT28	xentGT300	xentSAM1	xentSAM10	xentSAM28	xentSAM300
326.77	333.24	345.89	526.77	326.77	332.44	345.09	530.37
318.21	321.27	334.31	530.79	318.21	321.27	333.32	504.27
323.95	329.43	341.65	512.30	323.95	327.14	339.22	508.57
299.86	303.48	317.49	507.33	299.86	302.90	314.71	504.14
329.34	335.08	347.63	523.17	329.34	335.08	347.63	532.29
323.77	328.62	340.32	535.59	323.77	326.77	338.23	500.99
308.52	314.09	326.23	517.66	308.52	314.09	326.23	508.96
332.34	335.23	347.60	536.75	332.34	335.23	347.60	521.90
325.19	330.12	342.61	525.44	325.19	329.98	342.47	523.50
309.09	315.30	325.73	493.38	309.09	315.60	326.30	497.47
312.49	319.07	329.06	507.11	312.47	318.94	329.25	503.93
324.37	328.13	342.52	555.40	324.37	328.14	339.30	500.74
318.92	326.01	338.45	524.15	318.92	326.01	337.46	523.49
314.87	321.22	332.79	522.65	314.87	321.22	332.60	514.82
316.28	322.42	333.85	506.56	316.28	322.42	333.76	509.40
311.20	316.04	328.26	515.80	311.20	316.04	328.12	494.41
304.48	311.49	323.10	493.75	304.48	310.49	322.04	498.37
314.99	321.00	332.74	526.80	314.99	321.00	333.17	520.57
317.18	323.23	335.61	518.70	317.18	322.23	334.57	509.24
321.65	328.07	340.45	538.43	321.65	328.07	340.44	525.08
308.95	312.98	325.66	499.16	308.95	312.95	325.45	495.79
326.32	331.99	344.62	530.80	326.32	331.95	344.59	529.16
314.22	317.21	327.33	531.37	314.22	317.21	327.71	494.25
328.78	335.10	347.66	534.61	328.78	334.10	346.67	531.21

xentGT1	xentGT10	xentGT28	xentGT300	xentSAM1	xentSAM10	xentSAM28	xentSAM300
308.89	316.05	328.51	515.86	308.89	313.19	325.64	501.17
307.45	315.48	326.72	515.41	307.45	311.48	322.57	482.47
329.80	335.22	347.74	530.06	329.80	335.22	347.74	532.17
306.70	309.63	321.34	518.05	306.70	309.63	320.46	478.66
321.35	326.30	339.33	518.90	321.35	324.43	336.44	503.27
322.74	325.51	336.34	540.70	322.74	326.09	336.93	491.47
322.32	327.40	339.91	524.68	322.32	328.27	340.78	527.64
308.85	314.35	326.04	508.73	309.83	315.09	327.58	503.68
322.07	328.74	340.54	525.71	322.07	327.94	339.63	513.99
323.02	329.24	341.64	512.50	323.02	329.24	341.64	523.69
319.93	327.74	339.59	523.61	319.93	325.75	337.60	519.78
319.10	325.86	338.20	527.80	319.10	323.88	336.22	516.40
325.27	330.63	343.12	529.40	325.27	330.63	343.12	526.48
317.42	324.20	335.86	532.10	317.42	324.21	335.50	508.66
333.38	338.50	350.94	545.52	333.38	337.51	349.95	524.64
322.40	328.46	341.24	530.42	322.40	328.48	341.26	529.85
323.86	327.57	340.04	526.32	323.86	326.78	339.24	511.78
318.31	324.80	336.89	512.46	318.31	324.80	335.89	504.46
309.95	315.45	327.56	517.18	309.95	316.12	328.26	513.20
320.92	327.80	339.84	509.04	320.92	327.80	338.84	510.13
322.97	332.04	344.36	540.17	321.99	325.65	337.95	503.52
315.76	319.87	333.30	531.34	315.76	320.40	332.63	504.89
326.22	332.02	344.01	520.81	326.22	332.02	344.01	518.87
319.91	322.58	334.97	526.18	320.90	323.31	335.48	504.98
327.90	334.46	347.06	528.20	327.90	333.47	346.06	531.54
320.83	328.60	339.82	520.14	320.77	323.70	333.62	491.60
322.82	326.36	340.29	527.77	322.82	326.36	337.48	499.13
319.69	324.83	337.41	523.77	319.69	324.81	337.35	510.33
324.35	331.36	342.90	515.77	324.35	328.82	341.17	507.37
323.44	326.54	338.95	515.06	323.44	326.47	339.34	509.95
324.01	330.82	343.82	514.61	324.01	329.82	341.85	513.69
333.48	339.58	352.18	528.92	333.48	339.58	352.18	539.07

