Capstone Project

Neural translation model

Instructions

In this notebook, you will create a neural network that translates from English to German. You will use concepts from throughout this course, including building more flexible model architectures, freezing layers, data processing pipeline and sequence modelling.

This project is peer-assessed. Within this notebook you will find instructions in each section for how to complete the project. Pay close attention to the instructions as the peer review will be carried out according to a grading rubric that checks key parts of the project instructions. Feel free to add extra cells into the notebook as required.

How to submit

When you have completed the Capstone project notebook, you will submit a pdf of the notebook for peer review. First ensure that the notebook has been fully executed from beginning to end, and all of the cell outputs are visible. This is important, as the grading rubric depends on the reviewer being able to view the outputs of your notebook. Save the notebook as a pdf (you could download the notebook with File -> Download .ipynb, open the notebook locally, and then File -> Download as -> PDF via LaTeX), and then submit this pdf for review.

Let's get started!

We'll start by running some imports, and loading the dataset. For this project you are free to make further imports throughout the notebook as you wish.

In [2]:

```
import tensorflow as tf
import tensorflow_hub as hub
import unicodedata
import re
from IPython.display import Image
```

For the capstone project, you will use a language dataset from http://www.manythings.org/anki/ to build a neural translation model. This dataset consists of over 200,000 pairs of sentences in English and German. In order to make the training quicker, we will restrict to our dataset to 20,000 pairs. Feel free to change this if you wish - the size of the dataset used is not part of the grading rubric.

Your goal is to develop a neural translation model from English to German, making use of a pre-trained English word embedding module.

Import the data

The dataset is available for download as a zip file at the following link:

https://drive.google.com/open?id=1KczOciG7sYY7SB9UlBeRP1T9659b121Q

You should store the unzipped folder in Drive for use in this Colab notebook.

In [3]:

```
# Run this cell to connect to your Drive folder

# from google.colab import drive
# drive.mount('/content/gdrive')

from google.colab import drive
drive.mount('/content/drive')

import os
path = '/content/drive/My Drive/INSAID/TensorFlow/Customizing your models with Tensorflow 2/Week5'
os.chdir(path)
```

```
os.listdir()
```

```
Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/dri
ve", force_remount=True).
```

Out[3]:

```
['deu.txt',
 'english-german.pkl',
 'encoder.h5',
 'decoder.h5',
 'Capstone_Project.ipynb',
 'nmt_with_attention.ipynb',
 'machine_translation_tensorflow_orig.ipynb',
 'neural_translation_model.png']
```

In [4]:

```python
# Run this cell to load the dataset

NUM_EXAMPLES = 20000
data_examples = []
with open(os.path.join(path,'deu.txt'), 'r', encoding='utf8') as f:
    for line in f.readlines():
        if len(data_examples) < NUM_EXAMPLES:
            data_examples.append(line)
        else:
            break
```

In [5]:

```python
# These functions preprocess English and German sentences

def unicode_to_ascii(s):
    return ''.join(c for c in unicodedata.normalize('NFD', s) if unicodedata.category(c) != 'Mn')

def preprocess_sentence(sentence):
    sentence = sentence.lower().strip()
    sentence = re.sub(r"ü", 'ue', sentence)
    sentence = re.sub(r"ä", 'ae', sentence)
    sentence = re.sub(r"ö", 'oe', sentence)
    sentence = re.sub(r'ß', 'ss', sentence)

    sentence = unicode_to_ascii(sentence)
    sentence = re.sub(r"([?.!,])", r" \1 ", sentence)
    sentence = re.sub(r"[^a-z?.!,']+", " ", sentence)
    sentence = re.sub(r'[" "]+', " ", sentence)

    return sentence.strip()
```

The custom translation model

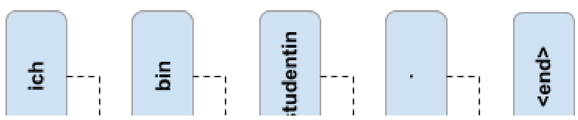
The following is a schematic of the custom translation model architecture you will develop in this project.

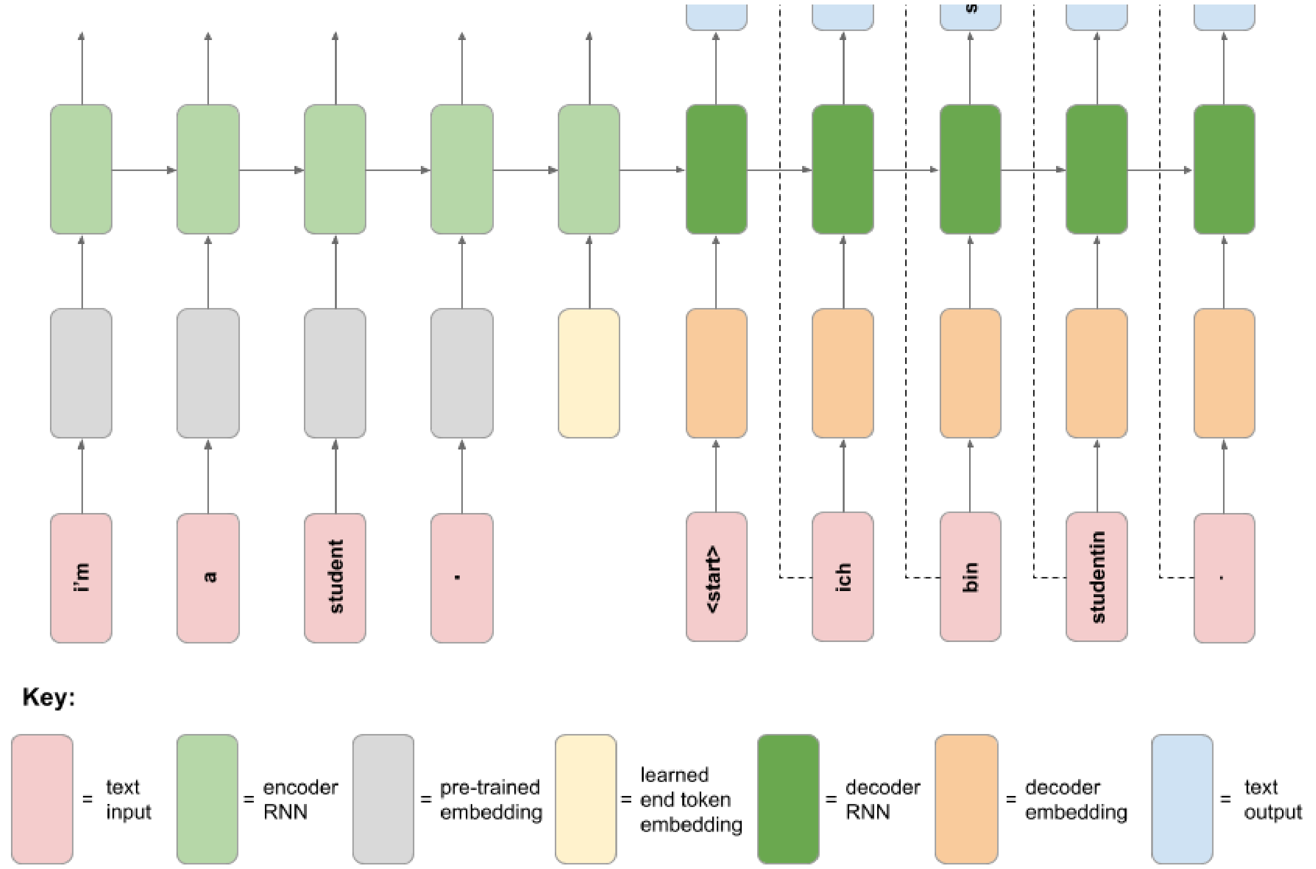
In [6]:

```python
# Run this cell to download and view a schematic diagram for the neural translation model

!wget -q -O neural_translation_model.png --no-check-certificate "https://docs.google.com/uc?export=down
load&id=1XsS1VlXoaEo-RbYNilJ9jcscNZvsSPmd"
Image("neural_translation_model.png")
```

Out[6]:





The custom model consists of an encoder RNN and a decoder RNN. The encoder takes words of an English sentence as input, and uses a pre-trained word embedding to embed the words into a 128-dimensional space. To indicate the end of the input sentence, a special end token (in the same 128-dimensional space) is passed in as an input. This token is a TensorFlow Variable that is learned in the training phase (unlike the pre-trained word embedding, which is frozen).

The decoder RNN takes the internal state of the encoder network as its initial state. A start token is passed in as the first input, which is embedded using a learned German word embedding. The decoder RNN then makes a prediction for the next German word, which during inference is then passed in as the following input, and this process is repeated until the special `<end>` token is emitted from the decoder.

1. Text preprocessing

- Create separate lists of English and German sentences, and preprocess them using the `preprocess_sentence` function provided for you above.
- Add a special `"<start>"` and `"<end>"` token to the beginning and end of every German sentence.
- Use the Tokenizer class from the `tf.keras.preprocessing.text` module to tokenize the German sentences, ensuring that no character filters are applied. *Hint: use the Tokenizer's "filter" keyword argument.*
- Print out at least 5 randomly chosen examples of (preprocessed) English and German sentence pairs. For the German sentence, print out the text (with start and end tokens) as well as the tokenized sequence.
- Pad the end of the tokenized German sequences with zeros, and batch the complete set of sequences into one numpy array.

In [7]:

```
# Inspect the first few data examples
data_examples[:5]
```

Out[7]:

```
['Hi.\tHallo!\tCC-BY 2.0 (France) Attribution: tatoeba.org #538123 (CM) & #380701 (cburgmer)\n',
 'Hi.\tGrüß Gott!\tCC-BY 2.0 (France) Attribution: tatoeba.org #538123 (CM) & #659813 (Esperantostern)\
n',
 'Run!\tLauf!\tCC-BY 2.0 (France) Attribution: tatoeba.org #906328 (papabear) & #941078 (Fingerhut)\n',
 'Wow!\tPotzdonner!\tCC-BY 2.0 (France) Attribution: tatoeba.org #52027 (Zifre) & #2122382 (Pfirsichbae
umchen)\n',
 'Wow!\tDonnerwetter!\tCC-BY 2.0 (France) Attribution: tatoeba.org #52027 (Zifre) & #2122391 (Pfirsichb
aeumchen)\n']
```

In [8]:

```
# Extract the English and the German sentences
english = [l.split("\t")[0] for l in data_examples]
german = [l.split("\t")[1] for l in data_examples]
```

In [9]:

```
# Preprocess the language datasets
import numpy as np

english = np.array([preprocess_sentence(s) for s in english])
german = np.array([preprocess_sentence(s) for s in german])
```

In [10]:

```
# Add start and end tokens
german = np.array(["<start> {} <end>".format(s) for s in german])
```

In [11]:

```
# Tokenize the German sequences
from tensorflow.keras.preprocessing.text import Tokenizer

german_tokenizer = Tokenizer(filters='')
german_tokenizer.fit_on_texts(german)
```

In [12]:

```
# Convert the German sentences to token sequences
tokenized_german_sequences = german_tokenizer.texts_to_sequences(german)
```

In [13]:

```
# Inspect a few examples of each
inx = np.random.choice(len(english), 5, replace=False)

for eng, ger in zip(english[inx], german[inx]):
    print(eng)
    print(ger)
    print(german_tokenizer.texts_to_sequences(ger.split()))
    print('')
```

```
i am out of work .
<start> ich bin arbeitslos . <end>
[[1], [4], [15], [2051], [3], [2]]

make it stop .
<start> mach , dass es aufhoert ! <end>
[[1], [79], [25], [132], [10], [2482], [9], [2]]

we miss you , too .
<start> wir vermissen dich auch . <end>
[[1], [17], [473], [28], [112], [3], [2]]

she helps us .
<start> sie hilft uns . <end>
[[1], [8], [545], [31], [3], [2]]

use your fist .
<start> nutze deine faust ! <end>
[[1], [1930], [117], [2557], [9], [2]]
```

In [14]:

```
# Pad the tokenized sequences with zeros
from tensorflow.keras.preprocessing.sequence import pad_sequences

padded_german_seq = pad_sequences(tokenized_german_sequences, padding="post")
```

In [15]:

```
# Get the number of tokens
num_german_tokens = max(german_tokenizer.index_word.keys()) + 1
print(num_german_tokens)
```

```
5744
```

2. Prepare the data

Load the embedding layer

As part of the dataset preproceessing for this project, you will use a pre-trained English word embedding module from TensorFlow Hub. The URL for the module is https://tfhub.dev/google/tf2-preview/nnlm-en-dim128-with-normalization/1.

This embedding takes a batch of text tokens in a 1-D tensor of strings as input. It then embeds the separate tokens into a 128-dimensional space.

The code to load and test the embedding layer is provided for you below.

NB: this model can also be used as a sentence embedding module. The module will process each token by removing punctuation and splitting on spaces. It then averages the word embeddings over a sentence to give a single embedding vector. However, we will use it only as a word embedding module, and will pass each word in the input sentence as a separate token.

In [16]:

```
# Load embedding module from Tensorflow Hub
embedding_layer = hub.KerasLayer("https://tfhub.dev/google/tf2-preview/nnlm-en-dim128/1",
                                 output_shape=[128], input_shape=[], dtype=tf.string)
```

In [17]:

```
# Test the layer
embedding_layer(tf.constant(["these", "aren't", "the", "droids", "you're", "looking", "for"])).shape
```

Out[17]:

```
TensorShape([7, 128])
```

You should now prepare the training and validation Datasets.

- Create a random training and validation set split of the data, reserving e.g. 20% of the data for validation (NB: each English dataset example is a single sentence string, and each German dataset example is a sequence of padded integer tokens).
- Load the training and validation sets into a tf.data.Dataset object, passing in a tuple of English and German data for both training and validation sets.
- Create a function to map over the datasets that splits each English sentence at spaces. Apply this function to both Dataset objects using the map method. *Hint: look at the tf.strings.split function.*
- Create a function to map over the datasets that embeds each sequence of English words using the loaded embedding layer/model. Apply this function to both Dataset objects using the map method.
- Create a function to filter out dataset examples where the English sentence is greater than or equal to than 13 (embedded) tokens in length. Apply this function to both Dataset objects using the filter method.
- Create a function to map over the datasets that pads each English sequence of embeddings with some distinct padding value before the sequence, so that each sequence is length 13. Apply this function to both Dataset objects using the map method. *Hint: look at the tf.pad function. You can extract a Tensor shape using tf.shape; you might also find the tf.math.maximum function useful.*

function useful.

- Batch both training and validation Datasets with a batch size of 16.
- Print the `element_spec` property for the training and validation Datasets.
- Using the Dataset `.take(1)` method, print the shape of the English data example from the training Dataset.
- Using the Dataset `.take(1)` method, print the German data example Tensor from the validation Dataset.

In [18]:

```
# Create training, test and validation splits
from random import shuffle
from sklearn.model_selection import train_test_split

english_train, english_valid, german_train, german_valid = train_test_split(english,
                                                                            padded_german_seq,
                                                                            test_size=0.2)
```

In [19]:

```
# Create a translation Dataset
eng_ger_train = tf.data.Dataset.from_tensor_slices((english_train, german_train))
eng_ger_valid = tf.data.Dataset.from_tensor_slices((english_valid, german_valid))
eng_ger_train.element_spec
```

Out[19]:

```
(TensorSpec(shape=(), dtype=tf.string, name=None),
 TensorSpec(shape=(14,), dtype=tf.int32, name=None))
```

In [20]:

```
# Split the English sentences at spaces
def split_at_spaces(eng, ger):
    return tf.strings.split(eng), ger

eng_ger_train = eng_ger_train.map(split_at_spaces)
eng_ger_valid = eng_ger_valid.map(split_at_spaces)
eng_ger_train.element_spec
```

Out[20]:

```
(TensorSpec(shape=(None,), dtype=tf.string, name=None),
 TensorSpec(shape=(14,), dtype=tf.int32, name=None))
```

In [21]:

```
# Get the embeddings for the English sentences
def get_embedding(eng, ger):
    return embedding_layer(eng), ger

eng_ger_train = eng_ger_train.map(get_embedding)
eng_ger_valid = eng_ger_valid.map(get_embedding)
eng_ger_train.element_spec
```

Out[21]:

```
(TensorSpec(shape=(None, 128), dtype=tf.float32, name=None),
 TensorSpec(shape=(14,), dtype=tf.int32, name=None))
```

In [22]:

```
# Filter out examples where the English sentence is longer than 13 tokens
max_english_sentence_len = 13

def filter_long_sentences(eng, ger):
    return tf.shape(eng)[0] <= max_english_sentence_len

eng_ger_train = eng_ger_train.filter(filter_long_sentences)
```

```
eng_ger_train = eng_ger_train.filter(filter_long_sentences)
eng_ger_valid = eng_ger_valid.filter(filter_long_sentences)
eng_ger_train.element_spec
```

Out[22]:

```
(TensorSpec(shape=(None, 128), dtype=tf.float32, name=None),
 TensorSpec(shape=(14,), dtype=tf.int32, name=None))
```

In [23]:

```
# Pad the English sentences
def pad_embedding_sequences(eng, ger):
    length_emb = tf.shape(eng)[0]
    padding = [[tf.math.maximum(0, max_english_sentence_len - length_emb), 0], [0, 0]]
    return tf.pad(eng, padding, 'CONSTANT', constant_values=-1000), ger

eng_ger_train = eng_ger_train.map(pad_embedding_sequences)
eng_ger_valid = eng_ger_valid.map(pad_embedding_sequences)
eng_ger_train.element_spec
```

Out[23]:

```
(TensorSpec(shape=(None, 128), dtype=tf.float32, name=None),
 TensorSpec(shape=(14,), dtype=tf.int32, name=None))
```

In [24]:

```
# Batch the Datasets
eng_ger_train = eng_ger_train.batch(16)
eng_ger_valid = eng_ger_valid.batch(16)
```

In [25]:

```
# Print the final training Dataset element_spec
eng_ger_train.element_spec
```

Out[25]:

```
(TensorSpec(shape=(None, None, 128), dtype=tf.float32, name=None),
 TensorSpec(shape=(None, 14), dtype=tf.int32, name=None))
```

In [26]:

```
# Print the shape of a batch of English data examples
for eng, ger in eng_ger_train.take(1):
    print(eng.shape)
```

```
(16, 13, 128)
```

In [27]:

```
# Print a batch of German data examples
for eng, ger in eng_ger_valid.take(1):
    print(ger)
```

```
tf.Tensor(
[[   1  763   17  381    7    2    0    0    0    0    0    0    0    0]
 [   1    5 4038   22   12    3    2    0    0    0    0    0    0    0]
 [   1    4 3434    3    2    0    0    0    0    0    0    0    0    0]
 [   1   10    6  625  401    3    2    0    0    0    0    0    0    0]
 [   1   11  152    6  611    3    2    0    0    0    0    0    0    0]
 [   1    4  918   10  141    3    2    0    0    0    0    0    0    0]
 [   1  406   53 1165    9    2    0    0    0    0    0    0    0    0]
```

```
 [   1  202 2485 1329   34    3    2    0    0    0    0    0    0    0]
 [   1    4   18   27   37 1368 4835    3    2    0    0    0    0    0]
 [   1    5  165 1735   20   54    3    2    0    0    0    0    0    0]
 [   1    5   16 2694    3    2    0    0    0    0    0    0    0    0]
 [   1    4   39   10  124  220    3    2    0    0    0    0    0    0]
 [   1    4   61  119    3    2    0    0    0    0    0    0    0    0]
 [   1    4   18   40 1593    3    2    0    0    0    0    0    0    0]
 [   1   14 5568  125    3    2    0    0    0    0    0    0    0    0]
 [   1   13  246  342   92    3    2    0    0    0    0    0    0    0]], shape=(16, 14), dtype=int32)
```

3. Create the custom layer

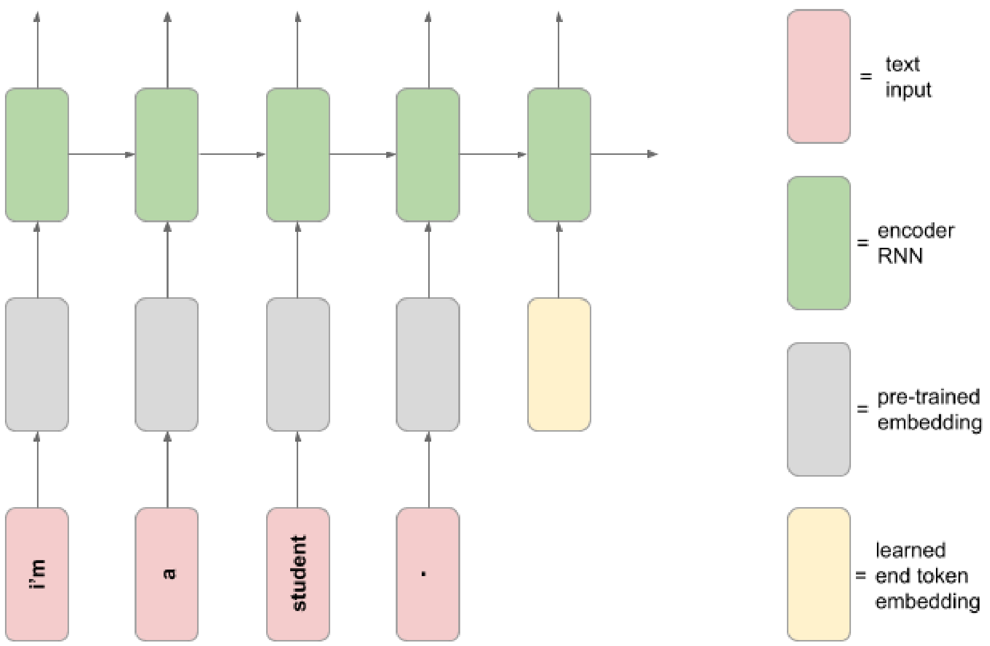
You will now create a custom layer to add the learned end token embedding to the encoder model:

In [28]:

```
# Run this cell to download and view a schematic diagram for the encoder model

!wget -q -O neural_translation_model.png --no-check-certificate "https://docs.google.com/uc?export=down
load&id=1JrtNOzUJDaOWrK4C-xv-4wUuZaI12sQI"
Image("neural_translation_model.png")
```

Out[28]:



You should now build the custom layer.

- Using layer subclassing, create a custom layer that takes a batch of English data examples from one of the Datasets, and adds a learned embedded 'end' token to the end of each sequence.
- This layer should create a TensorFlow Variable (that will be learned during training) that is 128-dimensional (the size of the embedding space). *Hint: you may find it helpful in the call method to use the tf.tile function to replicate the end token embedding across every element in the batch.*
- Using the Dataset `.take(1)` method, extract a batch of English data examples from the training Dataset and print the shape. Test the custom layer by calling the layer on the English data batch Tensor and print the resulting Tensor shape (the layer should increase the sequence length by one).

In [29]:

```
# Create the custom layer to add an end token embedding
class AddEndToken(tf.keras.layers.Layer):
    def __init__(self, **kwargs):
        super(AddEndToken, self).__init__(**kwargs)

    def build(self, input_shape):
        embedding_size = input_shape[-1]
```

```
        self.end_token = self.add_weight(shape=(1, embedding_size),
                                         initializer='random_normal',
                                         name='english_end_token')

    def call(self, inputs):
        batch_size = tf.shape(inputs)[0]
        token = tf.tile(tf.expand_dims(self.end_token, 0), [batch_size, 1, 1])
        return tf.concat([inputs, token], 1)
```

In [30]:

```
# Test the layer
add_end_token = AddEndToken()

for eng, ger in eng_ger_train.take(1):
    print(eng.shape)
    print(add_end_token(eng).shape)
```

```
(16, 13, 128)
(16, 14, 128)
```

In [30]:

```

```

4. Build the encoder network

The encoder network follows the schematic diagram above. You should now build the RNN encoder model.

- Using the functional API, build the encoder network according to the following spec:
  - The model will take a batch of sequences of embedded English words as input, as given by the Dataset objects.
  - The next layer in the encoder will be the custom layer you created previously, to add a learned end token embedding to the end of the English sequence.
  - This is followed by a Masking layer, with the `mask_value` set to the distinct padding value you used when you padded the English sequences with the Dataset preprocessing above.
  - The final layer is an LSTM layer with 512 units, which also returns the hidden and cell states.
  - The encoder is a multi-output model. There should be two output Tensors of this model: the hidden state and cell states of the LSTM layer. The output of the LSTM layer is unused.
- Using the Dataset `.take(1)` method, extract a batch of English data examples from the training Dataset and test the encoder model by calling it on the English data Tensor, and print the shape of the resulting Tensor outputs.
- Print the model summary for the encoder network.

In [31]:

```
# Build the encoder
from tensorflow.keras.models import Model
from tensorflow.keras.layers import Input, LSTM, Masking

def get_encoder():
    encoder_input = Input(shape=(None, 128))
    h = AddEndToken()(encoder_input)
    h = Masking(mask_value=-1000)(h)
    enc_output, enc_hidden, enc_cell = LSTM(512, return_state=True)(h)

    return Model(inputs=encoder_input, outputs=[enc_hidden, enc_cell])

encoder = get_encoder()
```

In [32]:

```
# Test the encoder
for eng, ger in eng_ger_train.take(1):
    print([t.shape for t in encoder(eng)])
```

```
[TensorShape([16, 512]), TensorShape([16, 512])]
```

In [33]:

```
# Print the encoder summary
encoder.summary()
```

```
Model: "functional_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_1 (InputLayer)         [(None, None, 128)]       0
_________________________________________________________________
add_end_token_1 (AddEndToken (None, None, 128)         128
_________________________________________________________________
masking (Masking)            (None, None, 128)         0
_________________________________________________________________
lstm (LSTM)                  [(None, 512), (None, 512) 1312768
=================================================================
Total params: 1,312,896
Trainable params: 1,312,896
Non-trainable params: 0
_________________________________________________________________
```

5. Build the decoder network

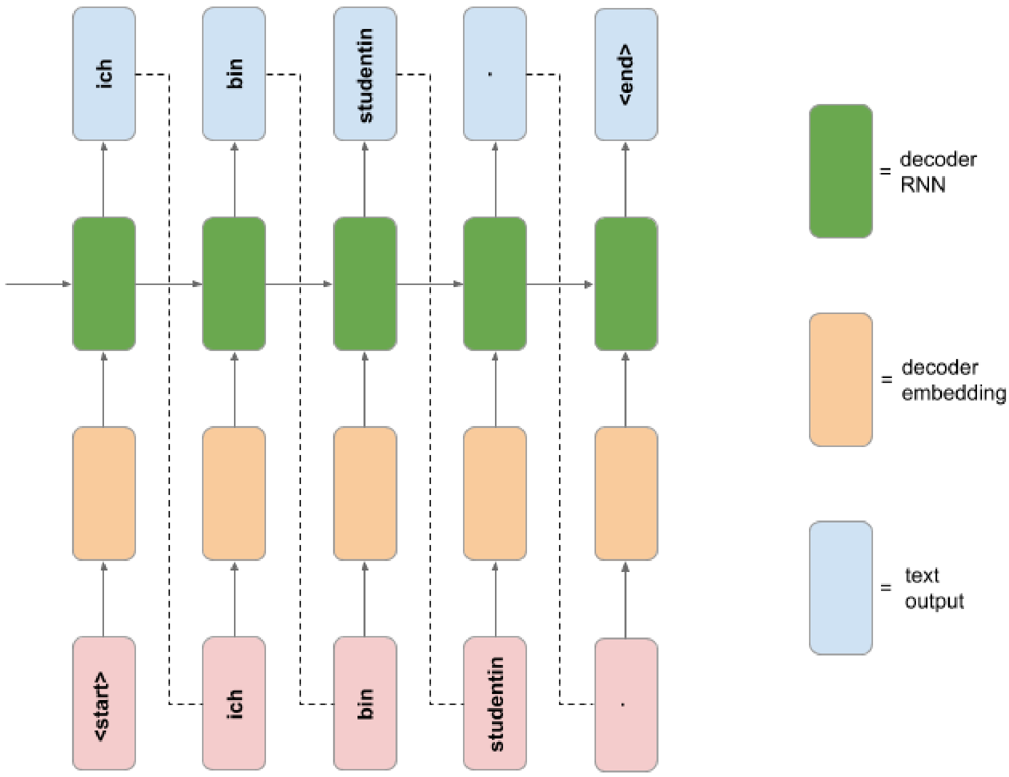
The decoder network follows the schematic diagram below.