xentGT1	xentGT10	xentGT28	xentGT300	xentSAM1	xentSAM10	xentSAM28	xentSAM300
335.00	341.18	353.64	539.29	335.00	341.10	353.56	540.26
315.18	320.41	331.79	500.43	315.18	321.39	332.41	507.94
317.43	323.55	336.03	516.70	317.43	323.55	336.03	523.43
313.80	320.21	331.15	495.73	313.80	316.78	327.74	490.26
322.94	328.69	341.20	531.04	322.94	327.70	340.21	522.66
316.59	323.44	335.51	522.86	315.75	318.96	330.99	513.68
318.18	325.29	338.04	534.83	317.67	321.57	334.30	508.69
331.03	335.75	348.32	533.74	331.03	336.61	349.19	534.14
319.81	326.14	338.48	528.48	319.81	325.34	337.69	518.70
305.20	314.42	328.59	530.71	305.20	308.25	319.41	482.52
319.07	325.14	337.72	519.35	319.07	324.15	336.73	517.32
321.22	327.49	339.84	523.79	321.22	327.49	339.87	527.33
305.48	309.60	321.41	507.61	305.48	308.68	320.29	480.55
309.25	312.84	325.67	497.01	310.07	313.14	325.98	497.10
322.68	328.85	341.06	528.59	322.68	328.75	340.93	525.53
319.25	326.31	337.90	522.01	319.25	322.38	333.95	505.27
336.24	342.27	354.79	536.22	336.24	342.27	354.79	541.80
303.77	310.43	324.72	525.63	303.77	307.64	319.64	492.62
290.22	297.75	309.36	485.88	290.22	293.02	304.64	469.72
304.99	310.84	323.00	503.43	304.99	310.84	323.00	499.32
327.11	333.08	345.54	529.14	327.11	333.08	345.54	530.88
314.43	317.62	327.25	527.66	314.43	317.62	327.31	489.00
323.65	329.30	341.72	531.04	323.65	329.30	341.72	524.08
314.24	319.30	332.99	513.14	313.24	317.81	330.09	501.72
324.21	329.18	341.59	534.94	324.21	329.03	341.42	517.76
323.54	328.98	340.98	514.88	323.54	328.79	340.78	521.84
314.26	319.18	331.82	516.15	314.26	319.18	331.83	513.52
325.81	331.13	343.88	531.33	325.81	331.13	343.88	527.59
329.42	334.68	347.17	530.77	329.42	334.68	347.15	528.40
319.44	326.93	338.68	525.12	319.44	324.27	335.99	509.97
322.02	326.48	339.35	522.19	322.02	326.48	339.35	518.33
328.61	335.03	347.62	544.45	328.61	334.01	346.61	529.27

xentGT1	xentGT10	xentGT28	xentGT300	xentSAM1	xentSAM10	xentSAM28	xentSAM300
312.20	318.56	331.20	511.93	312.15	318.37	331.01	507.91
322.69	327.91	341.05	518.42	321.69	325.00	338.15	508.43
311.75	318.21	331.25	517.74	311.75	318.21	330.26	517.57
318.55	325.38	337.43	519.70	318.55	325.38	336.63	512.97
320.01	325.91	338.05	528.47	320.01	323.38	335.44	497.24
305.33	308.21	318.80	508.22	305.33	308.21	318.47	482.65
296.39	300.16	311.97	518.40	296.39	299.21	310.54	480.56
312.49	316.88	329.21	511.75	312.49	315.83	328.15	494.43
315.63	321.10	333.57	525.39	315.63	320.12	332.56	512.21
310.59	316.66	331.00	508.10	310.59	314.99	325.20	490.87
320.76	325.65	338.25	527.15	320.76	324.72	337.22	513.34
312.90	318.41	330.11	514.69	312.90	315.79	327.53	492.04

2. Cross entropy table (GRU 64 unit, 1 layer)

Following are the values of the cross entropy of ground truth for 1, 10, 28, 300 step (xentGT1, xentGT10, xentGT28, xentGT300 respectively) and sample for 1, 10, 28, 300 steps (xentSAM1, xentSAM10, xentSAM28, xentSAM300 respectively)

Table 5: Cross entropy table (GRU 64 unit, 1 layer)

xentGT1	xentGT10	xentGT28	xentGT300	xentSAM1	xentSAM10	xentSAM28	xentSAM300
327.33	333.85	346.27	526.65	327.33	333.16	345.58	532.75
316.77	320.23	333.20	531.72	316.77	320.23	332.21	501.69
323.08	327.89	340.48	515.10	322.09	326.35	338.92	505.29
301.12	304.42	318.31	507.23	301.12	303.94	315.90	504.88
329.25	334.86	347.30	520.98	329.25	334.86	347.30	531.53
323.20	328.36	339.94	536.27	323.20	326.43	337.93	500.03
307.75	313.84	326.01	511.20	307.75	313.84	326.01	510.70
331.54	334.73	347.25	527.97	331.54	334.73	347.25	529.46
324.95	330.29	342.86	535.39	324.95	329.69	342.24	512.20
308.61	314.56	324.64	487.40	308.61	314.88	325.01	491.99
312.58	319.58	329.87	499.14	312.56	319.49	329.73	498.29
324.20	327.87	342.24	558.29	324.20	327.87	338.60	497.15
317.88	324.93	337.33	523.11	317.88	324.93	336.33	523.75
314.44	321.24	332.72	513.47	314.44	320.25	331.62	509.20
315.14	321.77	333.41	502.39	315.14	321.77	333.28	503.43
310.90	316.10	328.38	505.33	310.90	315.41	327.68	501.22
305.09	312.18	323.34	495.18	305.09	312.08	323.18	496.33
313.53	319.83	331.62	525.83	313.53	319.83	331.66	517.85
316.25	322.20	334.64	514.02	316.25	321.20	333.64	511.41
322.18	327.92	340.17	537.21	322.18	327.92	340.17	527.59
308.86	312.93	325.43	505.47	308.86	312.93	325.41	503.07
324.36	329.70	342.23	529.94	324.36	329.67	342.20	526.00
315.60	318.54	328.94	517.51	315.60	318.54	328.94	503.51
328.75	334.83	347.27	537.32	328.75	334.83	347.27	534.42
309.10	315.48	327.85	514.38	309.10	313.65	326.01	503.89
307.23	313.18	324.52	513.46	307.23	313.03	324.27	490.14
329.69	335.18	347.76	528.56	329.69	335.18	347.76	531.40

xentGT1	xentGT10	xentGT28	xentGT300	xentSAM1	xentSAM10	xentSAM28	xentSAM300
307.81	310.86	322.13	518.81	307.81	310.86	322.13	480.28
321.68	326.86	340.15	524.81	321.68	325.02	337.05	498.11
322.04	324.85	335.92	535.65	322.04	325.24	336.30	494.03
321.90	327.61	339.87	521.51	321.90	327.79	340.06	526.33
307.64	313.12	324.56	507.23	308.62	313.70	325.98	502.58
322.49	328.91	340.90	530.82	322.49	327.91	339.63	509.77
321.22	327.46	339.96	511.41	321.22	327.35	338.86	508.95
318.97	326.42	338.18	520.28	318.97	325.42	337.18	521.11
319.52	326.09	338.32	531.08	319.52	324.91	337.13	519.39
325.41	330.89	343.43	532.22	325.41	330.89	343.43	526.35
315.73	321.82	333.13	535.26	315.73	321.84	333.26	514.24
332.09	337.12	349.63	538.36	332.09	336.12	348.63	528.91
321.88	328.27	340.96	529.04	321.88	328.30	340.99	530.36
322.71	326.67	339.02	521.16	322.71	326.67	339.03	528.21
318.34	325.07	336.62	514.13	318.34	324.92	335.46	502.23
309.79	315.18	327.33	517.21	309.79	316.06	328.23	513.09
320.61	327.09	339.13	505.66	320.61	327.09	338.13	505.77
321.56	330.54	342.86	543.93	321.56	324.25	336.60	499.59
317.66	322.36	335.11	522.20	317.66	323.06	335.82	512.90
325.92	331.70	343.51	536.11	325.92	331.70	343.51	524.87
319.73	322.39	334.73	528.47	320.71	323.13	335.33	503.71
327.52	334.12	346.62	534.63	327.52	333.12	345.62	530.11
319.82	327.95	338.84	509.06	319.65	323.16	332.92	497.52
321.91	325.64	339.49	529.46	321.91	325.64	336.76	496.03
320.02	325.21	337.70	512.29	320.02	325.21	337.70	518.02
323.84	330.45	341.78	508.38	323.84	329.32	341.61	512.57
323.18	326.80	339.34	524.33	323.18	326.77	339.05	501.49
321.48	328.22	340.84	509.81	321.48	327.22	338.85	509.74
333.68	339.58	352.16	529.02	333.68	339.58	352.16	538.23
334.53	340.93	353.42	538.45	334.53	340.78	353.27	540.12
315.30	321.57	332.92	496.86	315.30	322.56	333.67	507.81
317.56	323.54	336.00	515.77	317.56	323.44	335.90	522.75