In [34]:

```
# Run this cell to download and view a schematic diagram for the decoder model

!wget -q -O neural_translation_model.png --no-check-certificate "https://docs.google.com/uc?export=down
load&id=1DTeaXD8tA8RjkpVrB2mr9csSBOY4LQiW"
Image("neural_translation_model.png")
```

Out[34]:



You should now build the RNN decoder model.

- Using Model subclassing, build the decoder network according to the following spec:
  - The initializer should create the following layers:
    - An Embedding layer with vocabulary size set to the number of unique German tokens, embedding dimension 128, and set to mask zero values in the input.
    - An LSTM layer with 512 units, that returns its hidden and cell states, and also returns sequences.
    - A Dense layer with number of units equal to the number of unique German tokens, and no activation function.
  - The call method should include the usual `inputs` argument, as well as the additional keyword arguments `hidden_state` and `cell_state`. The default value for these keyword arguments should be `None`.
  - The call method should pass the inputs through the Embedding layer, and then through the LSTM layer. If the `hidden_state` and `cell_state` arguments are provided, these should be used for the initial state of the LSTM layer. *Hint: use the* `initial_state` *keyword argument when calling the LSTM layer on its input.*
  - The call method should pass the LSTM output sequence through the Dense layer, and return the resulting Tensor, along with the hidden and cell states of the LSTM layer.
- Using the Dataset `.take(1)` method, extract a batch of English and German data examples from the training Dataset. Test the decoder model by first calling the encoder model on the English data Tensor to get the hidden and cell states, and then call the decoder model on the German data Tensor and hidden and cell states, and print the shape of the resulting decoder Tensor outputs.
- Print the model summary for the decoder network.

In [35]:

```
# Build the decoder
from tensorflow.keras.layers import Dense, Embedding

class Decoder(Model):
    def __init__(self, **kwargs):
        super(Decoder, self).__init__(**kwargs)
        self.embedding = Embedding(num_german_tokens, 128, mask_zero=True)
        self.lstm = LSTM(512, return_state=True, return_sequences=True)
        self.dense = Dense(num_german_tokens)

    def call(self, inputs, hidden_state=None, cell_state=None, training=True):
        if hidden_state is None:
            assert cell_state is None
            h = self.embedding(inputs)
            o, h, c = self.lstm(h)
        else:
            h = self.embedding(inputs)
            o, h, c = self.lstm(h, initial_state=[hidden_state, cell_state])
        return self.dense(o), h, c

decoder = Decoder()
```

In [36]:

```
# Test the decoder
for eng, ger in eng_ger_train.take(1):
    hidden, cell = encoder(eng)
    print([t.shape for t in decoder(ger, hidden_state=hidden, cell_state=cell)])
```

```
[TensorShape([16, 14, 5744]), TensorShape([16, 512]), TensorShape([16, 512])]
```

In [37]:

```
# Print the decoder summary
decoder.summary()
```

```
Model: "decoder"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
embedding (Embedding)        multiple                  735232
_________________________________________________________________
lstm_1 (LSTM)                multiple                  1312768
```

```
dense (Dense)                multiple                  2946672
=================================================================
Total params: 4,994,672
Trainable params: 4,994,672
Non-trainable params: 0
_________________________________________________________________
```

6. Make a custom training loop

You should now write a custom training loop to train your custom neural translation model.

- Define a function that takes a Tensor batch of German data (as extracted from the training Dataset), and returns a tuple containing German inputs and outputs for the decoder model (refer to schematic diagram above).
- Define a function that computes the forward and backward pass for your translation model. This function should take an English input, German input and German output as arguments, and should do the following:
  - Pass the English input into the encoder, to get the hidden and cell states of the encoder LSTM.
  - These hidden and cell states are then passed into the decoder, along with the German inputs, which returns a sequence of outputs (the hidden and cell state outputs of the decoder LSTM are unused in this function).
  - The loss should then be computed between the decoder outputs and the German output function argument.
  - The function returns the loss and gradients with respect to the encoder and decoder's trainable variables.
  - Decorate the function with `@tf.function`
- Define and run a custom training loop for a number of epochs (for you to choose) that does the following:
  - Iterates through the training dataset, and creates decoder inputs and outputs from the German sequences.
  - Updates the parameters of the translation model using the gradients of the function above and an optimizer object.
  - Every epoch, compute the validation loss on a number of batches from the validation and save the epoch training and validation losses.
- Plot the learning curves for loss vs epoch for both training and validation sets.

Hint: This model is computationally demanding to train. The quality of the model or length of training is not a factor in the grading rubric. However, to obtain a better model we recommend using the GPU accelerator hardware on Colab.

In [38]:

```
# Define the loss function and optimizer
loss_fn = tf.keras.losses.SparseCategoricalCrossentropy(from_logits=True)
optimizer = tf.keras.optimizers.Adam()
```

In [39]:

```
# Function to create German tokens inputs and outputs
def get_german_inputs_outputs(ger_in):
    return ger_in[:, :-1], ger_in[:, 1:]
```

In [40]:

```
# Define the forward and backward pass
@tf.function
def get_loss_and_grads(eng_in, ger_in, ger_out):
    with tf.GradientTape() as tape:
        h, c = encoder(eng_in)
        outputs, _1, _2 = decoder(ger_in, hidden_state=h, cell_state=c)
        current_loss = loss_fn(ger_out, outputs)
        grads = tape.gradient(current_loss,
                              encoder.trainable_variables + decoder.trainable_variables)
    return current_loss, grads
```

In [41]:

```
# For evaluation
@tf.function
def get_loss(eng_in, ger_in, ger_out):
    h, c = encoder(eng_in)
    outputs, _1, _2 = decoder(ger_in, hidden_state=h, cell_state=c)
    return loss_fn(ger_out, outputs)
```