xentGT1	xentGT10	xentGT28	xentGT300	xentSAM1	xentSAM10	xentSAM28	xentSAM300
313.60	320.29	331.20	489.16	313.60	316.84	327.81	498.19
322.58	328.42	340.81	528.22	322.58	327.82	340.22	523.41
316.95	323.00	335.36	520.43	316.02	318.76	330.83	514.28
320.45	327.38	340.11	533.40	320.11	323.76	336.52	514.15
330.17	335.00	347.56	532.48	330.17	335.85	348.41	532.34
319.65	325.77	337.88	517.02	319.65	325.67	337.78	522.07
304.83	314.15	328.39	529.00	304.83	308.20	319.29	482.18
318.76	324.98	337.43	520.39	318.76	324.09	336.54	517.90
319.65	325.72	338.14	523.06	319.65	325.72	338.19	524.72
306.04	309.85	321.68	501.05	306.04	308.89	320.66	483.78
309.73	313.46	326.25	495.64	310.52	313.79	326.59	496.57
322.75	328.87	341.26	524.26	322.75	327.97	340.36	523.72
319.58	326.70	338.20	525.59	319.58	322.49	333.99	497.78
335.85	341.90	354.50	534.50	335.85	341.90	354.50	540.89
305.73	312.06	326.15	518.49	305.73	309.21	321.44	502.41
289.34	297.23	308.67	483.52	289.34	292.53	304.00	467.96
303.99	309.65	321.69	492.83	303.99	309.65	321.69	492.51
326.95	333.25	345.76	528.00	326.95	333.25	345.76	531.70
315.91	318.78	328.58	523.28	315.91	318.78	328.58	492.30
324.60	330.06	342.57	530.37	324.60	330.06	342.57	525.47
315.64	320.68	334.01	514.73	314.64	319.62	332.84	511.96
324.63	329.53	342.04	529.31	324.63	329.52	342.03	522.74
321.76	327.33	339.59	518.93	321.76	326.93	339.19	519.03
313.20	318.40	330.88	514.81	313.20	318.40	330.89	512.43
326.33	331.60	344.18	529.81	326.33	331.60	344.18	529.92
329.39	334.44	346.89	529.15	329.39	334.44	346.89	527.47
320.78	327.73	339.50	521.76	320.78	325.53	337.27	511.97
321.22	326.14	339.40	533.97	321.22	325.86	338.56	513.86
330.35	336.37	349.27	549.72	330.35	335.86	347.96	529.92
311.48	317.26	329.26	523.40	311.48	318.02	329.98	515.38
322.53	328.02	340.62	519.33	322.53	326.04	338.65	510.05
309.79	315.92	328.65	515.66	309.79	315.92	327.87	515.39

xentGT1	xentGT10	xentGT28	xentGT300	xentSAM1	xentSAM10	xentSAM28	xentSAM300
318.49	325.12	337.55	526.38	318.49	325.12	335.84	505.90
321.00	326.59	338.87	514.61	321.00	323.91	335.79	506.30
307.44	309.83	320.51	502.38	307.44	309.83	320.39	488.59
295.89	300.12	311.88	501.51	295.89	299.14	310.91	493.68
312.81	317.48	330.05	522.75	312.81	315.69	328.26	485.25
315.31	320.71	333.02	527.09	315.31	321.22	333.55	515.61
311.58	317.25	331.61	507.46	311.58	315.53	326.51	486.47
320.99	325.78	338.38	520.88	320.99	325.76	338.29	521.01
313.35	318.86	330.61	502.43	313.35	316.42	328.15	499.69

3. Cross entropy table (GRU 128 unit, 1 layer)

Following are the values of the cross entropy of ground truth for 1, 10, 28, 300 step (xentGT1, xentGT10, xentGT28, xentGT300 respectively) and sample for 1, 10, 28, 300 steps (xentSAM1, xentSAM10, xentSAM28, xentSAM300 respectively)

Table 6: Cross entropy table (GRU 128 unit, 1 layer)

xentGT1	xentGT10	xentGT28	xentGT300	xentSAM1	xentSAM10	xentSAM28	xentSAM300
326.03	332.73	345.28	529.44	326.03	331.73	344.28	526.26
318.31	321.72	334.65	530.47	318.31	321.72	333.67	501.93
323.52	328.34	341.03	510.76	322.52	327.35	340.04	515.12
302.53	306.01	319.61	516.97	302.53	305.73	317.82	502.24
329.66	335.62	348.18	519.63	329.66	335.62	348.18	532.50
323.94	328.58	340.81	528.31	323.94	326.74	337.88	502.63
307.69	313.32	325.47	511.85	307.69	314.26	326.43	513.39
329.85	332.78	345.15	521.41	329.85	332.78	345.15	528.61
325.81	331.39	343.94	527.10	325.81	331.15	343.66	520.39
309.37	315.40	325.91	484.44	309.37	316.12	326.54	494.24
313.18	320.02	330.63	496.25	313.18	320.02	330.52	500.48
325.52	329.11	343.42	545.47	325.52	329.11	340.26	508.94
317.75	324.46	337.09	521.17	317.75	324.46	336.09	519.15
314.02	320.32	331.82	524.01	314.02	320.32	331.55	514.65
314.64	320.97	332.80	509.31	314.64	320.97	332.49	501.43
310.54	315.70	327.85	502.89	310.54	315.70	327.85	502.68
306.24	313.34	324.58	492.90	306.24	312.34	323.35	490.41
312.38	318.79	330.66	524.62	312.38	318.79	331.12	518.62
316.78	322.97	335.27	512.22	316.78	321.97	334.27	512.20
320.96	326.90	339.16	542.39	320.96	326.90	339.14	521.63
308.71	312.95	327.09	507.78	308.71	312.81	324.16	489.49
324.66	330.17	342.67	530.83	324.66	330.15	342.64	523.13
317.01	319.80	329.94	521.60	317.01	319.80	329.94	504.47
330.03	336.71	349.18	534.41	330.03	335.71	348.18	533.63
310.01	317.17	329.61	517.52	310.01	314.31	326.76	500.75
308.16	313.88	325.38	510.48	308.16	313.91	325.21	493.92
329.77	335.67	348.19	529.06	329.77	335.27	347.79	531.17

xentGT1	xentGT10	xentGT28	xentGT300	xentSAM1	xentSAM10	xentSAM28	xentSAM300
307.78	310.79	322.14	517.74	307.78	310.79	321.74	482.18
319.06	324.44	337.21	514.81	319.06	322.63	335.11	506.61
322.63	325.36	336.18	525.25	322.63	326.05	336.86	502.34
322.33	327.68	339.99	524.35	322.33	328.64	340.95	528.38
307.75	313.48	325.35	509.52	308.69	313.76	326.51	499.35
321.67	328.29	340.33	520.22	321.67	327.29	339.14	511.94
320.87	327.38	340.05	507.34	320.87	327.21	338.89	510.39
318.10	325.75	337.55	517.05	318.10	324.75	336.55	513.38
318.66	325.58	337.90	526.04	318.66	324.48	336.81	516.63
326.90	332.43	344.99	531.23	326.90	332.43	344.99	528.12
316.45	325.98	337.16	503.80	316.45	319.69	329.99	504.75
332.76	338.09	350.61	536.34	332.76	337.09	349.61	529.69
320.10	326.24	339.09	530.76	320.10	326.27	339.12	525.52
322.61	326.17	338.45	516.66	322.61	326.17	338.45	523.08
318.08	324.14	335.81	516.72	318.08	324.14	335.81	515.35
309.40	314.79	327.01	516.07	309.40	315.63	327.85	505.87
320.50	327.04	339.09	503.21	320.50	327.04	338.10	508.67
320.71	329.60	341.88	535.47	320.71	323.82	336.10	512.39
317.91	322.53	336.29	529.92	317.91	322.54	334.64	503.58
326.52	332.79	344.72	520.67	326.52	332.79	343.78	518.21
320.60	323.63	335.94	517.80	321.60	324.49	336.73	510.82
328.17	334.82	347.44	533.24	328.17	333.82	346.44	529.59
319.93	325.99	337.17	499.83	319.89	324.64	335.45	509.21
322.53	326.17	340.09	523.96	322.53	326.17	337.07	501.14
320.87	325.79	338.33	511.97	320.87	325.79	338.33	518.69
324.35	330.69	342.08	503.60	324.35	329.65	342.04	515.78
322.16	325.80	338.56	518.10	322.16	325.78	338.44	502.50
321.68	328.27	341.05	510.99	321.68	327.38	339.16	512.49
334.10	340.29	352.85	527.13	334.10	340.29	352.85	539.55
335.75	342.42	354.95	533.43	335.75	341.42	353.95	539.74
314.80	320.92	332.18	495.36	314.80	321.52	332.66	505.61
318.66	325.11	337.51	516.72	318.66	324.11	336.51	520.57