In [42]:

```
# Run the custom training loop
num_epochs = 8
num_valid_steps = 100

epoch_history = {
    'epoch': [],
    'loss': [],
    'val_loss': []
}
batch_history = {
    'iteration': [],
    'loss': []
}

for epoch in range(num_epochs):
    epoch_loss = tf.keras.metrics.Mean()

    for iteration, (eng, ger) in enumerate(eng_ger_train):
        ger_inputs, ger_outputs = get_german_inputs_outputs(ger)
        loss, grads = get_loss_and_grads(eng, ger_inputs, ger_outputs)
        epoch_loss(loss)
        print("Iteration {}, loss: {}".format(iteration, loss))
        if iteration % 10 == 0:
            batch_history['iteration'].append(iteration)
            batch_history['loss'].append(loss)
        optimizer.apply_gradients(zip(grads, encoder.trainable_variables + decoder.trainable_variables)
)
    print("End of epoch {}, loss: {}".format(epoch, epoch_loss.result() + 1))

    # Validation
    for (eng, ger) in eng_ger_valid.take(num_valid_steps):
        valid_loss = tf.keras.metrics.Mean()
        ger_inputs, ger_outputs = get_german_inputs_outputs(ger)
        valid_loss(get_loss(eng, ger_inputs, ger_outputs))
    print("Validation loss: {}".format(valid_loss.result()))

    epoch_history['epoch'].append(epoch)
    epoch_history['loss'].append(epoch_loss.result())
    epoch_history['val_loss'].append(valid_loss.result())
```

```
Iteration 146, loss: 2.8058855533599854
Iteration 147, loss: 2.7952470779418945
Iteration 148, loss: 2.9884378910064697
Iteration 149, loss: 2.950084686279297
Iteration 150, loss: 2.741946220397949
Iteration 151, loss: 2.887845754623413
Iteration 152, loss: 2.7155299186706543
Iteration 153, loss: 2.885842800140381
Iteration 154, loss: 2.812913656234741
Iteration 155, loss: 2.878556251525879
Iteration 156, loss: 2.9246485233306885
Iteration 157, loss: 2.775843620300293
Iteration 158, loss: 2.751404047012329
Iteration 159, loss: 2.5388360023498535
Iteration 160, loss: 2.8655099868774414
Iteration 161, loss: 2.964594841003418
Iteration 162, loss: 2.7003533840179443
Iteration 163, loss: 2.8629777431488037
Iteration 164, loss: 2.6988489627838135
Iteration 165, loss: 2.719658851623535
Iteration 166, loss: 2.761998414993286
Iteration 167, loss: 2.8764612674713135
Iteration 168, loss: 2.8896565437316895
Iteration 169, loss: 2.9137344360351562
Iteration 170, loss: 2.7605552673339844
Iteration 171, loss: 3.0624351501464844
Iteration 172, loss: 2.89264178276062
Iteration 173, loss: 2.909329652786255
Iteration 174, loss: 2.7756612300872803
Iteration 175, loss: 2.8397061824798584
Iteration 176, loss: 2.738278865814209
Iteration 177, loss: 2.8613011837005615
```

```
Iteration 178, loss: 2.8184030055999756
Iteration 179, loss: 2.764652729034424
Iteration 180, loss: 2.7360410690307617
Iteration 181, loss: 2.9279191493988037
Iteration 182, loss: 2.662417411804199
Iteration 183, loss: 2.952319860458374
Iteration 184, loss: 2.823753833770752
Iteration 185, loss: 2.8803470134735107
Iteration 186, loss: 2.7411394119262695
Iteration 187, loss: 2.968440532684326
Iteration 188, loss: 2.770155668258667
Iteration 189, loss: 2.831960439682007
Iteration 190, loss: 2.7174527645111084
Iteration 191, loss: 2.7775731086730957
Iteration 192, loss: 2.779301643371582
Iteration 193, loss: 2.9516406059265137
Iteration 194, loss: 2.685591697692871
Iteration 195, loss: 2.7380475997924805
Iteration 196, loss: 2.8906266689300537
Iteration 197, loss: 2.6989173889160156
Iteration 198, loss: 2.774268865585327
Iteration 199, loss: 2.5844743251800537
Iteration 200, loss: 2.8782551288604736
Iteration 201, loss: 2.8438570499420166
Iteration 202, loss: 2.8534159660339355
Iteration 203, loss: 2.6807470321655273
Iteration 204, loss: 2.7651727199554443
Iteration 205, loss: 2.9870288372039795
Iteration 206, loss: 2.721566677093506
Iteration 207, loss: 2.7276723384857178
Iteration 208, loss: 2.7635881900787354
Iteration 209, loss: 2.700345516204834
Iteration 210, loss: 2.8262813091278076
Iteration 211, loss: 2.8448362350463867
Iteration 212, loss: 2.8278651237487793
Iteration 213, loss: 2.7858762741088867
Iteration 214, loss: 2.769043445587158
Iteration 215, loss: 2.6045918464660645
Iteration 216, loss: 2.796858310699463
Iteration 217, loss: 2.8690969944000244
Iteration 218, loss: 2.7387683391571045
Iteration 219, loss: 2.903047800064087
Iteration 220, loss: 2.6427457332611084
Iteration 221, loss: 2.7259762287139893
Iteration 222, loss: 2.761227607727051
Iteration 223, loss: 2.768224000930786
Iteration 224, loss: 2.93489670753479
Iteration 225, loss: 2.963864803314209
Iteration 226, loss: 2.773711681365967
Iteration 227, loss: 2.933746576309204
Iteration 228, loss: 2.777250051498413
Iteration 229, loss: 2.8832955360412598
Iteration 230, loss: 2.678309440612793
Iteration 231, loss: 2.6090481281280518
Iteration 232, loss: 2.6914098262786865
Iteration 233, loss: 2.7928483486175537
Iteration 234, loss: 2.694908380508423
Iteration 235, loss: 2.6976118087768555
Iteration 236, loss: 2.702162027359009
Iteration 237, loss: 2.6881256103515625
Iteration 238, loss: 2.7879605293273926
Iteration 239, loss: 2.7370049953460693
Iteration 240, loss: 2.6644811630249023
Iteration 241, loss: 2.7737526893615723
Iteration 242, loss: 2.905168056488037
Iteration 243, loss: 2.71268367767334
Iteration 244, loss: 2.6856648921966553
Iteration 245, loss: 2.7538795471191406
Iteration 246, loss: 2.7061305046081543
Iteration 247, loss: 2.7244157791137695
Iteration 248, loss: 2.772280216217041
Iteration 249, loss: 2.5532479286193848
Iteration 250, loss: 2.7997145652770996
Iteration 251, loss: 2.684570550918579
Iteration 252, loss: 2.8336586952209473
Iteration 253, loss: 2.760101318359375
Iteration 254, loss: 2.7382092475891113
```

```
Iteration 255, loss: 2.678809642791748
Iteration 256, loss: 2.880034923553467
Iteration 257, loss: 2.7961180210113525
Iteration 258, loss: 2.662667751312256
Iteration 259, loss: 2.692384958267212
Iteration 260, loss: 2.685483455657959
Iteration 261, loss: 2.652174711227417
Iteration 262, loss: 2.734065055847168
Iteration 263, loss: 2.82511568069458
Iteration 264, loss: 2.7608489990234375
Iteration 265, loss: 2.8441872596740723
Iteration 266, loss: 2.9298136234283447
Iteration 267, loss: 2.6864585876464844
Iteration 268, loss: 2.779224157333374
Iteration 269, loss: 2.772735834121704
Iteration 270, loss: 2.572263479232788
Iteration 271, loss: 2.538558006286621
Iteration 272, loss: 2.795499563217163
Iteration 273, loss: 2.7873737812042236
Iteration 274, loss: 2.6034278869628906
Iteration 275, loss: 2.7432286739349365
Iteration 276, loss: 2.8046514987945557
Iteration 277, loss: 2.628849506378174
Iteration 278, loss: 2.598240852355957
Iteration 279, loss: 2.8291919231414795
Iteration 280, loss: 2.748882532119751
Iteration 281, loss: 2.695112705230713
Iteration 282, loss: 2.7628228664398193
Iteration 283, loss: 2.611643075942993
Iteration 284, loss: 2.7390408515930176
Iteration 285, loss: 2.7244884967803955
Iteration 286, loss: 2.944769859313965
Iteration 287, loss: 2.6985244750976562
Iteration 288, loss: 2.7365756034851074
Iteration 289, loss: 2.72965407371521
Iteration 290, loss: 2.575350046157837
Iteration 291, loss: 2.8057198524475098
Iteration 292, loss: 2.8132131099700928
Iteration 293, loss: 2.6902287006378174
Iteration 294, loss: 2.7132720947265625
Iteration 295, loss: 2.755549907684326
Iteration 296, loss: 2.735095500946045
Iteration 297, loss: 2.7840802669525146
Iteration 298, loss: 2.711946725845337
Iteration 299, loss: 2.5361416339874268
Iteration 300, loss: 2.6535654067993164
Iteration 301, loss: 2.661391019821167
Iteration 302, loss: 2.7047665119171143
Iteration 303, loss: 2.775317668914795
Iteration 304, loss: 2.6780574321746826
Iteration 305, loss: 2.535886287689209
Iteration 306, loss: 2.7856976985931396
Iteration 307, loss: 2.753690242767334
Iteration 308, loss: 2.67734694480896
Iteration 309, loss: 2.7393798828125
Iteration 310, loss: 2.7018213272094727
Iteration 311, loss: 2.7650911808013916
Iteration 312, loss: 2.648151159286499
Iteration 313, loss: 2.7134921550750732
Iteration 314, loss: 2.658704996109009
Iteration 315, loss: 2.697162628173828
Iteration 316, loss: 2.7884674072265625
Iteration 317, loss: 2.6886188983917236
Iteration 318, loss: 2.6746606826782227
Iteration 319, loss: 2.642329216003418
Iteration 320, loss: 2.6374807357788086
Iteration 321, loss: 2.5793967247009277
Iteration 322, loss: 2.7304470539093018
Iteration 323, loss: 2.7726428508758545
Iteration 324, loss: 2.9354238510131836
Iteration 325, loss: 2.6793224811553955
Iteration 326, loss: 2.707329273223877
Iteration 327, loss: 2.57969331741333
Iteration 328, loss: 2.753962278366089
Iteration 329, loss: 2.739173412322998
Iteration 330, loss: 2.728210687637329
Iteration 331, loss: 2.823061466217041
```

```
Iteration 332, loss: 2.4302220344543457
Iteration 333, loss: 2.661329984664917
Iteration 334, loss: 2.697338581085205
Iteration 335, loss: 2.6281442642211914
Iteration 336, loss: 2.796719789505005
Iteration 337, loss: 2.7102715969085693
Iteration 338, loss: 2.6990277767181396
Iteration 339, loss: 2.7730302810668945
Iteration 340, loss: 2.648810386657715
Iteration 341, loss: 2.744887351989746
Iteration 342, loss: 2.69644832611084
Iteration 343, loss: 2.6831018924713135
Iteration 344, loss: 2.6258089542388916
Iteration 345, loss: 2.7810277938842773
Iteration 346, loss: 2.641984462738037
Iteration 347, loss: 2.7387824058532715
Iteration 348, loss: 2.6554160118103027
Iteration 349, loss: 2.6406326293945312
Iteration 350, loss: 2.607513189315796
Iteration 351, loss: 2.730416774749756
Iteration 352, loss: 2.781599998474121
Iteration 353, loss: 2.682258129119873
Iteration 354, loss: 2.609081745147705
Iteration 355, loss: 2.5741631984710693
Iteration 356, loss: 2.8114442825317383
Iteration 357, loss: 2.7202723026275635
Iteration 358, loss: 2.575849771499634
Iteration 359, loss: 2.833207368850708
Iteration 360, loss: 2.750507116317749
Iteration 361, loss: 2.8019919395446777
Iteration 362, loss: 2.7111549377441406
Iteration 363, loss: 2.470707893371582
Iteration 364, loss: 2.6656904220581055
Iteration 365, loss: 2.801891326904297
Iteration 366, loss: 2.754974842071533
Iteration 367, loss: 2.743360757827759
Iteration 368, loss: 2.6225881576538086
Iteration 369, loss: 2.682474374771118
Iteration 370, loss: 2.5953118801116943
Iteration 371, loss: 2.6051089763641357
Iteration 372, loss: 2.5573856830596924
Iteration 373, loss: 2.5289931297302246
Iteration 374, loss: 2.6869924068450928
Iteration 375, loss: 2.6915626525878906
Iteration 376, loss: 2.5939881801605225
Iteration 377, loss: 2.626878261566162
Iteration 378, loss: 2.735008716583252
Iteration 379, loss: 2.7114129066467285
Iteration 380, loss: 2.7479493618011475
Iteration 381, loss: 2.6997451782226562
Iteration 382, loss: 2.5799543857574463
Iteration 383, loss: 2.6812498569488525
Iteration 384, loss: 2.6823365688323975
Iteration 385, loss: 2.727149486541748
Iteration 386, loss: 2.7752177715301514
Iteration 387, loss: 2.6985673904418945
Iteration 388, loss: 2.487088203430176
Iteration 389, loss: 2.6853203773498535
Iteration 390, loss: 2.558368682861328
Iteration 391, loss: 2.571427583694458
Iteration 392, loss: 2.634524345397949
Iteration 393, loss: 2.6853930950164795
Iteration 394, loss: 2.569978952407837
Iteration 395, loss: 2.74835205078125
Iteration 396, loss: 2.644124746322632
Iteration 397, loss: 2.5632286071777344
Iteration 398, loss: 2.6745734214782715
Iteration 399, loss: 2.651475429534912
Iteration 400, loss: 2.6888482570648193
Iteration 401, loss: 2.6574666500091553
Iteration 402, loss: 2.59905743598938
Iteration 403, loss: 2.7750298976898193
Iteration 404, loss: 2.7343294620513916
Iteration 405, loss: 2.7356696128845215
Iteration 406, loss: 2.5158069133758545
Iteration 407, loss: 2.584183692932129
Iteration 408, loss: 2.7411766052246094
```

```
Iteration 409, loss: 2.7197186946868896
Iteration 410, loss: 2.6401352882385254
Iteration 411, loss: 2.6552963256835938
Iteration 412, loss: 2.691652774810791
Iteration 413, loss: 2.5924603939056396
Iteration 414, loss: 2.5644145011901855
Iteration 415, loss: 2.5776939392089844
Iteration 416, loss: 2.668868064880371
Iteration 417, loss: 2.6655566692352295
Iteration 418, loss: 2.6532042026519775
Iteration 419, loss: 2.5972442626953125
Iteration 420, loss: 2.6295714378356934
Iteration 421, loss: 2.926191568374634
Iteration 422, loss: 2.4943933486938477
Iteration 423, loss: 2.4985101222991943
Iteration 424, loss: 2.77984619140625
Iteration 425, loss: 2.5693321228027344
Iteration 426, loss: 2.5253806114196777
Iteration 427, loss: 2.634136438369751
Iteration 428, loss: 2.5515928268432617
Iteration 429, loss: 2.5933122634887695
Iteration 430, loss: 2.583305835723877
Iteration 431, loss: 2.7466278076171875
Iteration 432, loss: 2.492119073867798
Iteration 433, loss: 2.6028361320495605
Iteration 434, loss: 2.546602964401245
Iteration 435, loss: 2.6060409545898438
Iteration 436, loss: 2.5813558101654053
Iteration 437, loss: 2.5947325229644775
Iteration 438, loss: 2.542424201965332
Iteration 439, loss: 2.679499626159668
Iteration 440, loss: 2.563939332962036
Iteration 441, loss: 2.527366876602173
Iteration 442, loss: 2.5607950687408447
Iteration 443, loss: 2.4798219203948975
Iteration 444, loss: 2.6128296852111816
Iteration 445, loss: 2.5772900581359863
Iteration 446, loss: 2.571507453918457
Iteration 447, loss: 2.5394225120544434
Iteration 448, loss: 2.7177975177764893
Iteration 449, loss: 2.4545276165008545
Iteration 450, loss: 2.552741050720215
Iteration 451, loss: 2.491490364074707
Iteration 452, loss: 2.4580013751983643
Iteration 453, loss: 2.4446702003479004
Iteration 454, loss: 2.598604679107666
Iteration 455, loss: 2.576826333999634
Iteration 456, loss: 2.5727155208587646
Iteration 457, loss: 2.6221556663513184
Iteration 458, loss: 2.578303337097168
Iteration 459, loss: 2.5340731143951416
Iteration 460, loss: 2.6854360103607178
Iteration 461, loss: 2.6336288452148438
Iteration 462, loss: 2.517927885055542
Iteration 463, loss: 2.54555082321167
Iteration 464, loss: 2.5371389389038086
Iteration 465, loss: 2.7088024616241455
Iteration 466, loss: 2.5894055366516113
Iteration 467, loss: 2.5617077350616455
Iteration 468, loss: 2.7493066787719727
Iteration 469, loss: 2.3555405139923096
Iteration 470, loss: 2.6002261638641357
Iteration 471, loss: 2.587358236312866
Iteration 472, loss: 2.572296142578125
Iteration 473, loss: 2.5675787925720215
Iteration 474, loss: 2.500217914581299
Iteration 475, loss: 2.5297343730926514
Iteration 476, loss: 2.68776273727417
Iteration 477, loss: 2.612861156463623
Iteration 478, loss: 2.4791150093078613
Iteration 479, loss: 2.6329057216644287
Iteration 480, loss: 2.7112956047058105
Iteration 481, loss: 2.675089120864868
Iteration 482, loss: 2.5147054195404053
Iteration 483, loss: 2.577122926712036
Iteration 484, loss: 2.646400213241577
Iteration 485, loss: 2.526621103286743
```

```
Iteration 486, loss: 2.542297840118408
Iteration 487, loss: 2.511404275894165
Iteration 488, loss: 2.540576457977295
Iteration 489, loss: 2.537771463394165
Iteration 490, loss: 2.5205283164978027
Iteration 491, loss: 2.5697884559631348
Iteration 492, loss: 2.5497708320617676
Iteration 493, loss: 2.577423095703125
Iteration 494, loss: 2.477466344833374
Iteration 495, loss: 2.550323247909546
Iteration 496, loss: 2.5230329036712646
Iteration 497, loss: 2.392677068710327
Iteration 498, loss: 2.592318058013916
Iteration 499, loss: 2.6535799503326416
Iteration 500, loss: 2.497476816177368
Iteration 501, loss: 2.670685052871704
Iteration 502, loss: 2.580314874649048
Iteration 503, loss: 2.571187734603882
Iteration 504, loss: 2.6002376079559326
Iteration 505, loss: 2.5957982540130615
Iteration 506, loss: 2.599477529525757
Iteration 507, loss: 2.5591135025024414
Iteration 508, loss: 2.4707422256469727
Iteration 509, loss: 2.5081980228424072
Iteration 510, loss: 2.605426549911499
Iteration 511, loss: 2.5642988681793213
Iteration 512, loss: 2.5913326740264893
Iteration 513, loss: 2.5970113277435303
Iteration 514, loss: 2.6232504844665527
Iteration 515, loss: 2.4810056686401367
Iteration 516, loss: 2.4604671001434326
Iteration 517, loss: 2.4407613277435303
Iteration 518, loss: 2.4931397438049316
Iteration 519, loss: 2.5485622882843018
Iteration 520, loss: 2.485074758529663
Iteration 521, loss: 2.6290972232818604
Iteration 522, loss: 2.4384820461273193
Iteration 523, loss: 2.559722423553467
Iteration 524, loss: 2.4816689491271973
Iteration 525, loss: 2.6331264972686768
Iteration 526, loss: 2.4288530349731445
Iteration 527, loss: 2.4886953830718994
Iteration 528, loss: 2.438131093978882
Iteration 529, loss: 2.6359708309173584
Iteration 530, loss: 2.737732172012329
Iteration 531, loss: 2.56608247756958
Iteration 532, loss: 2.5388548374176025
Iteration 533, loss: 2.598832607269287
Iteration 534, loss: 2.4635374546051025
Iteration 535, loss: 2.6304001808166504
Iteration 536, loss: 2.553523063659668
Iteration 537, loss: 2.526459217071533
Iteration 538, loss: 2.4130332469940186
Iteration 539, loss: 2.5378966331481934
Iteration 540, loss: 2.556283473968506
Iteration 541, loss: 2.5790112018585205
Iteration 542, loss: 2.5424957275390625
Iteration 543, loss: 2.583784580230713
Iteration 544, loss: 2.5371828079223633
Iteration 545, loss: 2.455618381500244
Iteration 546, loss: 2.6704866886138916
Iteration 547, loss: 2.4206743240356445
Iteration 548, loss: 2.4928066730499268
Iteration 549, loss: 2.475599765777588
Iteration 550, loss: 2.4843575954437256
Iteration 551, loss: 2.4981579780578613
Iteration 552, loss: 2.5211381912231445
Iteration 553, loss: 2.4452247619628906
Iteration 554, loss: 2.5109922885894775
Iteration 555, loss: 2.446045398712158
Iteration 556, loss: 2.5622076988220215
Iteration 557, loss: 2.5191574096679688
Iteration 558, loss: 2.5119411945343018
Iteration 559, loss: 2.7485532760620117
Iteration 560, loss: 2.4589762687683105
Iteration 561, loss: 2.34435772895813
Iteration 562, loss: 2.6070942878723145
```

```
Iteration 563, loss: 2.6939032077789307
Iteration 564, loss: 2.5815811157226562
Iteration 565, loss: 2.577340602874756
Iteration 566, loss: 2.508397102355957
Iteration 567, loss: 2.563157558441162
Iteration 568, loss: 2.422931432723999
Iteration 569, loss: 2.591735363006592
Iteration 570, loss: 2.4726362228393555
Iteration 571, loss: 2.5455307960510254
Iteration 572, loss: 2.5747389793395996
Iteration 573, loss: 2.435419797897339
Iteration 574, loss: 2.3633739948272705
Iteration 575, loss: 2.4810497760772705
Iteration 576, loss: 2.4277446269989014
Iteration 577, loss: 2.427529811859131
Iteration 578, loss: 2.475260019302368
Iteration 579, loss: 2.472126007080078
Iteration 580, loss: 2.479687213897705
Iteration 581, loss: 2.338590383529663
Iteration 582, loss: 2.378546714782715
Iteration 583, loss: 2.4547066688537598
Iteration 584, loss: 2.5110583305358887
Iteration 585, loss: 2.500746250152588
Iteration 586, loss: 2.461909055709839
Iteration 587, loss: 2.5816330909729004
Iteration 588, loss: 2.440631151199341
Iteration 589, loss: 2.5928218364715576
Iteration 590, loss: 2.4486207962036133
Iteration 591, loss: 2.525665760040283
Iteration 592, loss: 2.639780282974243
Iteration 593, loss: 2.5577733516693115
Iteration 594, loss: 2.4469828605651855
Iteration 595, loss: 2.3552327156066895
Iteration 596, loss: 2.5050337314605713
Iteration 597, loss: 2.494999885559082
Iteration 598, loss: 2.5906128883361816
Iteration 599, loss: 2.383867025375366
Iteration 600, loss: 2.471740245819092
Iteration 601, loss: 2.416203737258911
Iteration 602, loss: 2.4664673805236816
Iteration 603, loss: 2.507167339324951
Iteration 604, loss: 2.481844186782837
Iteration 605, loss: 2.382927894592285
Iteration 606, loss: 2.4857535362243652
Iteration 607, loss: 2.3792099952697754
Iteration 608, loss: 2.328310012817383
Iteration 609, loss: 2.526087522506714
Iteration 610, loss: 2.584588050842285
Iteration 611, loss: 2.4169256687164307
Iteration 612, loss: 2.435366630554199
Iteration 613, loss: 2.3860831260681152
Iteration 614, loss: 2.5117101669311523
Iteration 615, loss: 2.6495347023010254
Iteration 616, loss: 2.530392646789551
Iteration 617, loss: 2.5214943885803223
Iteration 618, loss: 2.404353141784668
Iteration 619, loss: 2.5219473838806152
Iteration 620, loss: 2.4179890155792236
Iteration 621, loss: 2.5360465049743652
Iteration 622, loss: 2.513360023498535
Iteration 623, loss: 2.3818578720092773
Iteration 624, loss: 2.4222683906555176
Iteration 625, loss: 2.538566827774048
Iteration 626, loss: 2.506270408630371
Iteration 627, loss: 2.6060075759887695
Iteration 628, loss: 2.42185640335083
Iteration 629, loss: 2.3757574558258057
Iteration 630, loss: 2.5229761600494385
Iteration 631, loss: 2.419801712036133
Iteration 632, loss: 2.3765199184417725
Iteration 633, loss: 2.4150826930999756
Iteration 634, loss: 2.3814761638641357
Iteration 635, loss: 2.3365094661712646
Iteration 636, loss: 2.493950128555298
Iteration 637, loss: 2.43508243560791
Iteration 638, loss: 2.4884212017059326
Iteration 639, loss: 2.4798777103424072
```

```
Iteration 640, loss: 2.462453603744507
Iteration 641, loss: 2.555490732192993
Iteration 642, loss: 2.4491915702819824
Iteration 643, loss: 2.5654497146606445
Iteration 644, loss: 2.5205931663513184
Iteration 645, loss: 2.4733030796051025
Iteration 646, loss: 2.55879807472229
Iteration 647, loss: 2.3640499114990234
Iteration 648, loss: 2.5290849208831787
Iteration 649, loss: 2.4390668869018555
Iteration 650, loss: 2.398433208465576
Iteration 651, loss: 2.536703586578369
Iteration 652, loss: 2.4193472862243652
Iteration 653, loss: 2.4177279472351074
Iteration 654, loss: 2.5249600410461426
Iteration 655, loss: 2.38769268989563
Iteration 656, loss: 2.440218925476074
Iteration 657, loss: 2.4162707328796387
Iteration 658, loss: 2.5415396690368652
Iteration 659, loss: 2.4363598823547363
Iteration 660, loss: 2.3057210445404053
Iteration 661, loss: 2.3733420372009277
Iteration 662, loss: 2.4889461994171143
Iteration 663, loss: 2.2447385787963867
Iteration 664, loss: 2.3961267471313477
Iteration 665, loss: 2.444265842437744
Iteration 666, loss: 2.4122800827026367
Iteration 667, loss: 2.378298282623291
Iteration 668, loss: 2.3694090843200684
Iteration 669, loss: 2.5461719036102295
Iteration 670, loss: 2.446324348449707
Iteration 671, loss: 2.4458985328674316
Iteration 672, loss: 2.4271292686462402
Iteration 673, loss: 2.506258726119995
Iteration 674, loss: 2.4267232418060303
Iteration 675, loss: 2.4077584743499756
Iteration 676, loss: 2.4075169563293457
Iteration 677, loss: 2.4531102180480957
Iteration 678, loss: 2.4807841777801514
Iteration 679, loss: 2.4742000102996826
Iteration 680, loss: 2.475003480911255
Iteration 681, loss: 2.381511926651001
Iteration 682, loss: 2.421236991882324
Iteration 683, loss: 2.373417377471924
Iteration 684, loss: 2.2967464923858643
Iteration 685, loss: 2.4385969638824463
Iteration 686, loss: 2.426978826522827
Iteration 687, loss: 2.4775187969207764
Iteration 688, loss: 2.436288595199585
Iteration 689, loss: 2.3779244422912598
Iteration 690, loss: 2.393120288848877
Iteration 691, loss: 2.417403221130371
Iteration 692, loss: 2.327580690383911
Iteration 693, loss: 2.4709625244140625
Iteration 694, loss: 2.407557249069214
Iteration 695, loss: 2.6003618240356445
Iteration 696, loss: 2.4755494594573975
Iteration 697, loss: 2.3914525508880615
Iteration 698, loss: 2.471292018890381
Iteration 699, loss: 2.477428436279297
Iteration 700, loss: 2.623873710632324
Iteration 701, loss: 2.4081857204437256
Iteration 702, loss: 2.4008543491363525
Iteration 703, loss: 2.492149591445923
Iteration 704, loss: 2.439507484436035
Iteration 705, loss: 2.4056191444396973
Iteration 706, loss: 2.464643716812134
Iteration 707, loss: 2.401803970336914
Iteration 708, loss: 2.4034650325775146
Iteration 709, loss: 2.411057710647583
Iteration 710, loss: 2.403853178024292
Iteration 711, loss: 2.4910314083099365
Iteration 712, loss: 2.329739570617676
Iteration 713, loss: 2.4085865020751953
Iteration 714, loss: 2.408437967300415
Iteration 715, loss: 2.3718199729919434
Iteration 716, loss: 2.3273870944976807
```

```
Iteration 717, loss: 2.413788318634033
Iteration 718, loss: 2.359253168106079
Iteration 719, loss: 2.4749059677124023
Iteration 720, loss: 2.440687417984009
Iteration 721, loss: 2.429455041885376
Iteration 722, loss: 2.354759693145752
Iteration 723, loss: 2.2116544246673584
Iteration 724, loss: 2.5147557258605957
Iteration 725, loss: 2.28410005569458
Iteration 726, loss: 2.417725086212158
Iteration 727, loss: 2.416877269744873
Iteration 728, loss: 2.3688414096832275
Iteration 729, loss: 2.2721269130706787
Iteration 730, loss: 2.380005121231079
Iteration 731, loss: 2.398561716079712
Iteration 732, loss: 2.427338123321533
Iteration 733, loss: 2.441525459289551
Iteration 734, loss: 2.4264087677001953
Iteration 735, loss: 2.3483870029449463
Iteration 736, loss: 2.5147626399993896
Iteration 737, loss: 2.396965503692627
Iteration 738, loss: 2.4619193077087402
Iteration 739, loss: 2.3009278774261475
Iteration 740, loss: 2.3383073806762695
Iteration 741, loss: 2.388958692550659
Iteration 742, loss: 2.4986772537231445
Iteration 743, loss: 2.405043125152588
Iteration 744, loss: 2.2770020961761475
Iteration 745, loss: 2.318166971206665
Iteration 746, loss: 2.3914594650268555
Iteration 747, loss: 2.4836361408233643
Iteration 748, loss: 2.3938050270080566
Iteration 749, loss: 2.273258686065674
Iteration 750, loss: 2.3800206184387207
Iteration 751, loss: 2.3024466037750244
Iteration 752, loss: 2.4294416904449463
Iteration 753, loss: 2.4456803798675537
Iteration 754, loss: 2.322568416595459
Iteration 755, loss: 2.2756311893463135
Iteration 756, loss: 2.3724868297576904
Iteration 757, loss: 2.311678409576416
Iteration 758, loss: 2.2518112659454346
Iteration 759, loss: 2.4383885860443115
Iteration 760, loss: 2.218719244003296
Iteration 761, loss: 2.3720197677612305
Iteration 762, loss: 2.364117383956909