xentGT1	xentGT10	xentGT28	xentGT300	xentSAM1	xentSAM10	xentSAM28	xentSAM300
313.85	320.40	331.43	489.29	313.85	316.93	327.96	500.73
321.43	327.11	339.63	526.97	321.43	327.11	339.63	521.95
316.57	322.42	334.19	514.53	315.65	319.77	331.59	516.77
320.37	327.41	339.97	526.06	320.00	324.12	336.71	519.87
330.83	335.58	348.02	530.16	330.83	336.54	348.96	533.15
317.99	324.32	336.48	516.45	317.99	324.32	336.48	520.48
305.40	313.63	327.68	521.98	305.40	309.17	320.05	496.88
318.76	324.63	337.11	518.05	318.76	323.63	336.12	516.71
318.57	325.01	337.42	520.34	318.57	324.02	336.53	514.86
307.02	310.86	323.05	489.41	307.02	309.99	321.28	492.57
310.29	313.89	326.94	498.95	311.27	314.52	327.60	495.97
321.54	327.94	340.34	518.12	321.54	327.34	339.74	520.92
319.17	326.40	337.90	526.37	319.17	322.10	333.62	496.57
336.58	342.71	355.24	533.08	336.58	342.71	355.24	541.92
305.90	311.32	323.90	508.70	305.90	311.16	323.73	503.28
291.38	299.02	310.35	482.89	291.38	294.38	305.71	468.02
303.86	309.63	321.65	497.11	303.86	309.63	321.91	495.25
328.13	334.19	346.66	526.29	328.13	334.19	346.66	530.00
315.16	319.49	329.47	530.44	315.16	320.00	329.98	495.97
326.05	331.42	343.98	529.58	326.05	331.42	343.98	526.77
316.00	321.15	334.44	519.70	315.00	319.97	332.26	507.71
324.78	329.95	342.47	518.54	324.78	329.95	342.47	523.44
321.05	326.43	338.75	511.72	321.05	326.43	338.75	520.76
313.08	318.20	330.73	519.08	313.08	317.90	330.43	503.31
325.18	332.28	346.12	542.43	325.18	329.65	341.01	509.04
330.33	335.65	348.18	526.09	330.33	335.65	348.18	529.48
322.13	328.60	340.05	523.02	322.13	328.53	339.98	520.00
321.58	326.89	339.98	536.37	321.58	325.40	338.05	508.52
328.49	334.27	346.78	546.60	328.49	334.16	346.66	531.01
311.63	317.21	329.08	524.95	311.63	318.13	329.98	510.48
322.90	328.04	340.54	513.48	322.90	328.04	340.54	518.70
308.81	314.80	327.49	515.26	308.81	314.80	327.01	512.40

xentGT1	xentGT10	xentGT28	xentGT300	xentSAM1	xentSAM10	xentSAM28	xentSAM300
316.84	323.92	336.28	530.88	316.84	323.92	334.76	506.03
322.21	328.25	340.30	506.83	322.21	326.46	338.68	511.46
307.75	310.41	321.11	498.54	307.75	310.41	320.95	492.67
298.16	302.08	313.86	498.55	298.16	301.14	312.92	493.85
313.72	318.65	331.33	514.83	313.72	316.81	329.49	497.53
315.26	320.39	332.84	525.30	315.26	320.19	332.64	511.54
311.58	318.22	332.82	504.51	311.58	315.32	325.43	485.21
321.29	326.40	338.87	523.31	321.29	325.48	337.89	515.69
314.38	318.91	330.60	498.03	314.38	318.70	330.36	502.02

4. Cross entropy table (GRU 32 unit, 3 layer)

Following are the values of the cross entropy of ground truth for 1, 10, 28, 300 step (xentGT1, xentGT10, xentGT28, xentGT300 respectively) and sample for 1, 10, 28, 300 steps (xentSAM1, xentSAM10, xentSAM28, xentSAM300 respectively)

Table 7: Cross entropy table (GRU 32 unit, 3 layer)

xentGT1	xentGT10	xentGT28	xentGT300	xentSAM1	xentSAM10	xentSAM28	xentSAM300
327.87	334.09	346.53	528.25	327.87	333.20	345.64	531.02
317.42	320.45	333.13	533.94	317.42	320.45	332.24	501.23
322.38	328.70	340.91	510.51	322.38	326.03	338.17	511.04
300.37	304.24	318.70	512.52	300.37	303.45	315.24	499.04
330.09	335.86	348.43	521.50	330.09	335.86	348.43	532.95
323.17	328.37	340.12	537.97	323.17	326.43	337.98	498.61
309.46	315.32	327.39	509.85	309.46	315.47	327.52	511.29
331.40	333.90	346.44	530.94	331.40	333.90	346.44	524.39
325.76	330.46	343.02	523.52	325.76	330.41	342.99	525.18
308.66	314.95	325.41	494.37	308.66	315.00	325.74	494.64
313.26	320.59	330.73	495.16	313.25	320.49	330.51	500.96
324.78	328.33	341.90	551.75	324.78	328.33	339.84	503.83
319.70	326.29	338.75	523.97	319.70	326.29	337.76	524.68
316.27	322.02	333.60	515.17	316.27	321.72	332.99	511.03
316.34	322.56	334.31	505.79	316.34	322.36	334.03	504.04
312.55	315.97	328.08	505.63	312.55	315.97	328.06	503.18
305.88	312.86	324.48	491.65	305.88	311.96	323.52	496.03
313.30	319.71	331.73	524.46	313.30	319.71	331.81	519.05
316.61	322.52	334.84	512.96	316.61	321.52	333.85	514.66
323.19	328.42	340.62	537.91	323.19	328.62	340.79	527.93
308.99	312.78	325.24	500.03	308.99	312.76	325.06	496.01
323.79	329.58	342.25	530.00	323.79	329.56	342.23	526.11
315.63	318.54	328.97	530.94	315.63	318.54	328.95	494.47
328.56	335.27	347.58	534.68	328.56	334.47	346.78	532.29
309.18	316.35	328.68	520.30	309.18	313.99	326.31	498.81
307.52	313.17	324.60	518.41	307.52	313.04	324.34	493.86
330.65	336.48	349.06	531.14	330.65	336.28	348.87	532.16

xentGT1	xentGT10	xentGT28	xentGT300	xentSAM1	xentSAM10	xentSAM28	xentSAM300
308.44	310.83	322.42	518.12	308.44	310.83	322.23	480.67
321.88	327.25	339.72	517.89	321.88	325.80	338.19	505.79
322.49	325.45	336.23	531.16	322.49	326.00	336.81	499.14
321.91	327.23	339.30	521.72	321.91	327.51	339.59	526.08
309.42	315.20	326.63	506.17	310.39	315.04	327.00	501.90
322.04	328.73	340.71	527.83	322.04	327.73	339.71	514.43
321.19	327.12	339.45	505.25	321.19	326.94	338.90	509.94
319.13	326.61	338.48	519.75	319.13	325.51	337.39	521.75
320.02	327.04	339.39	525.29	320.02	325.05	337.41	519.09
326.50	332.47	344.83	531.30	326.50	332.47	344.83	527.83
314.80	323.49	335.48	519.99	314.80	320.06	330.67	502.17
332.52	337.97	350.43	535.58	332.52	337.08	349.54	529.73
321.87	328.28	340.79	528.12	321.87	328.31	340.82	530.10
323.20	327.08	339.56	519.45	323.20	327.08	339.56	525.00
318.00	324.22	335.76	509.64	318.00	324.12	335.24	505.39
310.67	315.73	328.06	512.22	310.67	315.92	328.25	509.90
319.94	326.32	337.91	507.36	319.94	326.32	337.31	510.14
322.31	331.32	343.35	537.78	322.31	325.43	337.50	515.70
317.68	322.11	336.59	531.27	317.68	322.59	334.82	503.93
327.45	332.61	344.68	524.55	327.45	332.73	344.79	524.14
319.85	322.97	335.12	528.88	320.79	323.69	335.70	504.57
327.75	334.50	347.10	531.62	327.75	333.51	346.10	529.96
320.47	327.13	339.11	510.06	320.41	323.25	334.24	499.72
323.10	326.54	339.84	531.21	323.10	326.54	338.01	496.71
320.90	325.80	338.49	525.66	320.90	325.79	338.34	512.53
323.76	330.21	341.81	510.48	323.76	328.87	341.44	513.07
323.24	326.88	339.67	519.32	323.24	326.85	339.12	506.70
320.39	328.58	341.28	510.06	320.39	325.20	336.90	506.71
334.28	340.52	353.11	529.57	334.28	340.52	353.11	539.37
335.54	342.02	354.53	536.79	335.54	341.17	353.67	539.40
314.88	320.80	332.43	498.44	314.88	321.60	332.72	505.22
317.67	324.38	336.96	516.05	317.67	323.48	336.06	521.13