Iteration 763, loss: 2.437741994857788
Iteration 764, loss: 2.3979508876800537
Iteration 765, loss: 2.307569742202759
Iteration 766, loss: 2.2025418281555176
Iteration 767, loss: 2.44140625
Iteration 768, loss: 2.392359733581543
Iteration 769, loss: 2.349069118499756
Iteration 770, loss: 2.3636815547943115
Iteration 771, loss: 2.2344141006469727
Iteration 772, loss: 2.41587495803833
Iteration 773, loss: 2.282283306121826
Iteration 774, loss: 2.338346481323242
Iteration 775, loss: 2.393272638320923
Iteration 776, loss: 2.358548164367676
Iteration 777, loss: 2.473257303237915
Iteration 778, loss: 2.3705344200134277
Iteration 779, loss: 2.294370174407959
Iteration 780, loss: 2.431074380874634
Iteration 781, loss: 2.2830417156219482
Iteration 782, loss: 2.230095386505127
Iteration 783, loss: 2.261251926422119
Iteration 784, loss: 2.3642334938049316
Iteration 785, loss: 2.241039752960205
Iteration 786, loss: 2.2503621578216553
Iteration 787, loss: 2.366774082183838
Iteration 788, loss: 2.381105661392212
Iteration 789, loss: 2.446094274520874
Iteration 790, loss: 2.258274793624878
Iteration 791, loss: 2.435969114303589
Iteration 792, loss: 2.2644882202148438
Iteration 793, loss: 2.3073787689208984
```

```
Iteration 794, loss: 2.2528862953186035
Iteration 795, loss: 2.2989702224731445
Iteration 796, loss: 2.458319902420044
Iteration 797, loss: 2.275258779525757
Iteration 798, loss: 2.314300775527954
Iteration 799, loss: 2.2296628952026367
Iteration 800, loss: 2.3634634017944336
Iteration 801, loss: 2.21362042427063
Iteration 802, loss: 2.499089241027832
Iteration 803, loss: 2.113624095916748
Iteration 804, loss: 2.274649143218994
Iteration 805, loss: 2.3906261920928955
Iteration 806, loss: 2.2536234855651855
Iteration 807, loss: 2.2696382999420166
Iteration 808, loss: 2.4110970497131348
Iteration 809, loss: 2.389814853668213
Iteration 810, loss: 2.476771116256714
Iteration 811, loss: 2.3543992042541504
Iteration 812, loss: 2.42676043510437
Iteration 813, loss: 2.2530179023742676
Iteration 814, loss: 2.419952630996704
Iteration 815, loss: 2.428062915802002
Iteration 816, loss: 2.244847297668457
Iteration 817, loss: 2.3544187545776367
Iteration 818, loss: 2.2905354499816895
Iteration 819, loss: 2.2735087871551514
Iteration 820, loss: 2.3595564365386963
Iteration 821, loss: 2.52721905708313
Iteration 822, loss: 2.3701632022857666
Iteration 823, loss: 2.244309425354004
Iteration 824, loss: 2.3594095706939697
Iteration 825, loss: 2.317065477371216
Iteration 826, loss: 2.3047544956207275
Iteration 827, loss: 2.340508222579956
Iteration 828, loss: 2.32368803024292
Iteration 829, loss: 2.3214054107666016
Iteration 830, loss: 2.254434108734131
Iteration 831, loss: 2.4436228275299072
Iteration 832, loss: 2.284043073654175
Iteration 833, loss: 2.294623851776123
Iteration 834, loss: 2.294616460800171
Iteration 835, loss: 2.43818736076355
Iteration 836, loss: 2.1652956008911133
Iteration 837, loss: 2.3678929805755615
Iteration 838, loss: 2.3247358798980713
Iteration 839, loss: 2.321718692779541
Iteration 840, loss: 2.289543390274048
Iteration 841, loss: 2.430830717086792
Iteration 842, loss: 2.2566473484039307
Iteration 843, loss: 2.2349345684051514
Iteration 844, loss: 2.343703031539917
Iteration 845, loss: 2.2783069610595703
Iteration 846, loss: 2.2461585998535156
Iteration 847, loss: 2.3217129707336426
Iteration 848, loss: 2.431128978729248
Iteration 849, loss: 2.285231113433838
Iteration 850, loss: 2.219597816467285
Iteration 851, loss: 2.2898917198181152
Iteration 852, loss: 2.343888282775879
Iteration 853, loss: 2.4024555683135986
Iteration 854, loss: 2.285134792327881
Iteration 855, loss: 2.2966995239257812
Iteration 856, loss: 2.2341256141662598
Iteration 857, loss: 2.42673397064209
Iteration 858, loss: 2.363600015640259
Iteration 859, loss: 2.1875548362731934
Iteration 860, loss: 2.4533965587615967
Iteration 861, loss: 2.344062328338623
Iteration 862, loss: 2.2613165378570557
Iteration 863, loss: 2.3526663780212402
Iteration 864, loss: 2.215785503387451
Iteration 865, loss: 2.3571436405181885
Iteration 866, loss: 2.3011107444763184
Iteration 867, loss: 2.339046001434326
Iteration 868, loss: 2.2798702716827393
Iteration 869, loss: 2.264218330383301
Iteration 870, loss: 2.302513599395752
```

```
Iteration 871, loss: 2.2764170169830322
Iteration 872, loss: 2.2328379154205322
Iteration 873, loss: 2.3014116287231445
Iteration 874, loss: 2.343052625656128
Iteration 875, loss: 2.2669081687927246
Iteration 876, loss: 2.1752209663391113
Iteration 877, loss: 2.17584490776062
Iteration 878, loss: 2.4391844272613525
Iteration 879, loss: 2.320357322692871
Iteration 880, loss: 2.2319047451019287
Iteration 881, loss: 2.315286874771118
Iteration 882, loss: 2.345562219619751
Iteration 883, loss: 2.220599412918091
Iteration 884, loss: 2.3768081665039062
Iteration 885, loss: 2.2462785243988037
Iteration 886, loss: 2.3922834396362305
Iteration 887, loss: 2.217820644378662
Iteration 888, loss: 2.1748204231262207
Iteration 889, loss: 2.2884087562561035
Iteration 890, loss: 2.310228109359741
Iteration 891, loss: 2.3168890476226807
Iteration 892, loss: 2.2159924507141113
Iteration 893, loss: 2.3290302753448486
Iteration 894, loss: 2.3174333572387695
Iteration 895, loss: 2.237861156463623
Iteration 896, loss: 2.1991119384765625
Iteration 897, loss: 2.2866246700286865
Iteration 898, loss: 2.2299749851226807
Iteration 899, loss: 2.196929693222046
Iteration 900, loss: 2.278259754180908
Iteration 901, loss: 2.3309502601623535
Iteration 902, loss: 2.2048380374908447
Iteration 903, loss: 2.156127452850342
Iteration 904, loss: 2.294003963470459
Iteration 905, loss: 2.339667797088623
Iteration 906, loss: 2.2474138736724854
Iteration 907, loss: 2.286126136779785
Iteration 908, loss: 2.2803664207458496
Iteration 909, loss: 2.246030569076538
Iteration 910, loss: 2.264770984649658
Iteration 911, loss: 2.2544777393341064
Iteration 912, loss: 2.2400310039520264
Iteration 913, loss: 2.1334264278411865
Iteration 914, loss: 2.147890567779541
Iteration 915, loss: 2.288978338241577
Iteration 916, loss: 2.3371670246124268
Iteration 917, loss: 2.111406087875366
Iteration 918, loss: 2.185685634613037
Iteration 919, loss: 2.286878824234009
Iteration 920, loss: 2.2152082920074463
Iteration 921, loss: 2.2355763912200928
Iteration 922, loss: 2.1809844970703125
Iteration 923, loss: 2.227504253387451
Iteration 924, loss: 2.2476816177368164
Iteration 925, loss: 2.287135124206543
Iteration 926, loss: 2.283830165863037
Iteration 927, loss: 2.261441469192505
Iteration 928, loss: 2.182684898376465
Iteration 929, loss: 2.0933587551116943
Iteration 930, loss: 2.2704386711120605
Iteration 931, loss: 2.1917455196380615
Iteration 932, loss: 2.2507123947143555
Iteration 933, loss: 2.2388880252838135
Iteration 934, loss: 2.2662675380706787
Iteration 935, loss: 2.100637197494507
Iteration 936, loss: 2.1900835037231445
Iteration 937, loss: 2.2059762477874756
Iteration 938, loss: 2.243966817855835
Iteration 939, loss: 2.2287168502807617
Iteration 940, loss: 2.2849035263061523
Iteration 941, loss: 2.2174811363220215
Iteration 942, loss: 2.2017297744750977
Iteration 943, loss: 2.2840256690979004
Iteration 944, loss: 2.2671098709106445
Iteration 945, loss: 2.181126356124878
Iteration 946, loss: 2.2071869373321533
Iteration 947, loss: 2.204312324523926
```

```
Iteration 948, loss: 2.248568534851074
Iteration 949, loss: 2.127377986907959
Iteration 950, loss: 2.2054412364959717
Iteration 951, loss: 2.168154239654541
Iteration 952, loss: 2.2122342586517334
Iteration 953, loss: 2.241004705429077
Iteration 954, loss: 2.2205233573913574
Iteration 955, loss: 2.1682639122009277
Iteration 956, loss: 2.189659357070923
Iteration 957, loss: 2.2485733032226562
Iteration 958, loss: 2.0767571926116943
Iteration 959, loss: 2.166412353515625
Iteration 960, loss: 2.2613706588745117
Iteration 961, loss: 2.0905301570892334
Iteration 962, loss: 2.260551929473877
Iteration 963, loss: 2.3621339797973633
Iteration 964, loss: 2.2186532020568848
Iteration 965, loss: 2.242739677429199
Iteration 966, loss: 2.169273614883423
Iteration 967, loss: 2.2183382511138916
Iteration 968, loss: 2.1209404468536377
Iteration 969, loss: 2.221611976623535
Iteration 970, loss: 2.3314027786254883
Iteration 971, loss: 2.2218403816223145
Iteration 972, loss: 2.183612823486328
Iteration 973, loss: 2.230828285217285
Iteration 974, loss: 2.1381566524505615
Iteration 975, loss: 2.193277597427368
Iteration 976, loss: 2.1333541870117188
Iteration 977, loss: 2.3083322048187256
Iteration 978, loss: 2.161487579345703
Iteration 979, loss: 2.192385673522949
Iteration 980, loss: 2.0769262313842773
Iteration 981, loss: 2.155212163925171
Iteration 982, loss: 2.08565616607666
Iteration 983, loss: 2.0669825077056885
Iteration 984, loss: 2.182826042175293
Iteration 985, loss: 2.278489828109741
Iteration 986, loss: 2.2547402381896973
Iteration 987, loss: 2.254209518432617
Iteration 988, loss: 2.2173008918762207
Iteration 989, loss: 2.1785311698913574
Iteration 990, loss: 2.206315517425537
Iteration 991, loss: 2.245161294937134
Iteration 992, loss: 2.2169077396392822
Iteration 993, loss: 2.1678836345672607
Iteration 994, loss: 2.1389639377593994
Iteration 995, loss: 2.180595636367798
Iteration 996, loss: 2.12227463722229
Iteration 997, loss: 2.4014694690704346
Iteration 998, loss: 2.136626720428467
Iteration 999, loss: 2.094130277633667
End of epoch 3, loss: 3.567085027694702
Validation loss: 2.443629264831543
Iteration 0, loss: 2.254446268081665
Iteration 1, loss: 2.1170618534088135
Iteration 2, loss: 2.1007015705108643
Iteration 3, loss: 2.181905746459961
Iteration 4, loss: 2.156905174255371
Iteration 5, loss: 2.189368724822998
Iteration 6, loss: 2.224612236022949
Iteration 7, loss: 2.1189980506896973
Iteration 8, loss: 2.167840003967285
Iteration 9, loss: 2.1493146419525146
Iteration 10, loss: 2.1681530475616455
Iteration 11, loss: 2.202368974685669
Iteration 12, loss: 2.235257625579834
Iteration 13, loss: 2.1902976036071777
Iteration 14, loss: 2.2640342712402344
Iteration 15, loss: 2.219418525695801
Iteration 16, loss: 2.212517738342285
Iteration 17, loss: 2.1859631538391113
Iteration 18, loss: 2.2959864139556885
Iteration 19, loss: 2.2147140502929688
Iteration 20, loss: 2.2737345695495605
Iteration 21, loss: 2.2330756187438965
```

```
Iteration 22, loss: 2.2220215797424316
Iteration 23, loss: 2.0712525844573975
Iteration 24, loss: 2.0518171787261963
Iteration 25, loss: 2.241594076156616
Iteration 26, loss: 2.292088508605957
Iteration 27, loss: 2.2811620235443115
Iteration 28, loss: 2.1809675693511963
Iteration 29, loss: 2.1543099880218506
Iteration 30, loss: 2.224841594696045
Iteration 31, loss: 2.083775043487549
Iteration 32, loss: 2.071556329727173
Iteration 33, loss: 2.3499794006347656
Iteration 34, loss: 2.2095377445220947
Iteration 35, loss: 2.120694637298584
Iteration 36, loss: 2.1846139430999756
Iteration 37, loss: 2.085258722305298
Iteration 38, loss: 2.1588220596313477
Iteration 39, loss: 2.2094554901123047
Iteration 40, loss: 2.163623332977295
Iteration 41, loss: 2.1995606422424316
Iteration 42, loss: 2.040458917617798
Iteration 43, loss: 2.1589243412017822
Iteration 44, loss: 2.076413154602051
Iteration 45, loss: 2.3447165489196777
Iteration 46, loss: 2.2595815658569336
Iteration 47, loss: 2.15960431098938
Iteration 48, loss: 2.096033811569214
Iteration 49, loss: 2.1658053398132324
Iteration 50, loss: 2.2556090354919434
Iteration 51, loss: 2.264758586883545
Iteration 52, loss: 2.131124496459961
Iteration 53, loss: 2.175219774246216
Iteration 54, loss: 2.05540132522583
Iteration 55, loss: 2.345649480819702
Iteration 56, loss: 2.03993558883667
Iteration 57, loss: 2.1351709365844727
Iteration 58, loss: 2.2922587394714355
Iteration 59, loss: 2.0857746601104736
Iteration 60, loss: 2.0191922187805176
Iteration 61, loss: 2.229891300201416
Iteration 62, loss: 2.175981044769287
Iteration 63, loss: 2.161526679992676
Iteration 64, loss: 2.1500608921051025
Iteration 65, loss: 2.1830620765686035
Iteration 66, loss: 2.072347402572632
Iteration 67, loss: 2.181208372116089
Iteration 68, loss: 2.147399425506592
Iteration 69, loss: 2.176307201385498
Iteration 70, loss: 2.116468906402588
Iteration 71, loss: 2.188772678375244
Iteration 72, loss: 2.1006102561950684
Iteration 73, loss: 2.075373888015747
Iteration 74, loss: 2.228745698928833
Iteration 75, loss: 2.1417787075042725
Iteration 76, loss: 2.1211659908294678
Iteration 77, loss: 2.184507131576538
Iteration 78, loss: 2.264526844024658
Iteration 79, loss: 1.9952547550201416
Iteration 80, loss: 2.090491771697998
Iteration 81, loss: 2.1374876499176025
Iteration 82, loss: 1.9852485656738281
Iteration 83, loss: 2.1304845809936523
Iteration 84, loss: 2.149902105331421
Iteration 85, loss: 2.0733084678649902
Iteration 86, loss: 2.045771598815918
Iteration 87, loss: 2.074859142303467
Iteration 88, loss: 2.1331496238708496
Iteration 89, loss: 2.1361615657806396
Iteration 90, loss: 2.173232078552246
Iteration 91, loss: 2.203507661819458
Iteration 92, loss: 2.094926595687866
Iteration 93, loss: 2.1250267028808594
Iteration 94, loss: 2.10422945022583
Iteration 95, loss: 2.1507081985473633
Iteration 96, loss: 2.099017381668091
Iteration 97, loss: 2.0717194080352783
Iteration 98, loss: 2.0973660945892334
Iteration 99, loss: 2.128282070159912
```

```
Iteration 99, loss: 2.128282070159912
Iteration 100, loss: 2.1158080101013184
Iteration 101, loss: 2.132812976837158
Iteration 102, loss: 2.162557363510132
Iteration 103, loss: 2.001972198486328
Iteration 104, loss: 2.198202133178711
Iteration 105, loss: 2.1136720180511475
Iteration 106, loss: 2.0746870040893555
Iteration 107, loss: 2.029226541519165
Iteration 108, loss: 2.059389352798462
Iteration 109, loss: 2.190207004547119
Iteration 110, loss: 2.1731932163238525
Iteration 111, loss: 2.1978774070739746
Iteration 112, loss: 2.080306053161621
Iteration 113, loss: 2.1616477966308594
Iteration 114, loss: 2.052621603012085
Iteration 115, loss: 2.206796407699585
Iteration 116, loss: 2.1868748664855957
Iteration 117, loss: 2.111175298690796
Iteration 118, loss: 2.1499698162078857
Iteration 119, loss: 2.155113697052002
Iteration 120, loss: 2.053347587585449
Iteration 121, loss: 2.0289971828460693
Iteration 122, loss: 2.0346932411193848
Iteration 123, loss: 2.1065306663513184
Iteration 124, loss: 2.0788888931274414
Iteration 125, loss: 2.0650155544281006
Iteration 126, loss: 2.0917563438415527
Iteration 127, loss: 2.065673589706421
Iteration 128, loss: 2.0096263885498047
Iteration 129, loss: 2.1474897861480713
Iteration 130, loss: 2.010510206222534
Iteration 131, loss: 2.1294004917144775
Iteration 132, loss: 2.142483711242676
Iteration 133, loss: 2.0951578617095947
Iteration 134, loss: 2.0719921588897705
Iteration 135, loss: 2.0264816284179688
Iteration 136, loss: 2.0362765789031982
Iteration 137, loss: 2.074228286743164
Iteration 138, loss: 2.0536439418792725
Iteration 139, loss: 2.091203451156616
Iteration 140, loss: 1.9885934591293335
Iteration 141, loss: 2.030536413192749
Iteration 142, loss: 2.132439136505127
Iteration 143, loss: 2.09335994720459
Iteration 144, loss: 2.036078691482544
Iteration 145, loss: 2.0969738960266113
Iteration 146, loss: 2.0279226303100586
Iteration 147, loss: 2.0420174598693848
Iteration 148, loss: 2.158421277999878
Iteration 149, loss: 2.1520683765411377
Iteration 150, loss: 1.9779789447784424
Iteration 151, loss: 2.0861053466796875
Iteration 152, loss: 1.9809024333953857
Iteration 153, loss: 2.103965997695923
Iteration 154, loss: 2.0781919956207275
Iteration 155, loss: 2.131605625152588
Iteration 156, loss: 2.1335740089416504
Iteration 157, loss: 2.003981113433838
Iteration 158, loss: 1.9668030738830566
Iteration 159, loss: 1.8045355081558228
Iteration 160, loss: 2.088404417037964
Iteration 161, loss: 2.134291887283325
Iteration 162, loss: 1.9927113056182861
Iteration 163, loss: 2.0585904121398926
Iteration 164, loss: 1.9518983364105225
Iteration 165, loss: 1.9679287672042847
Iteration 166, loss: 2.0235939025878906
Iteration 167, loss: 2.0866246223449707
Iteration 168, loss: 2.089888095855713
Iteration 169, loss: 2.1225671768188477
Iteration 170, loss: 2.0245938301086426
Iteration 171, loss: 2.271146297454834
Iteration 172, loss: 2.0944483280181885
Iteration 173, loss: 2.1009469032287598
Iteration 174, loss: 2.011967420578003
Iteration 175, loss: 2.0542373657226562
```

```
Iteration 176, loss: 1.9508408308029175
Iteration 177, loss: 2.0472500324249268
Iteration 178, loss: 2.045583486557007
Iteration 179, loss: 1.9696439504623413
Iteration 180, loss: 2.0171608924865723
Iteration 181, loss: 2.1111767292022705
Iteration 182, loss: 1.935225248336792
Iteration 183, loss: 2.1378114223480225
Iteration 184, loss: 2.0426788330078125
Iteration 185, loss: 2.0888752937316895
Iteration 186, loss: 1.9742710590362549
Iteration 187, loss: 2.166198968887329
Iteration 188, loss: 2.0027518272399902
Iteration 189, loss: 2.068073272705078
Iteration 190, loss: 1.9823483228683472
Iteration 191, loss: 2.0196993350982666
Iteration 192, loss: 2.047586679458618
Iteration 193, loss: 2.1563448905944824
Iteration 194, loss: 1.9539859294891357
Iteration 195, loss: 1.986884355545044
Iteration 196, loss: 2.1059865951538086
Iteration 197, loss: 1.931540608406067
Iteration 198, loss: 2.0165505409240723
Iteration 199, loss: 1.8822137117385864
Iteration 200, loss: 2.077021360397339
Iteration 201, loss: 2.0530223846435547
Iteration 202, loss: 2.0787456035614014
Iteration 203, loss: 1.9242535829544067
Iteration 204, loss: 1.992577075958252
Iteration 205, loss: 2.1751043796539307
Iteration 206, loss: 1.9798588752746582
Iteration 207, loss: 1.9751088619232178
Iteration 208, loss: 2.0420751571655273
Iteration 209, loss: 1.9609736204147339
Iteration 210, loss: 2.0552968978881836
Iteration 211, loss: 2.081383228302002
Iteration 212, loss: 2.0781099796295166
Iteration 213, loss: 1.9857921600341797
Iteration 214, loss: 2.013315200805664
Iteration 215, loss: 1.8571289777755737
Iteration 216, loss: 2.0300519466400146
Iteration 217, loss: 2.112175464630127
Iteration 218, loss: 1.9699478149414062
Iteration 219, loss: 2.077479600906372
Iteration 220, loss: 1.9170165061950684
Iteration 221, loss: 1.9764288663864136
Iteration 222, loss: 1.961592197418213
Iteration 223, loss: 1.9829617738723755
Iteration 224, loss: 2.151228904724121
Iteration 225, loss: 2.18438720703125
Iteration 226, loss: 1.9649739265441895
Iteration 227, loss: 2.164018392562866
Iteration 228, loss: 1.9981420040130615
Iteration 229, loss: 2.092895269393921
Iteration 230, loss: 1.9253991842269897
Iteration 231, loss: 1.8545076847076416
Iteration 232, loss: 1.934483289718628
Iteration 233, loss: 2.0551199913024902
Iteration 234, loss: 1.9438329935073853
Iteration 235, loss: 1.9475306272506714
Iteration 236, loss: 1.9360616207122803
Iteration 237, loss: 1.9380791187286377
Iteration 238, loss: 2.006498336791992
Iteration 239, loss: 1.9326748847961426
Iteration 240, loss: 1.9211819171905518
Iteration 241, loss: 1.9796712398529053
Iteration 242, loss: 2.1458330154418945
Iteration 243, loss: 1.9610875844955444
Iteration 244, loss: 1.9238337278366089
Iteration 245, loss: 1.9284703731536865
Iteration 246, loss: 1.9159818887710571
Iteration 247, loss: 1.9629369974136353
Iteration 248, loss: 1.9994964599609375
Iteration 249, loss: 1.8399860858917236
Iteration 250, loss: 2.047532320022583
Iteration 251, loss: 1.9068814516067505
Iteration 252, loss: 2.0274126529693604
```

```
Iteration 253, loss: 1.9852104187011719
Iteration 254, loss: 1.9754513502120972
Iteration 255, loss: 1.9422882795333862
Iteration 256, loss: 2.1017847061157227
Iteration 257, loss: 2.036468267440796
Iteration 258, loss: 1.9556183815002441
Iteration 259, loss: 1.940093994140625
Iteration 260, loss: 1.9169731140136719
Iteration 261, loss: 1.9315234422683716
Iteration 262, loss: 1.9236990213394165
Iteration 263, loss: 2.0450692176818848
Iteration 264, loss: 1.9676927328109741
Iteration 265, loss: 2.0694739818573
Iteration 266, loss: 2.1179420948028564
Iteration 267, loss: 1.9304496049880981
Iteration 268, loss: 1.997780442237854
Iteration 269, loss: 1.9963716268539429
Iteration 270, loss: 1.843829870223999
Iteration 271, loss: 1.8577027320861816
Iteration 272, loss: 2.0689167976379395
Iteration 273, loss: 2.003418445587158
Iteration 274, loss: 1.8677170276641846
Iteration 275, loss: 1.9745173454284668
Iteration 276, loss: 2.0695877075195312
Iteration 277, loss: 1.8724172115325928
Iteration 278, loss: 1.8525981903076172
Iteration 279, loss: 2.0287840366363525
Iteration 280, loss: 1.9900346994400024
Iteration 281, loss: 1.91969895362854
Iteration 282, loss: 1.946536898612976
Iteration 283, loss: 1.8897203207015991
Iteration 284, loss: 1.952713131904602
Iteration 285, loss: 1.9445546865463257
Iteration 286, loss: 2.171313762664795
Iteration 287, loss: 1.9669554233551025
Iteration 288, loss: 1.9759389162063599
Iteration 289, loss: 1.9482213258743286
Iteration 290, loss: 1.8426718711853027
Iteration 291, loss: 2.0090112686157227
Iteration 292, loss: 2.0252938270568848
Iteration 293, loss: 1.9046759605407715
Iteration 294, loss: 1.9537148475646973
Iteration 295, loss: 1.9754656553268433
Iteration 296, loss: 1.9769532680511475
Iteration 297, loss: 2.0057661533355713
Iteration 298, loss: 1.9550414085388184
Iteration 299, loss: 1.861952304840088
Iteration 300, loss: 1.9058661460876465
Iteration 301, loss: 1.9591012001037598
Iteration 302, loss: 1.9591631889343262
Iteration 303, loss: 2.0115466117858887
Iteration 304, loss: 1.9146374464035034
Iteration 305, loss: 1.7828272581100464
Iteration 306, loss: 2.0220930576324463
Iteration 307, loss: 1.9557175636291504
Iteration 308, loss: 1.9095458984375
Iteration 309, loss: 1.97672438621521
Iteration 310, loss: 1.9152662754058838
Iteration 311, loss: 1.9989210367202759
Iteration 312, loss: 1.9215143918991089
Iteration 313, loss: 1.9340577125549316
Iteration 314, loss: 1.8858494758605957
Iteration 315, loss: 1.9690831899642944
Iteration 316, loss: 2.00799298286438
Iteration 317, loss: 1.9208744764328003
Iteration 318, loss: 1.887947678565979
Iteration 319, loss: 1.9070827960968018
Iteration 320, loss: 1.8705533742904663
Iteration 321, loss: 1.8514020442962646
Iteration 322, loss: 1.9556580781936646
Iteration 323, loss: 1.9275097846984863
Iteration 324, loss: 2.155219793319702
Iteration 325, loss: 1.9154573678970337
Iteration 326, loss: 1.9193761348724365
Iteration 327, loss: 1.825456142425537
Iteration 328, loss: 1.9249247312545776
Iteration 329, loss: 1.9966404438018799
```

```
Iteration 330, loss: 1.9787864685058594
Iteration 331, loss: 2.0151567459106445
Iteration 332, loss: 1.728676438331604
Iteration 333, loss: 1.8985272645950317
Iteration 334, loss: 1.9236539602279663
Iteration 335, loss: 1.896832823753357
Iteration 336, loss: 2.0040929317474365
Iteration 337, loss: 1.9387116432189941
Iteration 338, loss: 1.9463953971862793
Iteration 339, loss: 2.0070292949676514
Iteration 340, loss: 1.8939893245697021
Iteration 341, loss: 1.996390700340271
Iteration 342, loss: 1.922744870185852
Iteration 343, loss: 1.9392725229263306
Iteration 344, loss: 1.897303819656372
Iteration 345, loss: 1.9799346923828125
Iteration 346, loss: 1.919259786605835
Iteration 347, loss: 1.9738047122955322
Iteration 348, loss: 1.917752742767334
Iteration 349, loss: 1.8850655555725098
Iteration 350, loss: 1.796546459197998
Iteration 351, loss: 1.945879578590393
Iteration 352, loss: 2.009493827819824
Iteration 353, loss: 1.9662266969680786
Iteration 354, loss: 1.8639775514602661
Iteration 355, loss: 1.8245315551757812
Iteration 356, loss: 1.997154951095581
Iteration 357, loss: 1.9393823146820068
Iteration 358, loss: 1.818972110748291
Iteration 359, loss: 2.040086507797241
Iteration 360, loss: 1.9647421836853027
Iteration 361, loss: 2.0066378116607666
Iteration 362, loss: 1.964367389678955
Iteration 363, loss: 1.7420668601989746
Iteration 364, loss: 1.8858201503753662
Iteration 365, loss: 1.9928869009017944
Iteration 366, loss: 1.9949562549591064
Iteration 367, loss: 1.911798357963562
Iteration 368, loss: 1.880333662033081
Iteration 369, loss: 1.9128613471984863
Iteration 370, loss: 1.8601518869400024
Iteration 371, loss: 1.8502388000488281
Iteration 372, loss: 1.865505337715149
Iteration 373, loss: 1.7806504964828491
Iteration 374, loss: 1.9226895570755005
Iteration 375, loss: 1.938930630683899
Iteration 376, loss: 1.8561757802963257
Iteration 377, loss: 1.8742384910583496
Iteration 378, loss: 1.9330241680145264
Iteration 379, loss: 1.9315357208251953
Iteration 380, loss: 1.9850735664367676
Iteration 381, loss: 1.916341781616211
Iteration 382, loss: 1.8520195484161377
Iteration 383, loss: 1.8918577432632446
Iteration 384, loss: 1.8603967428207397
Iteration 385, loss: 1.9760215282440186
Iteration 386, loss: 1.9712920188903809
Iteration 387, loss: 1.9385504722595215
Iteration 388, loss: 1.7867205142974854
Iteration 389, loss: 1.9266002178192139
Iteration 390, loss: 1.8316614627838135
Iteration 391, loss: 1.832726240158081
Iteration 392, loss: 1.8818773031234741
Iteration 393, loss: 1.9032686948776245
Iteration 394, loss: 1.822462558746338
Iteration 395, loss: 1.9311624765396118
Iteration 396, loss: 1.883085012435913
Iteration 397, loss: 1.8428447246551514
Iteration 398, loss: 1.9074158668518066
Iteration 399, loss: 1.9079382419586182
Iteration 400, loss: 1.9118670225143433
Iteration 401, loss: 1.9126006364822388
Iteration 402, loss: 1.8430705070495605
Iteration 403, loss: 1.9717341661453247
Iteration 404, loss: 1.9527961015701294
Iteration 405, loss: 1.9442811012268066
Iteration 406, loss: 1.782196044921875
```

```
Iteration 407, loss: 1.8362542390823364
Iteration 408, loss: 1.9432324171066284
Iteration 409, loss: 1.9126372337341309
Iteration 410, loss: 1.8982182741165161
Iteration 411, loss: 1.9094939231872559
Iteration 412, loss: 1.9709644317626953
Iteration 413, loss: 1.8194655179977417
Iteration 414, loss: 1.8449349403381348
Iteration 415, loss: 1.841445803642273
Iteration 416, loss: 1.9139777421951294
Iteration 417, loss: 1.9075437784194946
Iteration 418, loss: 1.86604642868042
Iteration 419, loss: 1.8608489036560059
Iteration 420, loss: 1.8657888174057007
Iteration 421, loss: 2.126960039138794
Iteration 422, loss: 1.775199294090271
Iteration 423, loss: 1.7353171110153198
Iteration 424, loss: 1.9893672466278076
Iteration 425, loss: 1.8612778186798096
Iteration 426, loss: 1.8114755153656006
Iteration 427, loss: 1.9118030071258545
Iteration 428, loss: 1.8149930238723755
Iteration 429, loss: 1.8561921119689941
Iteration 430, loss: 1.8522385358810425
Iteration 431, loss: 1.9672925472259521
Iteration 432, loss: 1.7637816667556763
Iteration 433, loss: 1.860709309577942
Iteration 434, loss: 1.7934144735336304
Iteration 435, loss: 1.8362274169921875
Iteration 436, loss: 1.8293393850326538
Iteration 437, loss: 1.8477190732955933
Iteration 438, loss: 1.8270480632781982
Iteration 439, loss: 1.8988354206085205
Iteration 440, loss: 1.8462704420089722
Iteration 441, loss: 1.818006992340088
Iteration 442, loss: 1.80185067653656
Iteration 443, loss: 1.7920295000076294
Iteration 444, loss: 1.86992347240448
Iteration 445, loss: 1.838904619216919
Iteration 446, loss: 1.8346868753433228
Iteration 447, loss: 1.791083812713623
Iteration 448, loss: 1.960571527481079
Iteration 449, loss: 1.7511347532272339
Iteration 450, loss: 1.8079853057861328
Iteration 451, loss: 1.7638911008834839
Iteration 452, loss: 1.7640951871871948
Iteration 453, loss: 1.6984431743621826
Iteration 454, loss: 1.8302065134048462
Iteration 455, loss: 1.8377933502197266
Iteration 456, loss: 1.8074166774749756
Iteration 457, loss: 1.8425376415252686
Iteration 458, loss: 1.8093960285186768
Iteration 459, loss: 1.8207730054855347
Iteration 460, loss: 1.9162614345550537
Iteration 461, loss: 1.9027619361877441
Iteration 462, loss: 1.787271499633789
Iteration 463, loss: 1.789109468460083
Iteration 464, loss: 1.7844384908676147
Iteration 465, loss: 1.9527842998504639
Iteration 466, loss: 1.8752264976501465
Iteration 467, loss: 1.8237113952636719
Iteration 468, loss: 1.9727927446365356
Iteration 469, loss: 1.6330634355545044
Iteration 470, loss: 1.8344377279281616
Iteration 471, loss: 1.8457146883010864
Iteration 472, loss: 1.8093382120132446
Iteration 473, loss: 1.802375316619873
Iteration 474, loss: 1.7551759481430054
Iteration 475, loss: 1.810873031616211
Iteration 476, loss: 1.9113385677337646
Iteration 477, loss: 1.8843964338302612
Iteration 478, loss: 1.7544320821762085
Iteration 479, loss: 1.889259934425354
Iteration 480, loss: 1.9177924394607544
Iteration 481, loss: 1.9399218559265137
Iteration 482, loss: 1.7946110963821411
Iteration 483, loss: 1.7927063703536987
```

```
Iteration 484, loss: 1.8914827108383179
Iteration 485, loss: 1.787841558456421
Iteration 486, loss: 1.8201191425323486
Iteration 487, loss: 1.809130311012268
Iteration 488, loss: 1.791083812713623
Iteration 489, loss: 1.7881938219070435
Iteration 490, loss: 1.7745705842971802
Iteration 491, loss: 1.8065158128738403
Iteration 492, loss: 1.8565144538879395
Iteration 493, loss: 1.8446834087371826
Iteration 494, loss: 1.749017596244812
Iteration 495, loss: 1.826041340827942
Iteration 496, loss: 1.7928961515426636
Iteration 497, loss: 1.6975640058517456
Iteration 498, loss: 1.843784213066101
Iteration 499, loss: 1.8831288814544678
Iteration 500, loss: 1.7508533000946045
Iteration 501, loss: 1.8814703226089478
Iteration 502, loss: 1.8660956621170044
Iteration 503, loss: 1.854249119758606
Iteration 504, loss: 1.871775507926941
Iteration 505, loss: 1.8357876539230347
Iteration 506, loss: 1.8488332033157349
Iteration 507, loss: 1.7940953969955444
Iteration 508, loss: 1.7659389972686768
Iteration 509, loss: 1.7675493955612183
Iteration 510, loss: 1.8547027111053467
Iteration 511, loss: 1.8199049234390259
Iteration 512, loss: 1.8729853630065918
Iteration 513, loss: 1.8334286212921143
Iteration 514, loss: 1.8815772533416748
Iteration 515, loss: 1.7518577575683594
Iteration 516, loss: 1.7005528211593628
Iteration 517, loss: 1.7139356136322021
Iteration 518, loss: 1.720762014389038
Iteration 519, loss: 1.8360788822174072
Iteration 520, loss: 1.7358222007751465
Iteration 521, loss: 1.8208224773406982
Iteration 522, loss: 1.7083313465118408
Iteration 523, loss: 1.7552764415740967
Iteration 524, loss: 1.7543537616729736
Iteration 525, loss: 1.8634436130523682
Iteration 526, loss: 1.666386365890503
Iteration 527, loss: 1.7413324117660522
Iteration 528, loss: 1.7010294198989868
Iteration 529, loss: 1.8865656852722168
Iteration 530, loss: 1.914930820465088
Iteration 531, loss: 1.8004628419876099
Iteration 532, loss: 1.8074558973312378
Iteration 533, loss: 1.811671495437622
Iteration 534, loss: 1.7642267942428589
Iteration 535, loss: 1.8669885396957397
Iteration 536, loss: 1.8671796321868896
Iteration 537, loss: 1.79664146900177
Iteration 538, loss: 1.7020187377929688
Iteration 539, loss: 1.7718225717544556
Iteration 540, loss: 1.858567476272583
Iteration 541, loss: 1.8264944553375244
Iteration 542, loss: 1.817822813987732
Iteration 543, loss: 1.8081893920898438
Iteration 544, loss: 1.7967612743377686
Iteration 545, loss: 1.7332578897476196
Iteration 546, loss: 1.8747915029525757
Iteration 547, loss: 1.7044236660003662
Iteration 548, loss: 1.8020954132080078
Iteration 549, loss: 1.7537294626235962
Iteration 550, loss: 1.7448201179504395
Iteration 551, loss: 1.800747036933899
Iteration 552, loss: 1.7491343021392822
Iteration 553, loss: 1.7129591703414917
Iteration 554, loss: 1.7826734781265259
Iteration 555, loss: 1.6960053443908691
Iteration 556, loss: 1.812555193901062
Iteration 557, loss: 1.7847449779510498
Iteration 558, loss: 1.7912338972091675
Iteration 559, loss: 1.9417998790740967
Iteration 560, loss: 1.7269037961959839
```

```
Iteration 561, loss: 1.6466755867004395
Iteration 562, loss: 1.8096861839294434
Iteration 563, loss: 1.928317904472351
Iteration 564, loss: 1.816604733467102
Iteration 565, loss: 1.823573350906372
Iteration 566, loss: 1.7700307369232178
Iteration 567, loss: 1.7767698764801025
Iteration 568, loss: 1.7221386432647705
Iteration 569, loss: 1.873136281967163
Iteration 570, loss: 1.727462887763977
Iteration 571, loss: 1.7534401416778564
Iteration 572, loss: 1.8038703203201294
Iteration 573, loss: 1.6834185123443604
Iteration 574, loss: 1.6408659219741821
Iteration 575, loss: 1.7228827476501465
Iteration 576, loss: 1.6745890378952026
Iteration 577, loss: 1.6994588375091553
Iteration 578, loss: 1.7538235187530518
Iteration 579, loss: 1.7484697103500366
Iteration 580, loss: 1.7530972957611084
Iteration 581, loss: 1.6500391960144043
Iteration 582, loss: 1.6878306865692139
Iteration 583, loss: 1.7302221059799194
Iteration 584, loss: 1.7597966194152832
Iteration 585, loss: 1.743733286857605
Iteration 586, loss: 1.7368800640106201
Iteration 587, loss: 1.8615779876708984
Iteration 588, loss: 1.7006257772445679
Iteration 589, loss: 1.850282073020935
Iteration 590, loss: 1.7215704917907715
Iteration 591, loss: 1.754212498664856
Iteration 592, loss: 1.8662109375
Iteration 593, loss: 1.786583662033081
Iteration 594, loss: 1.6841362714767456
Iteration 595, loss: 1.638744592666626
Iteration 596, loss: 1.7569849491119385
Iteration 597, loss: 1.7364557981491089
Iteration 598, loss: 1.854649305343628
Iteration 599, loss: 1.6809213161468506
Iteration 600, loss: 1.7452471256256104
Iteration 601, loss: 1.691030740737915
Iteration 602, loss: 1.7228153944015503
Iteration 603, loss: 1.7534444332122803
Iteration 604, loss: 1.7169686555862427
Iteration 605, loss: 1.6717753410339355
Iteration 606, loss: 1.7172291278839111
Iteration 607, loss: 1.6606605052947998
Iteration 608, loss: 1.6365934610366821
Iteration 609, loss: 1.7875902652740479
Iteration 610, loss: 1.849261999130249
Iteration 611, loss: 1.6594562530517578
Iteration 612, loss: 1.685341715812683
Iteration 613, loss: 1.6661325693130493
Iteration 614, loss: 1.7497637271881104
Iteration 615, loss: 1.8504961729049683
Iteration 616, loss: 1.7558261156082153
Iteration 617, loss: 1.738899827003479
Iteration 618, loss: 1.7021713256835938
Iteration 619, loss: 1.7569983005523682
Iteration 620, loss: 1.7063905000686646
Iteration 621, loss: 1.783568024635315
Iteration 622, loss: 1.7522740364074707
Iteration 623, loss: 1.687239170074463
Iteration 624, loss: 1.7258727550506592
Iteration 625, loss: 1.7574234008789062
Iteration 626, loss: 1.7828433513641357
Iteration 627, loss: 1.8163145780563354
Iteration 628, loss: 1.6817100048065186
Iteration 629, loss: 1.6452363729476929
Iteration 630, loss: 1.7829769849777222
Iteration 631, loss: 1.6857120990753174
Iteration 632, loss: 1.6764971017837524
Iteration 633, loss: 1.7052401304244995
Iteration 634, loss: 1.6789957284927368
Iteration 635, loss: 1.6442315578460693
Iteration 636, loss: 1.7526841163635254
Iteration 637, loss: 1.715026617050171
```

```
Iteration 638, loss: 1.8074710369110107
Iteration 639, loss: 1.7319777011871338
Iteration 640, loss: 1.7360851764678955
Iteration 641, loss: 1.8275372982025146
Iteration 642, loss: 1.7309311628341675
Iteration 643, loss: 1.7815425395965576
Iteration 644, loss: 1.7537264823913574
Iteration 645, loss: 1.7387646436691284
Iteration 646, loss: 1.7857203483581543
Iteration 647, loss: 1.6257407665252686
Iteration 648, loss: 1.7799897193908691
Iteration 649, loss: 1.6965906620025635
Iteration 650, loss: 1.6513464450836182
Iteration 651, loss: 1.793963074684143
Iteration 652, loss: 1.6598498821258545
Iteration 653, loss: 1.7216734886169434
Iteration 654, loss: 1.7846757173538208
Iteration 655, loss: 1.6761748790740967
Iteration 656, loss: 1.7151185274124146
Iteration 657, loss: 1.7039496898651123
Iteration 658, loss: 1.7856545448303223
Iteration 659, loss: 1.7135803699493408
Iteration 660, loss: 1.5935108661651611
Iteration 661, loss: 1.6685192584991455
Iteration 662, loss: 1.7716542482376099
Iteration 663, loss: 1.5597821474075317
Iteration 664, loss: 1.672438144683838
Iteration 665, loss: 1.7776976823806763
Iteration 666, loss: 1.6900005340576172
Iteration 667, loss: 1.675328016281128
Iteration 668, loss: 1.6448076963424683
Iteration 669, loss: 1.7575318813323975
Iteration 670, loss: 1.7353510856628418
Iteration 671, loss: 1.6895956993103027
Iteration 672, loss: 1.698893427848816
Iteration 673, loss: 1.7804350852966309
Iteration 674, loss: 1.6880309581756592
Iteration 675, loss: 1.6729494333267212
Iteration 676, loss: 1.6940956115722656
Iteration 677, loss: 1.6905051469802856
Iteration 678, loss: 1.7579059600830078
Iteration 679, loss: 1.7366467714309692
Iteration 680, loss: 1.7326858043670654
Iteration 681, loss: 1.6839158535003662
Iteration 682, loss: 1.6758790016174316
Iteration 683, loss: 1.64851975440979
Iteration 684, loss: 1.6054816246032715
Iteration 685, loss: 1.7336820363998413
Iteration 686, loss: 1.704514980316162
Iteration 687, loss: 1.7177914381027222
Iteration 688, loss: 1.7292410135269165
Iteration 689, loss: 1.6531410217285156
Iteration 690, loss: 1.6481693983078003
Iteration 691, loss: 1.7095000743865967
Iteration 692, loss: 1.644285798072815
Iteration 693, loss: 1.7275818586349487
Iteration 694, loss: 1.7057503461837769
Iteration 695, loss: 1.8046005964279175
Iteration 696, loss: 1.7879524230957031
Iteration 697, loss: 1.653816819190979
Iteration 698, loss: 1.728151798248291
Iteration 699, loss: 1.7414188385009766
Iteration 700, loss: 1.8720206022262573
Iteration 701, loss: 1.673262357711792
Iteration 702, loss: 1.6831231117248535
Iteration 703, loss: 1.7178491353988647
Iteration 704, loss: 1.7069512605667114
Iteration 705, loss: 1.681898593902588
Iteration 706, loss: 1.6864262819290161
Iteration 707, loss: 1.6754294633865356
Iteration 708, loss: 1.6999335289001465
Iteration 709, loss: 1.695687174797058
Iteration 710, loss: 1.64314603805542
Iteration 711, loss: 1.7342942953109741
Iteration 712, loss: 1.6060187816619873
Iteration 713, loss: 1.6501052379608154
Iteration 714, loss: 1.7077555656433105
```

```
Iteration 715, loss: 1.6299359798431396
Iteration 716, loss: 1.6039025783538818
Iteration 717, loss: 1.6827114820480347
Iteration 718, loss: 1.6492348909378052
Iteration 719, loss: 1.6882356405258179
Iteration 720, loss: 1.672111988067627
Iteration 721, loss: 1.6936246156692505
Iteration 722, loss: 1.622063398361206
Iteration 723, loss: 1.5688153505325317
Iteration 724, loss: 1.7411452531814575
Iteration 725, loss: 1.604508638381958
Iteration 726, loss: 1.705819845199585
Iteration 727, loss: 1.7214484214782715
Iteration 728, loss: 1.6580103635787964
Iteration 729, loss: 1.6202794313430786
Iteration 730, loss: 1.6848310232162476
Iteration 731, loss: 1.6948062181472778
Iteration 732, loss: 1.6626616716384888
Iteration 733, loss: 1.6691229343414307
Iteration 734, loss: 1.680801510810852
Iteration 735, loss: 1.624714732170105
Iteration 736, loss: 1.7570329904556274
Iteration 737, loss: 1.6890361309051514
Iteration 738, loss: 1.7324529886245728
Iteration 739, loss: 1.6167423725128174
Iteration 740, loss: 1.6524068117141724
Iteration 741, loss: 1.6808794736862183
Iteration 742, loss: 1.757312536239624
Iteration 743, loss: 1.6786175966262817
Iteration 744, loss: 1.5668768882751465
Iteration 745, loss: 1.6253260374069214
Iteration 746, loss: 1.6921483278274536
Iteration 747, loss: 1.7465509176254272
Iteration 748, loss: 1.6544606685638428
Iteration 749, loss: 1.6000194549560547
Iteration 750, loss: 1.666516661643982
Iteration 751, loss: 1.5881530046463013
Iteration 752, loss: 1.6991100311279297
Iteration 753, loss: 1.728202223777771
Iteration 754, loss: 1.5935490131378174
Iteration 755, loss: 1.5823816061019897
Iteration 756, loss: 1.686063528060913
Iteration 757, loss: 1.6189271211624146
Iteration 758, loss: 1.592006802558899
Iteration 759, loss: 1.7155320644378662
Iteration 760, loss: 1.5271254777908325
Iteration 761, loss: 1.6667448282241821
Iteration 762, loss: 1.619864583015442
Iteration 763, loss: 1.6970059871673584
Iteration 764, loss: 1.6634738445281982
Iteration 765, loss: 1.5729811191558838
Iteration 766, loss: 1.5011234283447266
Iteration 767, loss: 1.6719696521759033
Iteration 768, loss: 1.6597671508789062
Iteration 769, loss: 1.6096491813659668
Iteration 770, loss: 1.642256498336792
Iteration 771, loss: 1.5271987915039062
Iteration 772, loss: 1.6960904598236084
Iteration 773, loss: 1.5809710025787354
Iteration 774, loss: 1.6470013856887817
Iteration 775, loss: 1.6569219827651978
Iteration 776, loss: 1.661598801612854
Iteration 777, loss: 1.7197331190109253
Iteration 778, loss: 1.662959337234497
Iteration 779, loss: 1.5705108642578125
Iteration 780, loss: 1.6720860004425049
Iteration 781, loss: 1.5969643592834473
Iteration 782, loss: 1.5280216932296753
Iteration 783, loss: 1.5827630758285522
Iteration 784, loss: 1.6087876558303833
Iteration 785, loss: 1.5496799945831299
Iteration 786, loss: 1.5628654956817627
Iteration 787, loss: 1.6576745510101318
Iteration 788, loss: 1.6425951719284058
Iteration 789, loss: 1.7294983863830566
Iteration 790, loss: 1.5460517406463623
Iteration 791, loss: 1.6754531860351562
```

```
Iteration 792, loss: 1.558158040046692
Iteration 793, loss: 1.5695953369140625
Iteration 794, loss: 1.536321759223938
Iteration 795, loss: 1.5984771251678467
Iteration 796, loss: 1.7256224155426025
Iteration 797, loss: 1.5431303977966309
Iteration 798, loss: 1.5820757150650024
Iteration 799, loss: 1.540722370147705
Iteration 800, loss: 1.599899411201477
Iteration 801, loss: 1.5161705017089844
Iteration 802, loss: 1.7196717262268066
Iteration 803, loss: 1.4474443197250366
Iteration 804, loss: 1.5923535823822021
Iteration 805, loss: 1.6449567079544067
Iteration 806, loss: 1.5457795858383179
Iteration 807, loss: 1.563801646232605
Iteration 808, loss: 1.6749184131622314
Iteration 809, loss: 1.667065978050232
Iteration 810, loss: 1.6874959468841553
Iteration 811, loss: 1.609415054321289
Iteration 812, loss: 1.6722314357757568
Iteration 813, loss: 1.5333107709884644
Iteration 814, loss: 1.6681946516036987
Iteration 815, loss: 1.685577154159546
Iteration 816, loss: 1.5258936882019043
Iteration 817, loss: 1.6454863548278809
Iteration 818, loss: 1.597969889640808
Iteration 819, loss: 1.6158740520477295
Iteration 820, loss: 1.658793568611145
Iteration 821, loss: 1.7647240161895752
Iteration 822, loss: 1.6512480974197388
Iteration 823, loss: 1.5291730165481567
Iteration 824, loss: 1.6482771635055542
Iteration 825, loss: 1.5673327445983887
Iteration 826, loss: 1.5947203636169434
Iteration 827, loss: 1.6228673458099365
Iteration 828, loss: 1.5929760932922363
Iteration 829, loss: 1.6101473569869995
Iteration 830, loss: 1.5482290983200073
Iteration 831, loss: 1.6944730281829834
Iteration 832, loss: 1.5294028520584106
Iteration 833, loss: 1.5736409425735474
Iteration 834, loss: 1.5646315813064575
Iteration 835, loss: 1.7275426387786865
Iteration 836, loss: 1.527368426322937
Iteration 837, loss: 1.6410268545150757
Iteration 838, loss: 1.5877171754837036
Iteration 839, loss: 1.5749685764312744
Iteration 840, loss: 1.5901143550872803
Iteration 841, loss: 1.672803282737732
Iteration 842, loss: 1.5624359846115112
Iteration 843, loss: 1.5423005819320679
Iteration 844, loss: 1.598609209060669
Iteration 845, loss: 1.5605394840240479
Iteration 846, loss: 1.537696123123169
Iteration 847, loss: 1.6064927577972412
Iteration 848, loss: 1.7485378980636597
Iteration 849, loss: 1.6341203451156616
Iteration 850, loss: 1.5064823627471924
Iteration 851, loss: 1.5916188955307007
Iteration 852, loss: 1.599907636642456
Iteration 853, loss: 1.6639949083328247
Iteration 854, loss: 1.5493050813674927
Iteration 855, loss: 1.5819497108459473
Iteration 856, loss: 1.556657314300537
Iteration 857, loss: 1.7058964967727661
Iteration 858, loss: 1.6475462913513184
Iteration 859, loss: 1.5398286581039429
Iteration 860, loss: 1.6963578462600708
Iteration 861, loss: 1.6180425882339478
Iteration 862, loss: 1.5438497066497803
Iteration 863, loss: 1.618552565574646
Iteration 864, loss: 1.5247511863708496
Iteration 865, loss: 1.620995044708252
Iteration 866, loss: 1.5899090766906738
Iteration 867, loss: 1.5861186981201172
Iteration 868, loss: 1.5923957824707031
```

```
Iteration 869, loss: 1.5310540199279785
Iteration 870, loss: 1.5747768878936768
Iteration 871, loss: 1.5803316831588745
Iteration 872, loss: 1.5264294147491455
Iteration 873, loss: 1.563932180404663
Iteration 874, loss: 1.6098709106445312
Iteration 875, loss: 1.5531322956085205
Iteration 876, loss: 1.4690011739730835
Iteration 877, loss: 1.4752405881881714
Iteration 878, loss: 1.7011809349060059
Iteration 879, loss: 1.579921841621399
Iteration 880, loss: 1.530731201171875
Iteration 881, loss: 1.5813783407211304
Iteration 882, loss: 1.5846726894378662
Iteration 883, loss: 1.5180550813674927
Iteration 884, loss: 1.6167486906051636
Iteration 885, loss: 1.5655100345611572
Iteration 886, loss: 1.6469289064407349
Iteration 887, loss: 1.5468273162841797
Iteration 888, loss: 1.4941961765289307
Iteration 889, loss: 1.569062352180481
Iteration 890, loss: 1.5733530521392822
Iteration 891, loss: 1.621049404144287
Iteration 892, loss: 1.5249228477478027
Iteration 893, loss: 1.6075317859649658
Iteration 894, loss: 1.6103771924972534
Iteration 895, loss: 1.528605580329895
Iteration 896, loss: 1.507011890411377
Iteration 897, loss: 1.5740243196487427
Iteration 898, loss: 1.544782042503357
Iteration 899, loss: 1.4925144910812378
Iteration 900, loss: 1.582515001296997
Iteration 901, loss: 1.5583593845367432
Iteration 902, loss: 1.5127589702606201
Iteration 903, loss: 1.4437251091003418
Iteration 904, loss: 1.5577139854431152
Iteration 905, loss: 1.6257935762405396
Iteration 906, loss: 1.540684461593628
Iteration 907, loss: 1.5503450632095337
Iteration 908, loss: 1.5593541860580444
Iteration 909, loss: 1.540150761604309
Iteration 910, loss: 1.5702906847000122
Iteration 911, loss: 1.5029104948043823
Iteration 912, loss: 1.5277358293533325
Iteration 913, loss: 1.4265391826629639
Iteration 914, loss: 1.4440492391586304
Iteration 915, loss: 1.5873775482177734
Iteration 916, loss: 1.5766297578811646
Iteration 917, loss: 1.441650152206421
Iteration 918, loss: 1.508951187133789
Iteration 919, loss: 1.5874321460723877
Iteration 920, loss: 1.523332118988037
Iteration 921, loss: 1.54618501663208
Iteration 922, loss: 1.5039705038070679
Iteration 923, loss: 1.5272941589355469
Iteration 924, loss: 1.5284206867218018
Iteration 925, loss: 1.59745454788208
Iteration 926, loss: 1.5279470682144165
Iteration 927, loss: 1.5298035144805908
Iteration 928, loss: 1.5288777351379395
Iteration 929, loss: 1.3748854398727417
Iteration 930, loss: 1.6093090772628784
Iteration 931, loss: 1.5034699440002441
Iteration 932, loss: 1.5052803754806519
Iteration 933, loss: 1.5363770723342896
Iteration 934, loss: 1.5834319591522217
Iteration 935, loss: 1.4403795003890991
Iteration 936, loss: 1.5133004188537598
Iteration 937, loss: 1.541167974472046
Iteration 938, loss: 1.5015475749969482
Iteration 939, loss: 1.5193945169448853
Iteration 940, loss: 1.5450172424316406
Iteration 941, loss: 1.5052679777145386
Iteration 942, loss: 1.4784754514694214
Iteration 943, loss: 1.577802062034607
Iteration 944, loss: 1.590395212173462
Iteration 945, loss: 1.5272852182388306
```

```
Iteration 946, loss: 1.5108590126037598
Iteration 947, loss: 1.5241749286651611
Iteration 948, loss: 1.5287175178527832
Iteration 949, loss: 1.3985568284988403
Iteration 950, loss: 1.5153908729553223
Iteration 951, loss: 1.487787127494812
Iteration 952, loss: 1.4931540489196777
Iteration 953, loss: 1.516353726387024
Iteration 954, loss: 1.513668179512024
Iteration 955, loss: 1.5079492330551147
Iteration 956, loss: 1.510338306427002
Iteration 957, loss: 1.5268434286117554
Iteration 958, loss: 1.4149394035339355
Iteration 959, loss: 1.4847992658615112
Iteration 960, loss: 1.553638219833374
Iteration 961, loss: 1.4014503955841064
Iteration 962, loss: 1.5352715253829956
Iteration 963, loss: 1.6394740343093872
Iteration 964, loss: 1.526606798171997
Iteration 965, loss: 1.5371994972229004
Iteration 966, loss: 1.5032404661178589
Iteration 967, loss: 1.4889321327209473
Iteration 968, loss: 1.434739351272583
Iteration 969, loss: 1.5285282135009766
Iteration 970, loss: 1.5829453468322754
Iteration 971, loss: 1.5106950998306274
Iteration 972, loss: 1.4775352478027344
Iteration 973, loss: 1.5359541177749634
Iteration 974, loss: 1.4459775686264038
Iteration 975, loss: 1.4938172101974487
Iteration 976, loss: 1.4744776487350464
Iteration 977, loss: 1.552668809890747
Iteration 978, loss: 1.4663116931915283
Iteration 979, loss: 1.4829108715057373
Iteration 980, loss: 1.4177342653274536
Iteration 981, loss: 1.4846497774124146
Iteration 982, loss: 1.4166477918624878
Iteration 983, loss: 1.406527042388916
Iteration 984, loss: 1.476143479347229
Iteration 985, loss: 1.546494960784912
Iteration 986, loss: 1.5424433946609497
Iteration 987, loss: 1.5307673215866089
Iteration 988, loss: 1.5015016794204712
Iteration 989, loss: 1.4350147247314453
Iteration 990, loss: 1.5085906982421875
Iteration 991, loss: 1.5317720174789429
Iteration 992, loss: 1.502355694770813
Iteration 993, loss: 1.4493427276611328
Iteration 994, loss: 1.4412955045700073
Iteration 995, loss: 1.472898244857788
Iteration 996, loss: 1.472540020942688
Iteration 997, loss: 1.7294604778289795
Iteration 998, loss: 1.4540975093841553
Iteration 999, loss: 1.3948289155960083
End of epoch 4, loss: 2.8222713470458984
Validation loss: 1.8756128549575806
Iteration 0, loss: 1.5596344470977783
Iteration 1, loss: 1.4363118410110474
Iteration 2, loss: 1.4087754487991333
Iteration 3, loss: 1.4420262575149536
Iteration 4, loss: 1.482550859451294
Iteration 5, loss: 1.5007649660110474
Iteration 6, loss: 1.5038306713104248
Iteration 7, loss: 1.4295140504837036
Iteration 8, loss: 1.4806791543960571
Iteration 9, loss: 1.4236963987350464
Iteration 10, loss: 1.4848933219909668
Iteration 11, loss: 1.512355089187622
Iteration 12, loss: 1.5288249254226685
Iteration 13, loss: 1.4739562273025513
Iteration 14, loss: 1.5672872066497803
Iteration 15, loss: 1.5133156776428223
Iteration 16, loss: 1.5366578102111816
Iteration 17, loss: 1.5183762311935425
Iteration 18, loss: 1.599089503288269
Iteration 19, loss: 1.5120564699172974
Iteration 20, loss: 1.548127293586731
```

```
Iteration 21, loss: 1.5427354574203491
Iteration 22, loss: 1.5443406105041504
Iteration 23, loss: 1.3945943117141724
Iteration 24, loss: 1.3805819749832153
Iteration 25, loss: 1.537406086921692
Iteration 26, loss: 1.5544918775558472
Iteration 27, loss: 1.5582598447799683
Iteration 28, loss: 1.4599518775939941
Iteration 29, loss: 1.4774705171585083
Iteration 30, loss: 1.5140715837478638
Iteration 31, loss: 1.4333800077438354
Iteration 32, loss: 1.4276316165924072
Iteration 33, loss: 1.6020232439041138
Iteration 34, loss: 1.5140113830566406
Iteration 35, loss: 1.4410645961761475
Iteration 36, loss: 1.4675196409225464
Iteration 37, loss: 1.3865522146224976
Iteration 38, loss: 1.451187252998352
Iteration 39, loss: 1.4989359378814697
Iteration 40, loss: 1.4738755226135254
Iteration 41, loss: 1.5305167436599731
Iteration 42, loss: 1.3728617429733276
Iteration 43, loss: 1.45182204246521
Iteration 44, loss: 1.3793678283691406
Iteration 45, loss: 1.586746335029602
Iteration 46, loss: 1.549001932144165
Iteration 47, loss: 1.452315092086792
Iteration 48, loss: 1.4356237649917603
Iteration 49, loss: 1.4932575225830078
Iteration 50, loss: 1.5238901376724243
Iteration 51, loss: 1.5234369039535522
Iteration 52, loss: 1.466599702835083
Iteration 53, loss: 1.4980690479278564
Iteration 54, loss: 1.4025919437408447
Iteration 55, loss: 1.6309431791305542
Iteration 56, loss: 1.335480809211731
Iteration 57, loss: 1.480594515800476
Iteration 58, loss: 1.5733723640441895
Iteration 59, loss: 1.4015860557556152
Iteration 60, loss: 1.3541455268859863
Iteration 61, loss: 1.5302332639694214