xentGT1	xentGT10	xentGT28	xentGT300	xentSAM1	xentSAM10	xentSAM28	xentSAM300
312.85	319.80	330.97	489.57	312.85	316.21	327.38	497.15
322.45	328.38	340.72	530.65	322.45	328.38	340.72	526.96
316.65	322.29	334.16	521.59	315.74	318.03	329.87	514.02
320.01	326.78	339.45	538.93	319.40	323.12	335.81	509.41
331.50	336.06	348.54	533.37	331.50	336.91	349.41	533.60
319.70	325.59	337.62	516.01	319.70	325.29	337.32	521.18
304.80	314.18	328.41	529.18	304.80	308.34	319.38	479.89
320.08	326.32	338.93	521.91	320.08	325.33	337.94	520.15
320.30	326.71	338.85	524.08	320.30	326.61	338.80	525.18
306.03	309.73	321.26	505.58	306.03	308.87	320.32	480.86
310.33	313.52	326.24	498.66	310.97	313.93	326.69	493.15
322.62	329.14	341.46	526.38	322.62	328.24	340.56	525.02
320.34	327.61	339.05	520.56	320.34	323.52	335.08	502.72
336.77	343.25	355.61	535.86	336.77	343.25	355.61	542.33
305.36	311.72	325.86	523.34	305.36	310.14	321.89	493.49
289.56	297.35	308.92	484.22	289.56	292.60	304.24	467.79
304.14	309.88	321.89	501.72	304.14	309.86	322.42	496.88
327.75	333.91	346.44	528.78	327.75	333.91	346.44	531.10
314.94	318.54	327.97	525.92	314.94	318.58	328.11	490.95
325.26	330.17	342.86	529.05	325.26	330.17	342.86	525.18
315.31	320.28	333.72	517.85	314.32	319.20	331.64	510.58
324.96	329.82	342.34	525.06	324.96	329.81	342.32	521.36
323.05	328.08	340.19	514.31	323.05	328.21	340.32	523.18
313.51	318.59	331.09	516.06	313.51	318.59	331.10	512.90
326.21	333.70	348.13	551.96	326.21	330.49	341.28	497.99
330.57	335.79	348.34	528.65	330.57	335.79	348.32	529.03
321.25	327.47	339.15	522.63	321.25	325.67	337.36	515.08
320.86	326.01	339.06	534.97	320.86	326.03	338.54	508.81
330.79	336.38	348.76	544.15	330.79	336.24	348.62	534.83
312.13	317.66	329.61	523.09	312.13	318.41	330.28	516.15
322.63	327.65	340.06	521.52	322.63	325.93	338.39	513.10
310.45	316.56	329.65	516.19	310.45	316.56	328.97	516.73

xentGT1	xentGT10	xentGT28	xentGT300	xentSAM1	xentSAM10	xentSAM28	xentSAM300
318.16	325.23	337.51	525.18	318.16	325.23	336.30	509.21
321.37	327.72	339.35	520.52	321.37	325.10	337.00	504.22
306.38	308.91	319.88	508.33	306.38	308.91	319.65	484.14
296.88	301.07	312.81	517.28	296.88	300.10	311.84	485.12
313.62	317.53	330.49	520.07	313.62	315.62	328.59	487.14
314.92	320.38	332.63	524.26	314.92	319.99	332.23	514.53
310.93	316.86	329.10	508.27	310.93	316.02	328.12	498.46
320.42	325.21	337.70	522.05	320.42	324.59	336.99	517.89
313.42	319.53	331.13	504.74	313.42	317.12	328.67	499.84

References

[1] C. Olah, “Colah blog,” 2015.

[2] O. Vinyals, “Sequences and recurrent neural nets,” February 2017.

[3] K. Simonyan, “Image recognition with convolutional networks,” February 2017.