Iteration 62, loss: 1.5073673725128174
Iteration 63, loss: 1.4592010974884033
Iteration 64, loss: 1.4266530275344849
Iteration 65, loss: 1.4815775156021118
Iteration 66, loss: 1.396858811378479
Iteration 67, loss: 1.5157992839813232
Iteration 68, loss: 1.473018765449524
Iteration 69, loss: 1.4902704954147339
Iteration 70, loss: 1.4174665212631226
Iteration 71, loss: 1.5275288820266724
Iteration 72, loss: 1.4554775953292847
Iteration 73, loss: 1.3886126279830933
Iteration 74, loss: 1.5533459186553955
Iteration 75, loss: 1.4461032152175903
Iteration 76, loss: 1.4298676252365112
Iteration 77, loss: 1.4853414297103882
Iteration 78, loss: 1.525928258895874
Iteration 79, loss: 1.334614872932434
Iteration 80, loss: 1.4032280445098877
Iteration 81, loss: 1.4280376434326172
Iteration 82, loss: 1.370916485786438
Iteration 83, loss: 1.4360642433166504
Iteration 84, loss: 1.4559741020202637
Iteration 85, loss: 1.3924181461334229
Iteration 86, loss: 1.3949127197265625
Iteration 87, loss: 1.416574478149414
Iteration 88, loss: 1.455509066581726
Iteration 89, loss: 1.4115333557128906
Iteration 90, loss: 1.4805482625961304
Iteration 91, loss: 1.5143005847930908
Iteration 92, loss: 1.370675802230835
Iteration 93, loss: 1.4782644510269165
Iteration 94, loss: 1.4186058044433594
Iteration 95, loss: 1.4576786756515503
Iteration 96, loss: 1.3872554302215576
Iteration 97, loss: 1.3792216777801514
```

```
Iteration 98, loss: 1.4561506509780884
Iteration 99, loss: 1.4434717893600464
Iteration 100, loss: 1.3750603199005127
Iteration 101, loss: 1.426074743270874
Iteration 102, loss: 1.4584252834320068
Iteration 103, loss: 1.331533432006836
Iteration 104, loss: 1.5206215381622314
Iteration 105, loss: 1.4331895112991333
Iteration 106, loss: 1.4607938528060913
Iteration 107, loss: 1.351362943649292
Iteration 108, loss: 1.399931788444519
Iteration 109, loss: 1.5471069812774658
Iteration 110, loss: 1.4756357669830322
Iteration 111, loss: 1.5066332817077637
Iteration 112, loss: 1.3842129707336426
Iteration 113, loss: 1.4927902221679688
Iteration 114, loss: 1.3858140707015991
Iteration 115, loss: 1.4830012321472168
Iteration 116, loss: 1.5299180746078491
Iteration 117, loss: 1.4592511653900146
Iteration 118, loss: 1.461091160774231
Iteration 119, loss: 1.453260898590088
Iteration 120, loss: 1.3798229694366455
Iteration 121, loss: 1.3559514284133911
Iteration 122, loss: 1.3662173748016357
Iteration 123, loss: 1.4584952592849731
Iteration 124, loss: 1.3927674293518066
Iteration 125, loss: 1.3981685638427734
Iteration 126, loss: 1.3728716373443604
Iteration 127, loss: 1.3638205528259277
Iteration 128, loss: 1.3524435758590698
Iteration 129, loss: 1.459883451461792
Iteration 130, loss: 1.34658944606781
Iteration 131, loss: 1.4112586975097656
Iteration 132, loss: 1.468291997909546
Iteration 133, loss: 1.3941371440887451
Iteration 134, loss: 1.3955460786819458
Iteration 135, loss: 1.3451939821243286
Iteration 136, loss: 1.354589581489563
Iteration 137, loss: 1.3734848499298096
Iteration 138, loss: 1.3950034379959106
Iteration 139, loss: 1.392844557762146
Iteration 140, loss: 1.3257349729537964
Iteration 141, loss: 1.386474609375
Iteration 142, loss: 1.4716626405715942
Iteration 143, loss: 1.3888087272644043
Iteration 144, loss: 1.3541430234909058
Iteration 145, loss: 1.4404524564743042
Iteration 146, loss: 1.3283915519714355
Iteration 147, loss: 1.4029462337493896
Iteration 148, loss: 1.4659130573272705
Iteration 149, loss: 1.4652491807937622
Iteration 150, loss: 1.3189036846160889
Iteration 151, loss: 1.40939462184906
Iteration 152, loss: 1.3183646202087402
Iteration 153, loss: 1.4451943635940552
Iteration 154, loss: 1.4434268474578857
Iteration 155, loss: 1.4608561992645264
Iteration 156, loss: 1.4298248291015625
Iteration 157, loss: 1.3357402086257935
Iteration 158, loss: 1.2985610961914062
Iteration 159, loss: 1.1856755018234253
Iteration 160, loss: 1.4303474426269531
Iteration 161, loss: 1.4052929878234863
Iteration 162, loss: 1.358593225479126
Iteration 163, loss: 1.3856474161148071
Iteration 164, loss: 1.3120135068893433
Iteration 165, loss: 1.3166123628616333
Iteration 166, loss: 1.3732839822769165
Iteration 167, loss: 1.4120169878005981
Iteration 168, loss: 1.3598361015319824
Iteration 169, loss: 1.4513111114501953
Iteration 170, loss: 1.3401639461517334
Iteration 171, loss: 1.566706895828247
Iteration 172, loss: 1.4134156703948975
Iteration 173, loss: 1.3724048137664795
Iteration 174, loss: 1.341186285018921
```

```
Iteration 174, loss: 1.3411660265016921
Iteration 175, loss: 1.3604718446731567
Iteration 176, loss: 1.2569516897201538
Iteration 177, loss: 1.3428795337677002
Iteration 178, loss: 1.3757586479187012
Iteration 179, loss: 1.2912869453430176
Iteration 180, loss: 1.3665167093276978
Iteration 181, loss: 1.4263685941696167
Iteration 182, loss: 1.3001879453659058
Iteration 183, loss: 1.4406723976135254
Iteration 184, loss: 1.3339976072311401
Iteration 185, loss: 1.3941019773483276
Iteration 186, loss: 1.3113702535629272
Iteration 187, loss: 1.4431469440460205
Iteration 188, loss: 1.340431809425354
Iteration 189, loss: 1.4054391384124756
Iteration 190, loss: 1.3246204853057861
Iteration 191, loss: 1.3625388145446777
Iteration 192, loss: 1.3755828142166138
Iteration 193, loss: 1.4629491567611694
Iteration 194, loss: 1.3242043256759644
Iteration 195, loss: 1.324126124382019
Iteration 196, loss: 1.4243966341018677
Iteration 197, loss: 1.2949836254119873
Iteration 198, loss: 1.3611228466033936
Iteration 199, loss: 1.273085117340088
Iteration 200, loss: 1.3692137002944946
Iteration 201, loss: 1.361983060836792
Iteration 202, loss: 1.3984875679016113
Iteration 203, loss: 1.2770100831985474
Iteration 204, loss: 1.3279340267181396
Iteration 205, loss: 1.4485691785812378
Iteration 206, loss: 1.3677120208740234
Iteration 207, loss: 1.3539810180664062
Iteration 208, loss: 1.38069748878479
Iteration 209, loss: 1.2931772470474243
Iteration 210, loss: 1.371057391166687
Iteration 211, loss: 1.4201878309249878
Iteration 212, loss: 1.4212806224822998
Iteration 213, loss: 1.3383328914642334
Iteration 214, loss: 1.376495361328125
Iteration 215, loss: 1.25276780128479
Iteration 216, loss: 1.368337869644165
Iteration 217, loss: 1.4391204118728638
Iteration 218, loss: 1.3177473545074463
Iteration 219, loss: 1.351011872291565
Iteration 220, loss: 1.3185685873031616
Iteration 221, loss: 1.328196406364441
Iteration 222, loss: 1.2886296510696411
Iteration 223, loss: 1.3218445777893066
Iteration 224, loss: 1.4405272006988525
Iteration 225, loss: 1.4975697994232178
Iteration 226, loss: 1.2726918458938599
Iteration 227, loss: 1.4481887817382812
Iteration 228, loss: 1.3411017656326294
Iteration 229, loss: 1.39577317237854
Iteration 230, loss: 1.283700704574585
Iteration 231, loss: 1.2297114133834839
Iteration 232, loss: 1.3012566566467285
Iteration 233, loss: 1.4092973470687866
Iteration 234, loss: 1.2899385690689087
Iteration 235, loss: 1.2945806980133057
Iteration 236, loss: 1.2913018465042114
Iteration 237, loss: 1.2862639427185059
Iteration 238, loss: 1.3368971347808838
Iteration 239, loss: 1.2690441608428955
Iteration 240, loss: 1.264349102973938
Iteration 241, loss: 1.2956994771957397
Iteration 242, loss: 1.479630708694458
Iteration 243, loss: 1.307304859161377
Iteration 244, loss: 1.2611281871795654
Iteration 245, loss: 1.214355230331421
Iteration 246, loss: 1.2566577196121216
Iteration 247, loss: 1.290147066116333
Iteration 248, loss: 1.2971123456954956
Iteration 249, loss: 1.211443543434143
Iteration 250, loss: 1.3706570863723755
Iteration 251, loss: 1.254279613494873
```

```
Iteration 251, loss: 1.254279613494875
Iteration 252, loss: 1.3693681955337524
Iteration 253, loss: 1.3251103162765503
Iteration 254, loss: 1.3278990983963013
Iteration 255, loss: 1.2945505380630493
Iteration 256, loss: 1.4039320945739746
Iteration 257, loss: 1.3737640380859375
Iteration 258, loss: 1.3376935720443726
Iteration 259, loss: 1.3181177377700806
Iteration 260, loss: 1.2726870775222778
Iteration 261, loss: 1.2973239421844482
Iteration 262, loss: 1.2451711893081665
Iteration 263, loss: 1.3571795225143433
Iteration 264, loss: 1.3003922700881958
Iteration 265, loss: 1.3976366519927979
Iteration 266, loss: 1.4114489555358887
Iteration 267, loss: 1.2829562425613403
Iteration 268, loss: 1.3131988048553467
Iteration 269, loss: 1.3315708637237549
Iteration 270, loss: 1.2109806537628174
Iteration 271, loss: 1.2425265312194824
Iteration 272, loss: 1.3806865215301514
Iteration 273, loss: 1.3468869924545288
Iteration 274, loss: 1.2135717868804932
Iteration 275, loss: 1.3244973421096802
Iteration 276, loss: 1.4099199771881104
Iteration 277, loss: 1.2317105531692505
Iteration 278, loss: 1.2278887033462524
Iteration 279, loss: 1.349281907081604
Iteration 280, loss: 1.3344238996505737
Iteration 281, loss: 1.264607310295105
Iteration 282, loss: 1.2837201356887817
Iteration 283, loss: 1.2742806673049927
Iteration 284, loss: 1.2911560535430908
Iteration 285, loss: 1.2966581583023071
Iteration 286, loss: 1.4802504777908325
Iteration 287, loss: 1.2877135276794434
Iteration 288, loss: 1.3144073486328125
Iteration 289, loss: 1.2843265533447266
Iteration 290, loss: 1.2066320180892944
Iteration 291, loss: 1.3289381265640259
Iteration 292, loss: 1.3446617126464844
Iteration 293, loss: 1.2396153211593628
Iteration 294, loss: 1.2997004985809326
Iteration 295, loss: 1.2826558351516724
Iteration 296, loss: 1.3087351322174072
Iteration 297, loss: 1.3398276567459106
Iteration 298, loss: 1.2915630340576172
Iteration 299, loss: 1.2616502046585083
Iteration 300, loss: 1.239378809928894
Iteration 301, loss: 1.3078416585922241
Iteration 302, loss: 1.2961161136627197
Iteration 303, loss: 1.3420733213424683
Iteration 304, loss: 1.295816421508789
Iteration 305, loss: 1.1671713590621948
Iteration 306, loss: 1.3669503927230835
Iteration 307, loss: 1.2865217924118042
Iteration 308, loss: 1.2947633266448975
Iteration 309, loss: 1.351732850074768
Iteration 310, loss: 1.2277852296829224
Iteration 311, loss: 1.3138283491134644
Iteration 312, loss: 1.3024380207061768
Iteration 313, loss: 1.2934436798095703
Iteration 314, loss: 1.2252461910247803
Iteration 315, loss: 1.3126182556152344
Iteration 316, loss: 1.3440895080566406
Iteration 317, loss: 1.2756716012954712
Iteration 318, loss: 1.2169066667556763
Iteration 319, loss: 1.2667380571365356
Iteration 320, loss: 1.2434049844741821
Iteration 321, loss: 1.2243998050689697
Iteration 322, loss: 1.3056082725524902
Iteration 323, loss: 1.244008183479309
Iteration 324, loss: 1.4624501466751099
Iteration 325, loss: 1.2641990184783936
Iteration 326, loss: 1.2718507051467896
Iteration 327, loss: 1.1798046827316284
Iteration 328, loss: 1.2342158555984497
```

```
Iteration 329, loss: 1.3535665273666382
Iteration 330, loss: 1.2824397087097168
Iteration 331, loss: 1.3226616382598877
Iteration 332, loss: 1.1242693662643433
Iteration 333, loss: 1.2426419258117676
Iteration 334, loss: 1.2675225734710693
Iteration 335, loss: 1.2723873853683472
Iteration 336, loss: 1.318410873413086
Iteration 337, loss: 1.2829434871673584
Iteration 338, loss: 1.3015074729919434
Iteration 339, loss: 1.3310887813568115
Iteration 340, loss: 1.2452123165130615
Iteration 341, loss: 1.3399327993392944
Iteration 342, loss: 1.2463390827178955
Iteration 343, loss: 1.2699381113052368
Iteration 344, loss: 1.2641698122024536
Iteration 345, loss: 1.2803676128387451
Iteration 346, loss: 1.2800250053405762
Iteration 347, loss: 1.3239773511886597
Iteration 348, loss: 1.259133219718933
Iteration 349, loss: 1.2330763339996338
Iteration 350, loss: 1.1518871784210205
Iteration 351, loss: 1.2929390668869019
Iteration 352, loss: 1.3645583391189575
Iteration 353, loss: 1.3050326108932495
Iteration 354, loss: 1.2428711652755737
Iteration 355, loss: 1.210695743560791
Iteration 356, loss: 1.3086867332458496
Iteration 357, loss: 1.3159281015396118
Iteration 358, loss: 1.1850711107254028
Iteration 359, loss: 1.3506615161895752
Iteration 360, loss: 1.2656203508377075
Iteration 361, loss: 1.3085416555404663
Iteration 362, loss: 1.2998642921447754
Iteration 363, loss: 1.1329289674758911
Iteration 364, loss: 1.213074803352356
Iteration 365, loss: 1.3137352466583252
Iteration 366, loss: 1.3136446475982666
Iteration 367, loss: 1.229452133178711
Iteration 368, loss: 1.2678298950195312
Iteration 369, loss: 1.215907335281372
Iteration 370, loss: 1.2027510404586792
Iteration 371, loss: 1.221210241317749
Iteration 372, loss: 1.246789813041687
Iteration 373, loss: 1.1423656940460205
Iteration 374, loss: 1.2651728391647339
Iteration 375, loss: 1.3038928508758545
Iteration 376, loss: 1.223472237586975
Iteration 377, loss: 1.213516116142273
Iteration 378, loss: 1.281554937362671
Iteration 379, loss: 1.2632198333740234
Iteration 380, loss: 1.3073925971984863
Iteration 381, loss: 1.2743185758590698
Iteration 382, loss: 1.2405496835708618
Iteration 383, loss: 1.2332203388214111
Iteration 384, loss: 1.199273943901062
Iteration 385, loss: 1.3453121185302734
Iteration 386, loss: 1.292420744895935
Iteration 387, loss: 1.2883355617523193
Iteration 388, loss: 1.1670379638671875
Iteration 389, loss: 1.2568199634552002
Iteration 390, loss: 1.216888666152954
Iteration 391, loss: 1.1993006467819214
Iteration 392, loss: 1.2386964559555054
Iteration 393, loss: 1.2619291543960571
Iteration 394, loss: 1.2043964862823486
Iteration 395, loss: 1.2396597862243652
Iteration 396, loss: 1.2249047756195068
Iteration 397, loss: 1.1915950775146484
Iteration 398, loss: 1.2272497415542603
Iteration 399, loss: 1.2738019227981567
Iteration 400, loss: 1.2891173362731934
Iteration 401, loss: 1.2573025226593018
Iteration 402, loss: 1.2022366523742676
Iteration 403, loss: 1.2872544527053833
Iteration 404, loss: 1.3057944774627686
```

```
Iteration 405, loss: 1.2463200092315674
Iteration 406, loss: 1.1677063703536987
Iteration 407, loss: 1.201696753501892
Iteration 408, loss: 1.2474784851074219
Iteration 409, loss: 1.285955548286438
Iteration 410, loss: 1.2496553659439087
Iteration 411, loss: 1.2703633308410645
Iteration 412, loss: 1.3326326608657837
Iteration 413, loss: 1.1876964569091797
Iteration 414, loss: 1.2310365438461304
Iteration 415, loss: 1.2417199611663818
Iteration 416, loss: 1.277040719985962
Iteration 417, loss: 1.2313684225082397
Iteration 418, loss: 1.1991915702819824
Iteration 419, loss: 1.250076413154602
Iteration 420, loss: 1.2285569906234741
Iteration 421, loss: 1.4355894327163696
Iteration 422, loss: 1.1647580862045288
Iteration 423, loss: 1.09876549243927
Iteration 424, loss: 1.3256151676177979
Iteration 425, loss: 1.2560869455337524
Iteration 426, loss: 1.2005255222320557
Iteration 427, loss: 1.2531933784484863
Iteration 428, loss: 1.1712459325790405
Iteration 429, loss: 1.236831784248352
Iteration 430, loss: 1.2289022207260132
Iteration 431, loss: 1.2934201955795288
Iteration 432, loss: 1.1443302631378174
Iteration 433, loss: 1.209831953048706
Iteration 434, loss: 1.1456509828567505
Iteration 435, loss: 1.176186203956604
Iteration 436, loss: 1.195239543914795
Iteration 437, loss: 1.2184364795684814
Iteration 438, loss: 1.1888279914855957
Iteration 439, loss: 1.2216317653656006
Iteration 440, loss: 1.218000888824463
Iteration 441, loss: 1.2025439739227295
Iteration 442, loss: 1.1504871845245361
Iteration 443, loss: 1.1765468120574951
Iteration 444, loss: 1.2422122955322266
Iteration 445, loss: 1.2331159114837646
Iteration 446, loss: 1.1906421184539795
Iteration 447, loss: 1.1590174436569214
Iteration 448, loss: 1.3289361000061035
Iteration 449, loss: 1.1316008567810059
Iteration 450, loss: 1.1879737377166748
Iteration 451, loss: 1.134770393371582
Iteration 452, loss: 1.1800544261932373
Iteration 453, loss: 1.1005626916885376
Iteration 454, loss: 1.1832157373428345
Iteration 455, loss: 1.1980338096618652
Iteration 456, loss: 1.1514993906021118
Iteration 457, loss: 1.238158941268921
Iteration 458, loss: 1.1752680540084839
Iteration 459, loss: 1.2299987077713013
Iteration 460, loss: 1.2835463285446167
Iteration 461, loss: 1.2638238668441772
Iteration 462, loss: 1.1874512434005737
Iteration 463, loss: 1.188206434249878
Iteration 464, loss: 1.1567931175231934
Iteration 465, loss: 1.302672028541565
Iteration 466, loss: 1.2716012001037598
Iteration 467, loss: 1.2016539573669434
Iteration 468, loss: 1.2897499799728394
Iteration 469, loss: 1.022905945777893
Iteration 470, loss: 1.1801130771636963
Iteration 471, loss: 1.2114723920822144
Iteration 472, loss: 1.198371410369873
Iteration 473, loss: 1.1692334413528442
Iteration 474, loss: 1.1594656705856323
Iteration 475, loss: 1.2128318548202515
Iteration 476, loss: 1.2652055025100708
Iteration 477, loss: 1.2267392873764038
Iteration 478, loss: 1.1509966850280762
Iteration 479, loss: 1.2600725889205933
Iteration 480, loss: 1.24248206615448
Iteration 481, loss: 1.2881144285202026
```

```
Iteration 482, loss: 1.1912283897399902
Iteration 483, loss: 1.139483094215393
Iteration 484, loss: 1.25863516330719
Iteration 485, loss: 1.1783066987991333
Iteration 486, loss: 1.2227368354797363
Iteration 487, loss: 1.1790791749954224
Iteration 488, loss: 1.1772751808166504
Iteration 489, loss: 1.1577999591827393
Iteration 490, loss: 1.1619477272033691
Iteration 491, loss: 1.1503453254699707
Iteration 492, loss: 1.2309956550598145
Iteration 493, loss: 1.2157814502716064
Iteration 494, loss: 1.1368249654769897
Iteration 495, loss: 1.1912105083465576
Iteration 496, loss: 1.1748156547546387
Iteration 497, loss: 1.1074984073638916
Iteration 498, loss: 1.1856106519699097
Iteration 499, loss: 1.2359354496002197
Iteration 500, loss: 1.1380553245544434
Iteration 501, loss: 1.2312742471694946
Iteration 502, loss: 1.2811795473098755
Iteration 503, loss: 1.2787009477615356
Iteration 504, loss: 1.2306426763534546
Iteration 505, loss: 1.2005608081817627
Iteration 506, loss: 1.245424747467041
Iteration 507, loss: 1.1476962566375732
Iteration 508, loss: 1.1752476692199707
Iteration 509, loss: 1.1507408618927002
Iteration 510, loss: 1.2146306037902832
Iteration 511, loss: 1.1867847442626953
Iteration 512, loss: 1.248974084854126
Iteration 513, loss: 1.2054009437561035
Iteration 514, loss: 1.2284390926361084
Iteration 515, loss: 1.132043480873108
Iteration 516, loss: 1.0807965993881226
Iteration 517, loss: 1.082429051399231
Iteration 518, loss: 1.099487543106079
Iteration 519, loss: 1.226918339729309
Iteration 520, loss: 1.144532561302185
Iteration 521, loss: 1.137030839920044
Iteration 522, loss: 1.0933712720870972
Iteration 523, loss: 1.0934659242630005
Iteration 524, loss: 1.1370049715042114
Iteration 525, loss: 1.2317042350769043
Iteration 526, loss: 1.0584871768951416
Iteration 527, loss: 1.130650281906128
Iteration 528, loss: 1.0983481407165527
Iteration 529, loss: 1.227114200592041
Iteration 530, loss: 1.2263531684875488
Iteration 531, loss: 1.1510754823684692
Iteration 532, loss: 1.1763157844543457
Iteration 533, loss: 1.1742242574691772
Iteration 534, loss: 1.1602730751037598
Iteration 535, loss: 1.2499819993972778
Iteration 536, loss: 1.2282023429870605
Iteration 537, loss: 1.1910545825958252
Iteration 538, loss: 1.1160292625427246
Iteration 539, loss: 1.1607310771942139
Iteration 540, loss: 1.2489686012268066
Iteration 541, loss: 1.180999517440796
Iteration 542, loss: 1.1977559328079224
Iteration 543, loss: 1.1911509037017822
Iteration 544, loss: 1.1965081691741943
Iteration 545, loss: 1.14201819896698
Iteration 546, loss: 1.2223944664001465
Iteration 547, loss: 1.0800741910934448
Iteration 548, loss: 1.2120871543884277
Iteration 549, loss: 1.1197959184646606
Iteration 550, loss: 1.1803340911865234
Iteration 551, loss: 1.204067349433899
Iteration 552, loss: 1.1378097534179688
Iteration 553, loss: 1.094783902168274
Iteration 554, loss: 1.1585633754730225
Iteration 555, loss: 1.0838991403579712
Iteration 556, loss: 1.1542279720306396
Iteration 557, loss: 1.1775779724121094
Iteration 558, loss: 1.188499093055725
```

```
Iteration 559, loss: 1.2837727069854736
Iteration 560, loss: 1.1456458568572998
Iteration 561, loss: 1.0714190006256104
Iteration 562, loss: 1.1696140766143799
Iteration 563, loss: 1.2602477073669434
Iteration 564, loss: 1.2070001363754272
Iteration 565, loss: 1.1865346431732178
Iteration 566, loss: 1.1715855598449707
Iteration 567, loss: 1.1546927690505981
Iteration 568, loss: 1.1303837299346924
Iteration 569, loss: 1.2513941526412964
Iteration 570, loss: 1.0925962924957275
Iteration 571, loss: 1.14694082736969
Iteration 572, loss: 1.143066167831421
Iteration 573, loss: 1.059761881828308
Iteration 574, loss: 1.0485848188400269
Iteration 575, loss: 1.1197459697723389
Iteration 576, loss: 1.0649296045303345
Iteration 577, loss: 1.0986480712890625
Iteration 578, loss: 1.1344834566116333
Iteration 579, loss: 1.1097103357315063
Iteration 580, loss: 1.1254907846450806
Iteration 581, loss: 1.0777817964553833
Iteration 582, loss: 1.0916781425476074
Iteration 583, loss: 1.133601427078247
Iteration 584, loss: 1.128501296043396
Iteration 585, loss: 1.1115416288375854
Iteration 586, loss: 1.1413185596466064
Iteration 587, loss: 1.2547156810760498
Iteration 588, loss: 1.0876965522766113
Iteration 589, loss: 1.2354023456573486
Iteration 590, loss: 1.102877140045166
Iteration 591, loss: 1.1341955661773682
Iteration 592, loss: 1.246012568473816
Iteration 593, loss: 1.1398494243621826
Iteration 594, loss: 1.0681016445159912
Iteration 595, loss: 1.0264866352081299
Iteration 596, loss: 1.145768642425537
Iteration 597, loss: 1.1331347227096558
Iteration 598, loss: 1.2220896482467651
Iteration 599, loss: 1.1068634986877441
Iteration 600, loss: 1.1641433238983154
Iteration 601, loss: 1.0650994777679443
Iteration 602, loss: 1.1248278617858887
Iteration 603, loss: 1.1243501901626587
Iteration 604, loss: 1.0687521696090698
Iteration 605, loss: 1.0880342721939087
Iteration 606, loss: 1.0793927907943726
Iteration 607, loss: 1.0619937181472778
Iteration 608, loss: 1.0482159852981567
Iteration 609, loss: 1.1495941877365112
Iteration 610, loss: 1.2394084930419922
Iteration 611, loss: 1.0513943433761597
Iteration 612, loss: 1.0749685764312744
Iteration 613, loss: 1.0721189975738525
Iteration 614, loss: 1.1197625398635864
Iteration 615, loss: 1.2119406461715698
Iteration 616, loss: 1.1667624711990356
Iteration 617, loss: 1.0974911451339722
Iteration 618, loss: 1.0978254079818726
Iteration 619, loss: 1.1192200183868408
Iteration 620, loss: 1.0914579629898071
Iteration 621, loss: 1.1539303064346313
Iteration 622, loss: 1.1021583080291748
Iteration 623, loss: 1.1055092811584473
Iteration 624, loss: 1.1514981985092163
Iteration 625, loss: 1.1169841289520264
Iteration 626, loss: 1.1801931858062744
Iteration 627, loss: 1.1871057748794556
Iteration 628, loss: 1.1262463331222534
Iteration 629, loss: 1.0346852540969849
Iteration 630, loss: 1.1666884422302246
Iteration 631, loss: 1.087511658668518
Iteration 632, loss: 1.1020256280899048
Iteration 633, loss: 1.1075620651245117
Iteration 634, loss: 1.0862973928451538
Iteration 635, loss: 1.0393834114074707
```

```
Iteration 636, loss: 1.1174603700637817
Iteration 637, loss: 1.1086238622665405
Iteration 638, loss: 1.1947872638702393
Iteration 639, loss: 1.1232150793075562
Iteration 640, loss: 1.0996606349945068
Iteration 641, loss: 1.228113055229187
Iteration 642, loss: 1.1069649457931519
Iteration 643, loss: 1.161025881767273
Iteration 644, loss: 1.1276618242263794
Iteration 645, loss: 1.1249877214431763
Iteration 646, loss: 1.1531163454055786
Iteration 647, loss: 1.0202864408493042
Iteration 648, loss: 1.1461710929870605
Iteration 649, loss: 1.0658353567123413
Iteration 650, loss: 1.0380897521972656
Iteration 651, loss: 1.1715261936187744
Iteration 652, loss: 1.0875729322433472
Iteration 653, loss: 1.1507939100265503
Iteration 654, loss: 1.1729927062988281
Iteration 655, loss: 1.0821788311004639
Iteration 656, loss: 1.0881692171096802
Iteration 657, loss: 1.1077637672424316
Iteration 658, loss: 1.1631710529327393
Iteration 659, loss: 1.1310889720916748
Iteration 660, loss: 1.0258698463439941
Iteration 661, loss: 1.0848933458328247
Iteration 662, loss: 1.1658432483673096
Iteration 663, loss: 1.005387306213379
Iteration 664, loss: 1.0858817100524902
Iteration 665, loss: 1.2114609479904175
Iteration 666, loss: 1.0655062198638916
Iteration 667, loss: 1.1105979681015015
Iteration 668, loss: 1.0834596157073975
Iteration 669, loss: 1.1350804567337036
Iteration 670, loss: 1.118699073791504
Iteration 671, loss: 1.0947967767715454
Iteration 672, loss: 1.095018982887268
Iteration 673, loss: 1.170716404914856
Iteration 674, loss: 1.0668412446975708
Iteration 675, loss: 1.090829849243164
Iteration 676, loss: 1.065618634223938
Iteration 677, loss: 1.0573941469192505
Iteration 678, loss: 1.1466177701950073
Iteration 679, loss: 1.1358447074890137
Iteration 680, loss: 1.1356970071792603
Iteration 681, loss: 1.070511817932129
Iteration 682, loss: 1.0617791414260864
Iteration 683, loss: 1.06786048412323
Iteration 684, loss: 1.035860300064087
Iteration 685, loss: 1.1536246538162231
Iteration 686, loss: 1.094815731048584
Iteration 687, loss: 1.1103852987289429
Iteration 688, loss: 1.1222221851348877
Iteration 689, loss: 1.0717021226882935
Iteration 690, loss: 1.0463733673095703
Iteration 691, loss: 1.1279406547546387
Iteration 692, loss: 1.0794492959976196
Iteration 693, loss: 1.1442692279815674
Iteration 694, loss: 1.0799674987792969
Iteration 695, loss: 1.1710176467895508
Iteration 696, loss: 1.2267630100250244
Iteration 697, loss: 1.0486440658569336
Iteration 698, loss: 1.1141186952590942
Iteration 699, loss: 1.1347706317901611
Iteration 700, loss: 1.2521637678146362
Iteration 701, loss: 1.113462209701538
Iteration 702, loss: 1.112152338027954
Iteration 703, loss: 1.1280207633972168
Iteration 704, loss: 1.110862135887146
Iteration 705, loss: 1.0741798877716064
Iteration 706, loss: 1.075499176979065
Iteration 707, loss: 1.0556889772415161
Iteration 708, loss: 1.108114242553711
Iteration 709, loss: 1.1104941368103027
Iteration 710, loss: 1.0548920631408691
Iteration 711, loss: 1.0940555334091187
Iteration 712, loss: 1.0081497430801392
```

```
Iteration 713, loss: 1.0403008460998535
Iteration 714, loss: 1.1013959646224976
Iteration 715, loss: 1.0781153440475464
Iteration 716, loss: 1.0408351421356201
Iteration 717, loss: 1.0657193660736084
Iteration 718, loss: 1.1158350706100464
Iteration 719, loss: 1.0708438158035278
Iteration 720, loss: 1.0585837364196777
Iteration 721, loss: 1.0707513093948364
Iteration 722, loss: 1.0323574542999268
Iteration 723, loss: 1.03209388256073
Iteration 724, loss: 1.0861375331878662
Iteration 725, loss: 1.0454503297805786
Iteration 726, loss: 1.0927516222000122
Iteration 727, loss: 1.1767109632492065
Iteration 728, loss: 1.1142809391021729
Iteration 729, loss: 1.092664122581482
Iteration 730, loss: 1.1476093530654907
Iteration 731, loss: 1.1186175346374512
Iteration 732, loss: 1.0891356468200684
Iteration 733, loss: 1.0761083364486694
Iteration 734, loss: 1.0955792665481567
Iteration 735, loss: 1.0448228120803833
Iteration 736, loss: 1.1345463991165161
Iteration 737, loss: 1.1052308082580566
Iteration 738, loss: 1.143764853477478
Iteration 739, loss: 1.0579222440719604
Iteration 740, loss: 1.080952525138855
Iteration 741, loss: 1.106429934501648
Iteration 742, loss: 1.1539363861083984
Iteration 743, loss: 1.108488917350769
Iteration 744, loss: 1.0262309312820435
Iteration 745, loss: 1.095263957977295
Iteration 746, loss: 1.1171326637268066
Iteration 747, loss: 1.1676369905471802
Iteration 748, loss: 1.0881270170211792
Iteration 749, loss: 1.0412163734436035
Iteration 750, loss: 1.0865135192871094
Iteration 751, loss: 1.0227241516113281
Iteration 752, loss: 1.1099351644515991
Iteration 753, loss: 1.1338355541229248
Iteration 754, loss: 1.0033769607543945
Iteration 755, loss: 1.0350732803344727
Iteration 756, loss: 1.1109051704406738
Iteration 757, loss: 1.045902967453003
Iteration 758, loss: 1.0522750616073608
Iteration 759, loss: 1.0837304592132568
Iteration 760, loss: 0.989939272403717
Iteration 761, loss: 1.0992544889450073
Iteration 762, loss: 1.0374137163162231
Iteration 763, loss: 1.1235358715057373
Iteration 764, loss: 1.0695466995239258
Iteration 765, loss: 1.0056312084197998
Iteration 766, loss: 0.9375207424163818
Iteration 767, loss: 1.078482747077942
Iteration 768, loss: 1.0676639080047607
Iteration 769, loss: 1.038724422454834
Iteration 770, loss: 1.0594605207443237
Iteration 771, loss: 0.9725439548492432
Iteration 772, loss: 1.0932565927505493
Iteration 773, loss: 1.0177327394485474
Iteration 774, loss: 1.0883033275604248
Iteration 775, loss: 1.0929051637649536
Iteration 776, loss: 1.0697842836380005
Iteration 777, loss: 1.1253612041473389
Iteration 778, loss: 1.095884919166565
Iteration 779, loss: 1.0007946491241455
Iteration 780, loss: 1.1104931831359863
Iteration 781, loss: 1.0235929489135742
Iteration 782, loss: 0.9616188406944275
Iteration 783, loss: 0.9900128841400146
Iteration 784, loss: 1.0114974975585938
Iteration 785, loss: 0.9759792685508728
Iteration 786, loss: 1.0039387941360474
Iteration 787, loss: 1.0641041994094849
Iteration 788, loss: 1.059782862663269
Iteration 789, loss: 1.1464123725891113
```

```
Iteration 790, loss: 1.033055305480957
Iteration 791, loss: 1.0516890287399292
Iteration 792, loss: 1.0324759483337402
Iteration 793, loss: 1.031847357749939
Iteration 794, loss: 0.96539306640625
Iteration 795, loss: 1.037819743156433
Iteration 796, loss: 1.1315970420837402
Iteration 797, loss: 0.9769617915153503
Iteration 798, loss: 1.0127092599868774
Iteration 799, loss: 1.0224846601486206
Iteration 800, loss: 1.000767707824707
Iteration 801, loss: 0.9691574573516846
Iteration 802, loss: 1.1102871894836426
Iteration 803, loss: 0.9069645404815674
Iteration 804, loss: 1.0200625658035278
Iteration 805, loss: 1.0427144765853882
Iteration 806, loss: 0.9803057312965393
Iteration 807, loss: 0.9992090463638306
Iteration 808, loss: 1.0651081800460815
Iteration 809, loss: 1.0645346641540527
Iteration 810, loss: 1.0867940187454224
Iteration 811, loss: 1.0010786056518555
Iteration 812, loss: 1.0726287364959717
Iteration 813, loss: 0.9574878215789795
Iteration 814, loss: 1.0588793754577637
Iteration 815, loss: 1.0765992403030396
Iteration 816, loss: 0.9646514058113098
Iteration 817, loss: 1.0525171756744385
Iteration 818, loss: 1.0419856309890747
Iteration 819, loss: 1.071085810661316
Iteration 820, loss: 1.0462937355041504
Iteration 821, loss: 1.136576533317566
Iteration 822, loss: 1.1033360958099365
Iteration 823, loss: 0.9792537093162537
Iteration 824, loss: 1.0599274635314941
Iteration 825, loss: 0.9919744729995728
Iteration 826, loss: 1.0399147272109985
Iteration 827, loss: 1.050578236579895
Iteration 828, loss: 0.9962809681892395
Iteration 829, loss: 1.046187162399292
Iteration 830, loss: 0.9891966581344604
Iteration 831, loss: 1.1353589296340942
Iteration 832, loss: 0.9546563625335693
Iteration 833, loss: 0.9818232655525208
Iteration 834, loss: 0.985697865486145
Iteration 835, loss: 1.1572197675704956
Iteration 836, loss: 0.983454704284668
Iteration 837, loss: 1.0309392213821411
Iteration 838, loss: 0.9922838807106018
Iteration 839, loss: 0.9989707469940186
Iteration 840, loss: 1.0082650184631348
Iteration 841, loss: 1.0596593618392944
Iteration 842, loss: 1.004831075668335
Iteration 843, loss: 0.9747257232666016
Iteration 844, loss: 1.002037763595581
Iteration 845, loss: 0.9866189956665039
Iteration 846, loss: 0.9943544864654541
Iteration 847, loss: 1.0333805084228516
Iteration 848, loss: 1.137121319770813
Iteration 849, loss: 1.1088879108428955
Iteration 850, loss: 0.9335642457008362
Iteration 851, loss: 1.0199995040893555
Iteration 852, loss: 1.0189582109451294
Iteration 853, loss: 1.0619728565216064
Iteration 854, loss: 0.984007716178894
Iteration 855, loss: 0.9985535740852356
Iteration 856, loss: 0.9908421039581299
Iteration 857, loss: 1.1250654458999634
Iteration 858, loss: 1.0674293041229248
Iteration 859, loss: 1.0037264823913574
Iteration 860, loss: 1.0948410034179688
Iteration 861, loss: 1.0418384075164795
Iteration 862, loss: 0.9903180003166199
Iteration 863, loss: 1.0363224744796753
Iteration 864, loss: 0.9704148173332214
Iteration 865, loss: 1.0656566619873047
Iteration 866, loss: 1.0128123760223389
```

```
Iteration 867, loss: 1.0257151126861572
Iteration 868, loss: 1.0159071683883667
Iteration 869, loss: 0.9614028930664062
Iteration 870, loss: 1.0109275579452515
Iteration 871, loss: 1.0173031091690063
Iteration 872, loss: 0.9866231679916382
Iteration 873, loss: 0.9947823882102966
Iteration 874, loss: 1.032542109489441
Iteration 875, loss: 0.9715680480003357
Iteration 876, loss: 0.9265117049217224
Iteration 877, loss: 0.9150563478469849
Iteration 878, loss: 1.1126152276992798
Iteration 879, loss: 0.9948216676712036
Iteration 880, loss: 0.9694774150848389
Iteration 881, loss: 0.9983922243118286
Iteration 882, loss: 0.9850447773933411
Iteration 883, loss: 0.9667062163352966
Iteration 884, loss: 1.014098882675171
Iteration 885, loss: 1.007206916809082
Iteration 886, loss: 1.0747041702270508
Iteration 887, loss: 1.0025367736816406
Iteration 888, loss: 0.937017560005188
Iteration 889, loss: 0.9806214570999146
Iteration 890, loss: 1.008898138999939
Iteration 891, loss: 1.0559779405593872
Iteration 892, loss: 0.9865044951438904
Iteration 893, loss: 1.008971095085144
Iteration 894, loss: 1.0506705045700073
Iteration 895, loss: 0.9891240000724792
Iteration 896, loss: 0.9714454412460327
Iteration 897, loss: 0.9837590456008911
Iteration 898, loss: 1.014839768409729
Iteration 899, loss: 0.9376802444458008
Iteration 900, loss: 1.0147075653076172
Iteration 901, loss: 0.9912852048873901
Iteration 902, loss: 0.986790657043457
Iteration 903, loss: 0.8947153687477112
Iteration 904, loss: 1.0114187002182007
Iteration 905, loss: 1.0754412412643433
Iteration 906, loss: 1.005761981010437
Iteration 907, loss: 1.0060688257217407
Iteration 908, loss: 1.0203347206115723
Iteration 909, loss: 0.9941647052764893
Iteration 910, loss: 1.0148733854293823
Iteration 911, loss: 0.9642781019210815
Iteration 912, loss: 0.9867672324180603
Iteration 913, loss: 0.889981210231781
Iteration 914, loss: 0.8923453092575073
Iteration 915, loss: 1.013697624206543
Iteration 916, loss: 0.9909569621086121
Iteration 917, loss: 0.8884525895118713
Iteration 918, loss: 0.9814977049827576
Iteration 919, loss: 1.006693720817566
Iteration 920, loss: 0.9485213160514832
Iteration 921, loss: 0.9991140961647034
Iteration 922, loss: 0.973479688167572
Iteration 923, loss: 0.9836858510971069
Iteration 924, loss: 0.9822402596473694
Iteration 925, loss: 1.0495641231536865
Iteration 926, loss: 0.9680442810058594
Iteration 927, loss: 0.9582633972167969
Iteration 928, loss: 1.007190465927124
Iteration 929, loss: 0.83547043800354
Iteration 930, loss: 1.0606186389923096
Iteration 931, loss: 0.9660081267356873
Iteration 932, loss: 0.925424337387085
Iteration 933, loss: 0.9820288419723511
Iteration 934, loss: 1.0433297157287598
Iteration 935, loss: 0.9466943144798279
Iteration 936, loss: 0.9892946481704712
Iteration 937, loss: 0.9938830137252808
Iteration 938, loss: 0.9421875476837158
Iteration 939, loss: 1.0117024183273315
Iteration 940, loss: 0.9662874937057495
Iteration 941, loss: 0.9688053727149963
Iteration 942, loss: 0.9305317401885986
Iteration 943, loss: 1.0092110633850098
```

```
Iteration 944, loss: 1.0608879327774048
Iteration 945, loss: 1.0059542655944824
Iteration 946, loss: 0.9651960730552673
Iteration 947, loss: 0.9539811015129089
Iteration 948, loss: 0.9463104605674744
Iteration 949, loss: 0.8627492189407349
Iteration 950, loss: 0.9553599953651428
Iteration 951, loss: 0.9486192464828491
Iteration 952, loss: 0.9403049945831299
Iteration 953, loss: 0.9425152540206909
Iteration 954, loss: 0.9703810214996338
Iteration 955, loss: 0.9836364388465881
Iteration 956, loss: 0.9511006474494934
Iteration 957, loss: 0.9630511999130249
Iteration 958, loss: 0.8860827088356018
Iteration 959, loss: 0.9539110660552979
Iteration 960, loss: 0.9940687417984009
Iteration 961, loss: 0.8565061092376709
Iteration 962, loss: 0.9694800972938538
Iteration 963, loss: 1.0459938049316406
Iteration 964, loss: 0.9500558376312256
Iteration 965, loss: 0.9824050068855286
Iteration 966, loss: 1.007138967514038
Iteration 967, loss: 0.9338066577911377
Iteration 968, loss: 0.9045533537864685
Iteration 969, loss: 0.9914922118186951
Iteration 970, loss: 1.0180631875991821
Iteration 971, loss: 0.9367100596427917
Iteration 972, loss: 0.9499215483665466
Iteration 973, loss: 0.9956591129302979
Iteration 974, loss: 0.9441712498664856
Iteration 975, loss: 0.9560216069221497
Iteration 976, loss: 0.9894440770149231
Iteration 977, loss: 0.9818942546844482
Iteration 978, loss: 0.9236741662025452
Iteration 979, loss: 0.9459839463233948
Iteration 980, loss: 0.864730715751648
Iteration 981, loss: 0.9284086227416992
Iteration 982, loss: 0.8822657465934753
Iteration 983, loss: 0.8922047019004822
Iteration 984, loss: 0.9243815541267395
Iteration 985, loss: 0.9781865477561951
Iteration 986, loss: 0.9680315256118774
Iteration 987, loss: 0.9671944975852966
Iteration 988, loss: 0.9228289723396301
Iteration 989, loss: 0.9196102619171143
Iteration 990, loss: 0.9619083404541016
Iteration 991, loss: 1.0114213228225708
Iteration 992, loss: 0.9738110303878784
Iteration 993, loss: 0.9120689630508423
Iteration 994, loss: 0.90351402759552
Iteration 995, loss: 0.9014826416969299
Iteration 996, loss: 0.9408146142959595
Iteration 997, loss: 1.1411586999893188
Iteration 998, loss: 0.9159939289093018
Iteration 999, loss: 0.8669404983520508
End of epoch 5, loss: 2.1984479427337646
Validation loss: 1.4300923347473145
Iteration 0, loss: 1.0218559503555298
Iteration 1, loss: 0.9072954654693604
Iteration 2, loss: 0.889885663986206
Iteration 3, loss: 0.8973697423934937
Iteration 4, loss: 0.962757408618927
Iteration 5, loss: 1.003585696220398
Iteration 6, loss: 0.9648170471191406
Iteration 7, loss: 0.8942257761955261
Iteration 8, loss: 0.9265689253807068
Iteration 9, loss: 0.8994598388671875
Iteration 10, loss: 0.9685699939727783
Iteration 11, loss: 0.9619287848472595
Iteration 12, loss: 1.0070717334747314
Iteration 13, loss: 0.9200350046157837
Iteration 14, loss: 1.0055615901947021
Iteration 15, loss: 0.9440730810165405
Iteration 16, loss: 0.9829254746437073
Iteration 17, loss: 0.9933240413665771
Iteration 18, loss: 1.0272390842437744
```

```
Iteration 19, loss: 0.9790473580360413
Iteration 20, loss: 1.024342656135559
Iteration 21, loss: 1.0100280046463013
Iteration 22, loss: 0.9830559492111206
Iteration 23, loss: 0.8786798119544983
Iteration 24, loss: 0.8832809925079346
Iteration 25, loss: 1.0188575983047485
Iteration 26, loss: 1.024062156677246
Iteration 27, loss: 1.0134778022766113
Iteration 28, loss: 0.9288766384124756
Iteration 29, loss: 0.9520953297615051
Iteration 30, loss: 0.9542974829673767
Iteration 31, loss: 0.9148282408714294
Iteration 32, loss: 0.9214850068092346
Iteration 33, loss: 1.0167758464813232
Iteration 34, loss: 0.9834414124488831
Iteration 35, loss: 0.9154990315437317
Iteration 36, loss: 0.897251546382904
Iteration 37, loss: 0.8554385900497437
Iteration 38, loss: 0.9193137288093567
Iteration 39, loss: 0.9528558254241943
Iteration 40, loss: 0.9410148859024048
Iteration 41, loss: 1.0019587278366089
Iteration 42, loss: 0.8760501146316528
Iteration 43, loss: 0.9111246466636658
Iteration 44, loss: 0.8843404054641724
Iteration 45, loss: 1.0232337713241577
Iteration 46, loss: 0.9777405261993408
Iteration 47, loss: 0.9098526239395142
Iteration 48, loss: 0.9357474446296692
Iteration 49, loss: 0.9551281332969666
Iteration 50, loss: 0.9651819467544556
Iteration 51, loss: 0.9292752146720886
Iteration 52, loss: 0.921195924282074
Iteration 53, loss: 0.9435252547264099
Iteration 54, loss: 0.8728234767913818
Iteration 55, loss: 1.0499809980392456
Iteration 56, loss: 0.8203508853912354
Iteration 57, loss: 0.9703998565673828
Iteration 58, loss: 1.0189473628997803
Iteration 59, loss: 0.8647748231887817
Iteration 60, loss: 0.8718922138214111
Iteration 61, loss: 1.0282715559005737
Iteration 62, loss: 0.9866800904273987
Iteration 63, loss: 0.9342921376228333
Iteration 64, loss: 0.8812845945358276
Iteration 65, loss: 0.959103524684906
Iteration 66, loss: 0.8857120275497437
Iteration 67, loss: 0.9995700120925903
Iteration 68, loss: 0.9256178140640259
Iteration 69, loss: 0.9898574352264404
Iteration 70, loss: 0.8806710839271545
Iteration 71, loss: 0.9817469716072083
Iteration 72, loss: 0.9009026885032654
Iteration 73, loss: 0.8771053552627563
Iteration 74, loss: 1.0495879650115967
Iteration 75, loss: 0.910538375377655
Iteration 76, loss: 0.9076036214828491
Iteration 77, loss: 0.9393948316574097
Iteration 78, loss: 0.9744197130203247
Iteration 79, loss: 0.8230503797531128
Iteration 80, loss: 0.9033164978027344
Iteration 81, loss: 0.8981243371963501
Iteration 82, loss: 0.8834357857704163
Iteration 83, loss: 0.9389496445655823
Iteration 84, loss: 0.9297940135002136
Iteration 85, loss: 0.866331160068512
Iteration 86, loss: 0.9052814245223999
Iteration 87, loss: 0.9253038763999939
Iteration 88, loss: 0.9552360773086548
Iteration 89, loss: 0.8957363367080688
Iteration 90, loss: 0.9213762283325195
Iteration 91, loss: 0.964959979057312
Iteration 92, loss: 0.8412952423095703
Iteration 93, loss: 0.9718412160873413
Iteration 94, loss: 0.8923993706703186
Iteration 95, loss: 0.9320775270462036
```

```
Iteration 96, loss: 0.8754532933235168
Iteration 97, loss: 0.8459824919700623
Iteration 98, loss: 0.9491638541221619
Iteration 99, loss: 0.9000946879386902
Iteration 100, loss: 0.8542159199714661
Iteration 101, loss: 0.8915119767189026
Iteration 102, loss: 0.8864759802818298
Iteration 103, loss: 0.8143819570541382
Iteration 104, loss: 1.002389669418335
Iteration 105, loss: 0.916386604309082
Iteration 106, loss: 0.9435262084007263
Iteration 107, loss: 0.8353556394577026
Iteration 108, loss: 0.8851874470710754
Iteration 109, loss: 1.022578477859497
Iteration 110, loss: 0.9594384431838989
Iteration 111, loss: 0.9755893349647522
Iteration 112, loss: 0.849632203578949
Iteration 113, loss: 0.9841350317001343
Iteration 114, loss: 0.8745803236961365
Iteration 115, loss: 0.9198826551437378
Iteration 116, loss: 1.0227468013763428
Iteration 117, loss: 0.9579765796661377
Iteration 118, loss: 0.927623987197876
Iteration 119, loss: 0.9014514088630676
Iteration 120, loss: 0.8557427525520325
Iteration 121, loss: 0.8305995464324951
Iteration 122, loss: 0.8607120513916016
Iteration 123, loss: 0.9177263975143433
Iteration 124, loss: 0.8770326375961304
Iteration 125, loss: 0.8800392746925354
Iteration 126, loss: 0.879586935043335
Iteration 127, loss: 0.842131495475769
Iteration 128, loss: 0.8657596111297607
Iteration 129, loss: 0.9327220320701599
Iteration 130, loss: 0.8382186889648438
Iteration 131, loss: 0.8800725340843201
Iteration 132, loss: 0.9313675165176392
Iteration 133, loss: 0.8372120261192322
Iteration 134, loss: 0.8849490284919739
Iteration 135, loss: 0.8319264650344849
Iteration 136, loss: 0.8309022188186646
Iteration 137, loss: 0.8278471231460571
Iteration 138, loss: 0.9034378528594971
Iteration 139, loss: 0.8830604553222656
Iteration 140, loss: 0.8267036080360413
Iteration 141, loss: 0.8912827372550964
Iteration 142, loss: 0.9639756679534912
Iteration 143, loss: 0.8459849953651428
Iteration 144, loss: 0.8278381824493408
Iteration 145, loss: 0.9027568101882935
Iteration 146, loss: 0.8138877153396606
Iteration 147, loss: 0.8994622826576233
Iteration 148, loss: 0.9521878957748413
Iteration 149, loss: 0.948165774345398
Iteration 150, loss: 0.825919508934021
Iteration 151, loss: 0.9141203761100769
Iteration 152, loss: 0.8403248190879822
Iteration 153, loss: 0.92961186170578
Iteration 154, loss: 0.9220077991485596
Iteration 155, loss: 0.9256978631019592
Iteration 156, loss: 0.9097658395767212
Iteration 157, loss: 0.8269466161727905
Iteration 158, loss: 0.7852127552032471
Iteration 159, loss: 0.7351556420326233
Iteration 160, loss: 0.9373709559440613
Iteration 161, loss: 0.8829022645950317
Iteration 162, loss: 0.8888434767723083
Iteration 163, loss: 0.882045328617096
Iteration 164, loss: 0.8350791335105896
Iteration 165, loss: 0.8553070425987244
Iteration 166, loss: 0.8679037094116211
Iteration 167, loss: 0.9166418313980103
Iteration 168, loss: 0.8434067368507385
Iteration 169, loss: 0.9282661080360413
Iteration 170, loss: 0.8502291440963745
Iteration 171, loss: 1.030239224433899
Iteration 172, loss: 0.9046831130981445
```

```
Iteration 173, loss: 0.8558511734008789
Iteration 174, loss: 0.8481331467628479
Iteration 175, loss: 0.8635778427124023
Iteration 176, loss: 0.7575655579566956
Iteration 177, loss: 0.8058710098266602
Iteration 178, loss: 0.869042158126831
Iteration 179, loss: 0.7842137217521667
Iteration 180, loss: 0.8932709097862244
Iteration 181, loss: 0.9253327250480652
Iteration 182, loss: 0.8097310662269592
Iteration 183, loss: 0.8983021974563599
Iteration 184, loss: 0.8021138310432434
Iteration 185, loss: 0.8705583214759827
Iteration 186, loss: 0.8030154705047607
Iteration 187, loss: 0.9188892841339111
Iteration 188, loss: 0.8471412658691406
Iteration 189, loss: 0.9097527861595154
Iteration 190, loss: 0.8510033488273621
Iteration 191, loss: 0.8496848344802856
Iteration 192, loss: 0.8838393688201904
Iteration 193, loss: 0.9197355508804321
Iteration 194, loss: 0.8301395773887634
Iteration 195, loss: 0.8423118591308594
Iteration 196, loss: 0.9319328665733337
Iteration 197, loss: 0.8284026384353638
Iteration 198, loss: 0.8733347654342651
Iteration 199, loss: 0.7938292622566223
Iteration 200, loss: 0.8384371995925903
Iteration 201, loss: 0.8697353601455688
Iteration 202, loss: 0.8786749243736267
Iteration 203, loss: 0.8004320859909058
Iteration 204, loss: 0.8430666327476501
Iteration 205, loss: 0.9304803013801575
Iteration 206, loss: 0.9124780297279358
Iteration 207, loss: 0.8797286748886108
Iteration 208, loss: 0.8669863343238831
Iteration 209, loss: 0.7972903847694397
Iteration 210, loss: 0.8533755540847778
Iteration 211, loss: 0.9193031787872314
Iteration 212, loss: 0.9119314551353455
Iteration 213, loss: 0.8383331894874573
Iteration 214, loss: 0.9057743549346924
Iteration 215, loss: 0.799924910068512
Iteration 216, loss: 0.8311851024627686
Iteration 217, loss: 0.9053977131843567
Iteration 218, loss: 0.8461582660675049
Iteration 219, loss: 0.8327049016952515
Iteration 220, loss: 0.8389177918434143
Iteration 221, loss: 0.8519670963287354
Iteration 222, loss: 0.7870400547981262
Iteration 223, loss: 0.819198489189148
Iteration 224, loss: 0.9158393144607544
Iteration 225, loss: 0.982645571231842
Iteration 226, loss: 0.7698707580566406
Iteration 227, loss: 0.9211000204086304
Iteration 228, loss: 0.8516718149185181
Iteration 229, loss: 0.8753018975257874
Iteration 230, loss: 0.8160852193832397
Iteration 231, loss: 0.7740980386734009
Iteration 232, loss: 0.8211933970451355
Iteration 233, loss: 0.9024377465248108
Iteration 234, loss: 0.8059256076812744
Iteration 235, loss: 0.8345785737037659
Iteration 236, loss: 0.7970333695411682
Iteration 237, loss: 0.8043502569198608
Iteration 238, loss: 0.8375906348228455
Iteration 239, loss: 0.8186228275299072
Iteration 240, loss: 0.8060745000839233
Iteration 241, loss: 0.787233293056488
Iteration 242, loss: 0.9303598403930664
Iteration 243, loss: 0.8451740741729736
Iteration 244, loss: 0.7623157501220703
Iteration 245, loss: 0.720420777797699
Iteration 246, loss: 0.7508468627929688
Iteration 247, loss: 0.8174037933349609
Iteration 248, loss: 0.8054824471473694
Iteration 249, loss: 0.7605491876602173
```

```
Iteration 250, loss: 0.8778208494186401
Iteration 251, loss: 0.7874135375022888
Iteration 252, loss: 0.8541730642318726
Iteration 253, loss: 0.8151340484619141
Iteration 254, loss: 0.8474769592285156
Iteration 255, loss: 0.809559166431427
Iteration 256, loss: 0.9137589931488037
Iteration 257, loss: 0.8581850528717041
Iteration 258, loss: 0.8727610111236572
Iteration 259, loss: 0.8257952332496643
Iteration 260, loss: 0.7886568903923035
Iteration 261, loss: 0.841705322265625
Iteration 262, loss: 0.7750676274299622
Iteration 263, loss: 0.8366562724113464
Iteration 264, loss: 0.8227968811988831
Iteration 265, loss: 0.8624306321144104
Iteration 266, loss: 0.8899823427200317
Iteration 267, loss: 0.7937912344932556
Iteration 268, loss: 0.8373154401779175
Iteration 269, loss: 0.8341085910797119
Iteration 270, loss: 0.7303786873817444
Iteration 271, loss: 0.7898350954055786
Iteration 272, loss: 0.8895102739334106
Iteration 273, loss: 0.8389474153518677
Iteration 274, loss: 0.7612175345420837
Iteration 275, loss: 0.8445191979408264
Iteration 276, loss: 0.9233758449554443
Iteration 277, loss: 0.7417351603507996
Iteration 278, loss: 0.7735482454299927
Iteration 279, loss: 0.8390327095985413
Iteration 280, loss: 0.8469772934913635
Iteration 281, loss: 0.7893149256706238
Iteration 282, loss: 0.78995281457901
Iteration 283, loss: 0.8130956888198853
Iteration 284, loss: 0.8172838687896729
Iteration 285, loss: 0.8090448975563049
Iteration 286, loss: 0.95307856798172
Iteration 287, loss: 0.7950757741928101
Iteration 288, loss: 0.8163489103317261
Iteration 289, loss: 0.8075975775718689
Iteration 290, loss: 0.7537503838539124
Iteration 291, loss: 0.8341715335845947
Iteration 292, loss: 0.8572565913200378
Iteration 293, loss: 0.7615233063697815
Iteration 294, loss: 0.794562041759491
Iteration 295, loss: 0.8047879934310913
Iteration 296, loss: 0.8168977499008179
Iteration 297, loss: 0.8731963634490967
Iteration 298, loss: 0.7930285334587097
Iteration 299, loss: 0.7886051535606384
Iteration 300, loss: 0.792873740196228
Iteration 301, loss: 0.8532399535179138
Iteration 302, loss: 0.8077738881111145
Iteration 303, loss: 0.8281524181365967
Iteration 304, loss: 0.8172215223312378
Iteration 305, loss: 0.7208656072616577
Iteration 306, loss: 0.8698428273200989
Iteration 307, loss: 0.7811132669448853
Iteration 308, loss: 0.829380452632904
Iteration 309, loss: 0.8314145803451538
Iteration 310, loss: 0.7412247061729431
Iteration 311, loss: 0.8345276117324829
Iteration 312, loss: 0.8445759415626526
Iteration 313, loss: 0.8073834776878357
Iteration 314, loss: 0.7548609375953674
Iteration 315, loss: 0.8667017817497253
Iteration 316, loss: 0.8378124237060547
Iteration 317, loss: 0.7975214719772339
Iteration 318, loss: 0.7231908440589905
Iteration 319, loss: 0.786652684211731
Iteration 320, loss: 0.8041003942489624
Iteration 321, loss: 0.7442213892936707
Iteration 322, loss: 0.8028680086135864
Iteration 323, loss: 0.7452332973480225
Iteration 324, loss: 0.919165849685669
Iteration 325, loss: 0.780604362487793
Iteration 326, loss: 0.7862886190414429
```

```
Iteration 326, loss: 0.7[illegible]
Iteration 327, loss: 0.7037765979766846
Iteration 328, loss: 0.7629006505012512
Iteration 329, loss: 0.8739849328994751
Iteration 330, loss: 0.8015835285186768
Iteration 331, loss: 0.818517804145813
Iteration 332, loss: 0.7004196643829346
Iteration 333, loss: 0.804462194442749
Iteration 334, loss: 0.7806278467178345
Iteration 335, loss: 0.8060370087623596
Iteration 336, loss: 0.8116986155509949
Iteration 337, loss: 0.8363903760910034
Iteration 338, loss: 0.8230047821998596
Iteration 339, loss: 0.8685699701309204
Iteration 340, loss: 0.7926853895187378
Iteration 341, loss: 0.8410834074020386
Iteration 342, loss: 0.7723391056060791
Iteration 343, loss: 0.7780662178993225
Iteration 344, loss: 0.7869287729263306
Iteration 345, loss: 0.8106381893157959
Iteration 346, loss: 0.7951616644859314
Iteration 347, loss: 0.8310299515724182
Iteration 348, loss: 0.7820034027099609
Iteration 349, loss: 0.7644932866096497
Iteration 350, loss: 0.7136849164962769
Iteration 351, loss: 0.7896157503128052
Iteration 352, loss: 0.89383864402771
Iteration 353, loss: 0.8264636993408203
Iteration 354, loss: 0.8411709070205688
Iteration 355, loss: 0.7507946491241455
Iteration 356, loss: 0.822133481502533
Iteration 357, loss: 0.8634814023971558
Iteration 358, loss: 0.7303842902183533
Iteration 359, loss: 0.8222361207008362
Iteration 360, loss: 0.7917464375495911
Iteration 361, loss: 0.7994983196258545
Iteration 362, loss: 0.8351473212242126
Iteration 363, loss: 0.696782112121582
Iteration 364, loss: 0.7027661204338074
Iteration 365, loss: 0.8460829257965088
Iteration 366, loss: 0.8240472078323364
Iteration 367, loss: 0.7513972520828247
Iteration 368, loss: 0.8548148274421692
Iteration 369, loss: 0.7602236270904541
Iteration 370, loss: 0.7484174966812134
Iteration 371, loss: 0.7359906435012817
Iteration 372, loss: 0.7702592015266418
Iteration 373, loss: 0.669430673122406
Iteration 374, loss: 0.7529183626174927
Iteration 375, loss: 0.8302730321884155
Iteration 376, loss: 0.7562160491943359
Iteration 377, loss: 0.7294586896896362
Iteration 378, loss: 0.7893180847167969
Iteration 379, loss: 0.7615473866462708
Iteration 380, loss: 0.8391824960708618
Iteration 381, loss: 0.7955535054206848
Iteration 382, loss: 0.8178874254226685
Iteration 383, loss: 0.7696306705474854
Iteration 384, loss: 0.7320446372032166
Iteration 385, loss: 0.8937121033668518
Iteration 386, loss: 0.8273982405662537
Iteration 387, loss: 0.8067731261253357
Iteration 388, loss: 0.7172713875770569
Iteration 389, loss: 0.7768769860267639
Iteration 390, loss: 0.7661257386207581
Iteration 391, loss: 0.7531622648239136
Iteration 392, loss: 0.7826283574104309
Iteration 393, loss: 0.8018662929534912
Iteration 394, loss: 0.7446836233139038
Iteration 395, loss: 0.7697896957397461
Iteration 396, loss: 0.7786036729812622
Iteration 397, loss: 0.7242592573165894
Iteration 398, loss: 0.7580749988555908
Iteration 399, loss: 0.7876268029212952
Iteration 400, loss: 0.8672468662261963
Iteration 401, loss: 0.8023996353149414
Iteration 402, loss: 0.7513532042503357
Iteration 403, loss: 0.807580828666687
```

```
Iteration 404, loss: 0.8073652982711792
Iteration 405, loss: 0.8021061420440674
Iteration 406, loss: 0.7004503607749939
Iteration 407, loss: 0.7381940484046936
Iteration 408, loss: 0.7757216095924377
Iteration 409, loss: 0.8037087321281433
Iteration 410, loss: 0.8095930814743042
Iteration 411, loss: 0.7864009141921997
Iteration 412, loss: 0.8592515587806702
Iteration 413, loss: 0.7331390976905823
Iteration 414, loss: 0.7745409607887268
Iteration 415, loss: 0.811752438545227
Iteration 416, loss: 0.8201002478599548
Iteration 417, loss: 0.7333618998527527
Iteration 418, loss: 0.7427974939346313
Iteration 419, loss: 0.8085097074508667
Iteration 420, loss: 0.7580959796905518
Iteration 421, loss: 0.9324063062667847
Iteration 422, loss: 0.752455472946167
Iteration 423, loss: 0.6598034501075745
Iteration 424, loss: 0.8277746438980103
Iteration 425, loss: 0.8187055587768555
Iteration 426, loss: 0.7531478404998779
Iteration 427, loss: 0.7882517576217651
Iteration 428, loss: 0.7225852608680725
Iteration 429, loss: 0.785750687122345
Iteration 430, loss: 0.7719271183013916
Iteration 431, loss: 0.826183557510376
Iteration 432, loss: 0.7022621631622314
Iteration 433, loss: 0.7248627543449402
Iteration 434, loss: 0.6981092691421509
Iteration 435, loss: 0.7346714735031128
Iteration 436, loss: 0.7434219121932983
Iteration 437, loss: 0.7736176252365112
Iteration 438, loss: 0.7533012628555298
Iteration 439, loss: 0.7410306930541992
Iteration 440, loss: 0.7627169489860535
Iteration 441, loss: 0.7774797081947327
Iteration 442, loss: 0.6939547657966614
Iteration 443, loss: 0.7600044012069702
Iteration 444, loss: 0.7843095064163208
Iteration 445, loss: 0.768522322177887
Iteration 446, loss: 0.7418997883796692
Iteration 447, loss: 0.7233248353004456
Iteration 448, loss: 0.8772900700569153
Iteration 449, loss: 0.6932507157325745
Iteration 450, loss: 0.7482211589813232
Iteration 451, loss: 0.701405942440033
Iteration 452, loss: 0.7335278391838074
Iteration 453, loss: 0.6651041507720947
Iteration 454, loss: 0.7338510751724243
Iteration 455, loss: 0.7564821243286133
Iteration 456, loss: 0.6954296827316284
Iteration 457, loss: 0.8163073062896729
Iteration 458, loss: 0.7283530831336975
Iteration 459, loss: 0.7662979960441589
Iteration 460, loss: 0.8451722860336304
Iteration 461, loss: 0.7928105592727661
Iteration 462, loss: 0.7491579651832581
Iteration 463, loss: 0.7463624477386475
Iteration 464, loss: 0.7037604451179504
Iteration 465, loss: 0.8019777536392212
Iteration 466, loss: 0.8298947811126709
Iteration 467, loss: 0.760809063911438
Iteration 468, loss: 0.7883173823356628
Iteration 469, loss: 0.6005494594573975
Iteration 470, loss: 0.7251055836677551
Iteration 471, loss: 0.7892777919769287
Iteration 472, loss: 0.7420789003372192
Iteration 473, loss: 0.7308151721954346
Iteration 474, loss: 0.7447219491004944
Iteration 475, loss: 0.7615166902542114
Iteration 476, loss: 0.8075329661369324
Iteration 477, loss: 0.7851430177688599
Iteration 478, loss: 0.7234639525413513
Iteration 479, loss: 0.8037726283073425
Iteration 480, loss: 0.7840582728385925
```

```
Iteration 481, loss: 0.8280483484268188
Iteration 482, loss: 0.7366863489151001
Iteration 483, loss: 0.7078185677528381
Iteration 484, loss: 0.796392560005188
Iteration 485, loss: 0.7258666157722473
Iteration 486, loss: 0.7630958557128906
Iteration 487, loss: 0.7364858984947205
Iteration 488, loss: 0.7406833171844482
Iteration 489, loss: 0.7265803217887878
Iteration 490, loss: 0.7174005508422852
Iteration 491, loss: 0.7064065933227539
Iteration 492, loss: 0.7564345002174377
Iteration 493, loss: 0.7808047533035278
Iteration 494, loss: 0.7304587960243225
Iteration 495, loss: 0.774482011795044
Iteration 496, loss: 0.7356421947479248
Iteration 497, loss: 0.6804494261741638
Iteration 498, loss: 0.7374099493026733
Iteration 499, loss: 0.7709011435508728
Iteration 500, loss: 0.7140483856201172
Iteration 501, loss: 0.76512610912323
Iteration 502, loss: 0.8166424632072449
Iteration 503, loss: 0.837768018245697
Iteration 504, loss: 0.7683976888656616
Iteration 505, loss: 0.7644465565681458
Iteration 506, loss: 0.8027762174606323
Iteration 507, loss: 0.6771716475486755
Iteration 508, loss: 0.7530030012130737
Iteration 509, loss: 0.7437784671783447
Iteration 510, loss: 0.7532981038093567
Iteration 511, loss: 0.7455888390541077
Iteration 512, loss: 0.8109118342399597
Iteration 513, loss: 0.7744772434234619
Iteration 514, loss: 0.7715717554092407
Iteration 515, loss: 0.6946709156036377
Iteration 516, loss: 0.6542145013809204
Iteration 517, loss: 0.6766499280929565
Iteration 518, loss: 0.677808403968811
Iteration 519, loss: 0.7756017446517944
Iteration 520, loss: 0.7111420035362244
Iteration 521, loss: 0.6908289194107056
Iteration 522, loss: 0.6742972731590271
Iteration 523, loss: 0.6544961333274841
Iteration 524, loss: 0.7329006195068359
Iteration 525, loss: 0.7680884003639221
Iteration 526, loss: 0.6430705189704895
Iteration 527, loss: 0.7182035446166992
Iteration 528, loss: 0.675407350063324
Iteration 529, loss: 0.7771501541137695
Iteration 530, loss: 0.7526319026947021
Iteration 531, loss: 0.7219126224517822
Iteration 532, loss: 0.7314809560775757
Iteration 533, loss: 0.714753270149231
Iteration 534, loss: 0.7365383505821228
Iteration 535, loss: 0.7956476807594299
Iteration 536, loss: 0.7882826924324036
Iteration 537, loss: 0.731028139591217
Iteration 538, loss: 0.7031902074813843
Iteration 539, loss: 0.7266252040863037
Iteration 540, loss: 0.8155847787857056
Iteration 541, loss: 0.7190093994140625
Iteration 542, loss: 0.7429361343383789
Iteration 543, loss: 0.7333409786224365
Iteration 544, loss: 0.8055260181427002
Iteration 545, loss: 0.734062910079956
Iteration 546, loss: 0.771371603012085
Iteration 547, loss: 0.6719564199447632
Iteration 548, loss: 0.7855721116065979
Iteration 549, loss: 0.6826034188270569
Iteration 550, loss: 0.8239426016807556
Iteration 551, loss: 0.7993717193603516
Iteration 552, loss: 0.6913309097290039
Iteration 553, loss: 0.6508862376213074
Iteration 554, loss: 0.7136095762252808
Iteration 555, loss: 0.671123743057251
Iteration 556, loss: 0.7064279913902283
Iteration 557, loss: 0.7594830989837646
```

```
Iteration 557, loss: 0.7594850969857646
Iteration 558, loss: 0.7489360570907593
Iteration 559, loss: 0.7947816252708435
Iteration 560, loss: 0.7437063455581665
Iteration 561, loss: 0.6724151968955994
Iteration 562, loss: 0.7485176920890808
Iteration 563, loss: 0.8243453502655029
Iteration 564, loss: 0.7873124480247498
Iteration 565, loss: 0.751327633857727
Iteration 566, loss: 0.7570149898529053
Iteration 567, loss: 0.7265599966049194
Iteration 568, loss: 0.6963227391242981
Iteration 569, loss: 0.8178691864013672
Iteration 570, loss: 0.6864162683486938
Iteration 571, loss: 0.6999011635780334
Iteration 572, loss: 0.7186193466186523
Iteration 573, loss: 0.6530869603157043
Iteration 574, loss: 0.667032778263092
Iteration 575, loss: 0.7186435461044312
Iteration 576, loss: 0.6777693629264832
Iteration 577, loss: 0.7087386846542358
Iteration 578, loss: 0.7307935357093811
Iteration 579, loss: 0.7268592119216919
Iteration 580, loss: 0.7085362076759338
Iteration 581, loss: 0.6868529915809631
Iteration 582, loss: 0.6969954967498779
Iteration 583, loss: 0.7122099995613098
Iteration 584, loss: 0.6987886428833008
Iteration 585, loss: 0.6629202961921692
Iteration 586, loss: 0.7072740197181702
Iteration 587, loss: 0.8117886781692505
Iteration 588, loss: 0.6516919732093811
Iteration 589, loss: 0.8332081437110901
Iteration 590, loss: 0.6763355731964111
Iteration 591, loss: 0.7180760502815247
Iteration 592, loss: 0.8345312476158142
Iteration 593, loss: 0.7124290466308594
Iteration 594, loss: 0.679806649684906
Iteration 595, loss: 0.6584721803665161
Iteration 596, loss: 0.782778799533844
Iteration 597, loss: 0.7158442139625549
Iteration 598, loss: 0.7965906858444214
Iteration 599, loss: 0.6853176355361938
Iteration 600, loss: 0.7460921406745911
Iteration 601, loss: 0.6472346782684326
Iteration 602, loss: 0.7277358174324036
Iteration 603, loss: 0.710342526435852
Iteration 604, loss: 0.6136643290519714
Iteration 605, loss: 0.6520496606826782
Iteration 606, loss: 0.6363950967788696
Iteration 607, loss: 0.6683605909347534
Iteration 608, loss: 0.668572187423706
Iteration 609, loss: 0.7300101518630981
Iteration 610, loss: 0.8031737208366394
Iteration 611, loss: 0.6416765451431274
Iteration 612, loss: 0.6593784093856812
Iteration 613, loss: 0.6849708557128906
Iteration 614, loss: 0.6952992081642151
Iteration 615, loss: 0.7833395004272461
Iteration 616, loss: 0.7776546478271484
Iteration 617, loss: 0.7024088501930237
Iteration 618, loss: 0.6874266266822815
Iteration 619, loss: 0.7103649973869324
Iteration 620, loss: 0.6995881199836731
Iteration 621, loss: 0.7480615973472595
Iteration 622, loss: 0.6808388233184814
Iteration 623, loss: 0.6954696178436279
Iteration 624, loss: 0.7460908889770508
Iteration 625, loss: 0.6711447238922119
Iteration 626, loss: 0.7219721078872681
Iteration 627, loss: 0.7516813278198242
Iteration 628, loss: 0.7104179859161377
Iteration 629, loss: 0.6181185841560364
Iteration 630, loss: 0.7525041103363037
Iteration 631, loss: 0.7110704183578491
Iteration 632, loss: 0.6886788606643677
Iteration 633, loss: 0.679343581199646
Iteration 634, loss: 0.708004891872406
```

```
Iteration 634, loss: 0.7080048918724060
Iteration 635, loss: 0.6852505207061768
Iteration 636, loss: 0.6882498860359192
Iteration 637, loss: 0.6979963779449463
Iteration 638, loss: 0.7806491851806641
Iteration 639, loss: 0.715377926826477
Iteration 640, loss: 0.688399076461792
Iteration 641, loss: 0.832912802696228
Iteration 642, loss: 0.7176449298858643
Iteration 643, loss: 0.7194563150405884
Iteration 644, loss: 0.7357338666915894
Iteration 645, loss: 0.7400308847427368
Iteration 646, loss: 0.7181733250617981
Iteration 647, loss: 0.6086116433143616
Iteration 648, loss: 0.7070028781890869
Iteration 649, loss: 0.6602521538734436
Iteration 650, loss: 0.6506534814834595
Iteration 651, loss: 0.7665795683860779
Iteration 652, loss: 0.7049797773361206
Iteration 653, loss: 0.8030596375465393
Iteration 654, loss: 0.7149679660797119
Iteration 655, loss: 0.6718314290046692
Iteration 656, loss: 0.6822750568389893
Iteration 657, loss: 0.6982957124710083
Iteration 658, loss: 0.7454436421394348
Iteration 659, loss: 0.7071444392204285
Iteration 660, loss: 0.6322072148323059
Iteration 661, loss: 0.6794332265853882
Iteration 662, loss: 0.7289373874664307
Iteration 663, loss: 0.629483163356781
Iteration 664, loss: 0.7212527394294739
Iteration 665, loss: 0.8355041146278381
Iteration 666, loss: 0.6658487915992737
Iteration 667, loss: 0.7326784133911133
Iteration 668, loss: 0.6663771271705627
Iteration 669, loss: 0.7349395155906677
Iteration 670, loss: 0.7148011326789856
Iteration 671, loss: 0.6875154972076416
Iteration 672, loss: 0.7218987345695496
Iteration 673, loss: 0.7569683790206909
Iteration 674, loss: 0.6747726798057556
Iteration 675, loss: 0.7108549475669861
Iteration 676, loss: 0.6681036353111267
Iteration 677, loss: 0.6461194753646851
Iteration 678, loss: 0.7291165590286255
Iteration 679, loss: 0.7180659770965576
Iteration 680, loss: 0.7365707159042358
Iteration 681, loss: 0.6790676116943359
Iteration 682, loss: 0.6569891571998596
Iteration 683, loss: 0.7140350937843323
Iteration 684, loss: 0.6396714448928833
Iteration 685, loss: 0.7377550601959229
Iteration 686, loss: 0.6707032322883606
Iteration 687, loss: 0.6980258226394653
Iteration 688, loss: 0.7156659364700317
Iteration 689, loss: 0.6852745413780212
Iteration 690, loss: 0.6615939736366272
Iteration 691, loss: 0.7227219939231873
Iteration 692, loss: 0.7053489089012146
Iteration 693, loss: 0.769404411315918
Iteration 694, loss: 0.7060059309005737
Iteration 695, loss: 0.727593719959259
Iteration 696, loss: 0.8456971049308777
Iteration 697, loss: 0.6727162599563599
Iteration 698, loss: 0.7071546912193298
Iteration 699, loss: 0.7256215810775757
Iteration 700, loss: 0.8135788440704346
Iteration 701, loss: 0.7112051248550415
Iteration 702, loss: 0.6991366147994995
Iteration 703, loss: 0.6920918822288513
Iteration 704, loss: 0.7059406638145447
Iteration 705, loss: 0.6862785816192627
Iteration 706, loss: 0.6858927011489868
Iteration 707, loss: 0.6758274435997009
Iteration 708, loss: 0.6984758377075195
Iteration 709, loss: 0.7262438535690308
Iteration 710, loss: 0.6463958024978638
Iteration 711, loss: 0.6920474171628480
```

```
Iteration 711, loss: 0.6920474171638489
Iteration 712, loss: 0.6354871392250061
Iteration 713, loss: 0.663760244846344
Iteration 714, loss: 0.6777093410491943
Iteration 715, loss: 0.682064950466156
Iteration 716, loss: 0.6780463457107544
Iteration 717, loss: 0.6911497712135315
Iteration 718, loss: 0.7468351721763611
Iteration 719, loss: 0.716536283493042
Iteration 720, loss: 0.6650218367576599
Iteration 721, loss: 0.6604272127151489
Iteration 722, loss: 0.6506472826004028
Iteration 723, loss: 0.6769847869873047
Iteration 724, loss: 0.6955105662345886
Iteration 725, loss: 0.6420494914054871
Iteration 726, loss: 0.6676886677742004
Iteration 727, loss: 0.7675946354866028
Iteration 728, loss: 0.701223611831665
Iteration 729, loss: 0.7056828737258911
Iteration 730, loss: 0.7546629309654236
Iteration 731, loss: 0.7333252429962158
Iteration 732, loss: 0.692010223865509
Iteration 733, loss: 0.6961026787757874
Iteration 734, loss: 0.7094614505767822
Iteration 735, loss: 0.6997223496437073
Iteration 736, loss: 0.7444192171096802
Iteration 737, loss: 0.7125311493873596
Iteration 738, loss: 0.7605011463165283
Iteration 739, loss: 0.6860173344612122
Iteration 740, loss: 0.7029454112052917
Iteration 741, loss: 0.7393723130226135
Iteration 742, loss: 0.7316151857376099
Iteration 743, loss: 0.7203035354614258
Iteration 744, loss: 0.6507731676101685
Iteration 745, loss: 0.7211644053459167
Iteration 746, loss: 0.7431187033653259
Iteration 747, loss: 0.7935252785682678
Iteration 748, loss: 0.7336196303367615
Iteration 749, loss: 0.6648390889167786
Iteration 750, loss: 0.7227330803871155
Iteration 751, loss: 0.6469438076019287
Iteration 752, loss: 0.703100860118866
Iteration 753, loss: 0.7177599668502808
Iteration 754, loss: 0.641130805015564
Iteration 755, loss: 0.6925540566444397
Iteration 756, loss: 0.7293780446052551
Iteration 757, loss: 0.6673065423965454
Iteration 758, loss: 0.6830065846443176
Iteration 759, loss: 0.6763544678688049
Iteration 760, loss: 0.6160793304443359
Iteration 761, loss: 0.6957293152809143
Iteration 762, loss: 0.6694835424423218
Iteration 763, loss: 0.7216524481773376
Iteration 764, loss: 0.6640816926956177
Iteration 765, loss: 0.6498915553092957
Iteration 766, loss: 0.5766963362693787
Iteration 767, loss: 0.6733747720718384
Iteration 768, loss: 0.6725877523422241
Iteration 769, loss: 0.6483296155929565
Iteration 770, loss: 0.6541985869407654
Iteration 771, loss: 0.631926953792572
Iteration 772, loss: 0.7123006582260132
Iteration 773, loss: 0.6496162414550781
Iteration 774, loss: 0.6913379430770874
Iteration 775, loss: 0.7083628177642822
Iteration 776, loss: 0.6534729599952698
Iteration 777, loss: 0.7217881679534912
Iteration 778, loss: 0.705180287361145
Iteration 779, loss: 0.6445081233978271
Iteration 780, loss: 0.7258228063583374
Iteration 781, loss: 0.6459706425666809
Iteration 782, loss: 0.6171401739120483
Iteration 783, loss: 0.6563276052474976
Iteration 784, loss: 0.6406208872795105
Iteration 785, loss: 0.6094147562980652
Iteration 786, loss: 0.6435518264770508
Iteration 787, loss: 0.6661306619644165
```

```
Iteration 788, loss: 0.682715117931366
Iteration 789, loss: 0.7402273416519165
Iteration 790, loss: 0.6529261469841003
Iteration 791, loss: 0.6531516313552856
Iteration 792, loss: 0.6166003942489624
Iteration 793, loss: 0.6401227712631226
Iteration 794, loss: 0.6161490082740784
Iteration 795, loss: 0.6925870180130005
Iteration 796, loss: 0.7332859039306641
Iteration 797, loss: 0.6242310404777527
Iteration 798, loss: 0.6547471284866333
Iteration 799, loss: 0.6839675903320312
Iteration 800, loss: 0.6273068785667419
Iteration 801, loss: 0.6107051372528076
Iteration 802, loss: 0.6984473466873169
Iteration 803, loss: 0.5711179375648499
Iteration 804, loss: 0.6600040197372437
Iteration 805, loss: 0.658733069896698
Iteration 806, loss: 0.6270389556884766
Iteration 807, loss: 0.6485360264778137
Iteration 808, loss: 0.6893457174301147
Iteration 809, loss: 0.7261873483657837
Iteration 810, loss: 0.6868555545806885
Iteration 811, loss: 0.6209429502487183
Iteration 812, loss: 0.6698790788650513
Iteration 813, loss: 0.6056420803070068
Iteration 814, loss: 0.6866874694824219
Iteration 815, loss: 0.6838915944099426
Iteration 816, loss: 0.5896062254905701
Iteration 817, loss: 0.6769949793815613
Iteration 818, loss: 0.6825377941131592
Iteration 819, loss: 0.7188329100608826
Iteration 820, loss: 0.6612623929977417
Iteration 821, loss: 0.763526976108551
Iteration 822, loss: 0.7135801315307617
Iteration 823, loss: 0.6087768077850342
Iteration 824, loss: 0.6865737438201904
Iteration 825, loss: 0.610679030418396
Iteration 826, loss: 0.6724193096160889
Iteration 827, loss: 0.6877216696739197
Iteration 828, loss: 0.6308498978614807
Iteration 829, loss: 0.6674543023109436
Iteration 830, loss: 0.6392418742179871
Iteration 831, loss: 0.7530924081802368
Iteration 832, loss: 0.5790587663650513
Iteration 833, loss: 0.6290337443351746
Iteration 834, loss: 0.611750066280365
Iteration 835, loss: 0.7544031739234924
Iteration 836, loss: 0.6315282583236694
Iteration 837, loss: 0.6455240249633789
Iteration 838, loss: 0.6205216646194458
Iteration 839, loss: 0.6264448165893555
Iteration 840, loss: 0.6332340836524963
Iteration 841, loss: 0.6562846899032593
Iteration 842, loss: 0.6218592524528503
Iteration 843, loss: 0.6052379012107849
Iteration 844, loss: 0.5911052227020264
Iteration 845, loss: 0.6194950938224792
Iteration 846, loss: 0.6399196982383728
Iteration 847, loss: 0.6635149717330933
Iteration 848, loss: 0.7806838750839233
Iteration 849, loss: 0.7677615284919739
Iteration 850, loss: 0.5692253708839417
Iteration 851, loss: 0.654346227645874
Iteration 852, loss: 0.6463546752929688
Iteration 853, loss: 0.6775157451629639
Iteration 854, loss: 0.6132296323776245
Iteration 855, loss: 0.6258076429367065
Iteration 856, loss: 0.6255158185958862
Iteration 857, loss: 0.7361761927604675
Iteration 858, loss: 0.701903223991394
Iteration 859, loss: 0.6347763538360596
Iteration 860, loss: 0.6859128475189209
Iteration 861, loss: 0.6604819893836975
Iteration 862, loss: 0.6440842747688293
Iteration 863, loss: 0.675784170627594
Iteration 864, loss: 0.6253531575202942
```

```
Iteration 865, loss: 0.6790532469749451
Iteration 866, loss: 0.6416808366775513
Iteration 867, loss: 0.6449558734893799
Iteration 868, loss: 0.6430408954620361
Iteration 869, loss: 0.5700759887695312
Iteration 870, loss: 0.6220682859420776
Iteration 871, loss: 0.6491774320602417
Iteration 872, loss: 0.6462789177894592
Iteration 873, loss: 0.6562796235084534
Iteration 874, loss: 0.6651768088340759
Iteration 875, loss: 0.6118944883346558
Iteration 876, loss: 0.5898662209510803
Iteration 877, loss: 0.5816580057144165
Iteration 878, loss: 0.7585504055023193
Iteration 879, loss: 0.6145409941673279
Iteration 880, loss: 0.6046310663223267
Iteration 881, loss: 0.6427987217903137
Iteration 882, loss: 0.5953057408332825
Iteration 883, loss: 0.6020224094390869
Iteration 884, loss: 0.6382684707641602
Iteration 885, loss: 0.6528434157371521
Iteration 886, loss: 0.6656027436256409
Iteration 887, loss: 0.6259041428565979
Iteration 888, loss: 0.5931633114814758
Iteration 889, loss: 0.6199208498001099
Iteration 890, loss: 0.6520683765411377
Iteration 891, loss: 0.6644653081893921
Iteration 892, loss: 0.6152803897857666
Iteration 893, loss: 0.6250174641609192
Iteration 894, loss: 0.6879395842552185
Iteration 895, loss: 0.6113541722297668
Iteration 896, loss: 0.6033629775047302
Iteration 897, loss: 0.6183633208274841
Iteration 898, loss: 0.6465592980384827
Iteration 899, loss: 0.5855289101600647
Iteration 900, loss: 0.6604378819465637
Iteration 901, loss: 0.6159229278564453
Iteration 902, loss: 0.5990491509437561
Iteration 903, loss: 0.5478845834732056
Iteration 904, loss: 0.6429821848869324
Iteration 905, loss: 0.6718733906745911
Iteration 906, loss: 0.6284554600715637
Iteration 907, loss: 0.6557324528694153
Iteration 908, loss: 0.6515026092529297
Iteration 909, loss: 0.6339667439460754
Iteration 910, loss: 0.6709883809089661
Iteration 911, loss: 0.6233908534049988
Iteration 912, loss: 0.6170549392700195
Iteration 913, loss: 0.5756374597549438
Iteration 914, loss: 0.5608857274055481
Iteration 915, loss: 0.6212491989135742
Iteration 916, loss: 0.636346161365509
Iteration 917, loss: 0.5154281854629517
Iteration 918, loss: 0.6286675333976746
Iteration 919, loss: 0.6040599346160889
Iteration 920, loss: 0.5958192348480225
Iteration 921, loss: 0.6139214038848877
Iteration 922, loss: 0.6521099805831909
Iteration 923, loss: 0.6542444825172424
Iteration 924, loss: 0.5998026132583618
Iteration 925, loss: 0.6956421732902527
Iteration 926, loss: 0.6055550575256348
Iteration 927, loss: 0.5932872295379639
Iteration 928, loss: 0.6646223068237305
Iteration 929, loss: 0.5304097533226013
Iteration 930, loss: 0.7000669240951538
Iteration 931, loss: 0.6222591996192932
Iteration 932, loss: 0.5618321299552917
Iteration 933, loss: 0.6338673830032349
Iteration 934, loss: 0.6910197734832764
Iteration 935, loss: 0.6088632941246033
Iteration 936, loss: 0.6512609124183655
Iteration 937, loss: 0.6557268500328064
Iteration 938, loss: 0.5881652235984802
Iteration 939, loss: 0.6593648195266724
Iteration 940, loss: 0.5938503742218018
Iteration 941, loss: 0.6506609320640564
```

```
Iteration 942, loss: 0.5733956098556519
Iteration 943, loss: 0.6429315805435181
Iteration 944, loss: 0.7025004029273987
Iteration 945, loss: 0.7072393894195557
Iteration 946, loss: 0.621221661567688
Iteration 947, loss: 0.6195839643478394
Iteration 948, loss: 0.5950924158096313
Iteration 949, loss: 0.535594642162323
Iteration 950, loss: 0.613554835319519
Iteration 951, loss: 0.6030067801475525
Iteration 952, loss: 0.5800262689590454
Iteration 953, loss: 0.5938462615013123
Iteration 954, loss: 0.5888075232505798
Iteration 955, loss: 0.6332883834838867
Iteration 956, loss: 0.5757916569709778
Iteration 957, loss: 0.6163014769554138
Iteration 958, loss: 0.5706685781478882
Iteration 959, loss: 0.6005257964134216
Iteration 960, loss: 0.6424334645271301
Iteration 961, loss: 0.5215511918067932
Iteration 962, loss: 0.6007929444313049
Iteration 963, loss: 0.6532762050628662
Iteration 964, loss: 0.6027713418006897
Iteration 965, loss: 0.6460216045379639
Iteration 966, loss: 0.6875719428062439
Iteration 967, loss: 0.5549365282058716
Iteration 968, loss: 0.5580839514732361
Iteration 969, loss: 0.638715386390686
Iteration 970, loss: 0.657795786857605
Iteration 971, loss: 0.585892915725708
Iteration 972, loss: 0.5989444255828857
Iteration 973, loss: 0.6235386729240417
Iteration 974, loss: 0.580352246761322
Iteration 975, loss: 0.625419020652771
Iteration 976, loss: 0.6665743589401245
Iteration 977, loss: 0.6254605650901794
Iteration 978, loss: 0.5843326449394226
Iteration 979, loss: 0.6281428933143616
Iteration 980, loss: 0.5385528206825256
Iteration 981, loss: 0.5817055106163025
Iteration 982, loss: 0.5806183815002441
Iteration 983, loss: 0.572575032711029
Iteration 984, loss: 0.5999234914779663
Iteration 985, loss: 0.6061955690383911
Iteration 986, loss: 0.6107160449028015
Iteration 987, loss: 0.6107631325721741
Iteration 988, loss: 0.5403700470924377
Iteration 989, loss: 0.6098780632019043
Iteration 990, loss: 0.5992317795753479
Iteration 991, loss: 0.6532313823699951
Iteration 992, loss: 0.6107380986213684
Iteration 993, loss: 0.583556592464447
Iteration 994, loss: 0.5398576259613037
Iteration 995, loss: 0.5699387788772583
Iteration 996, loss: 0.6399946212768555
Iteration 997, loss: 0.7611187696456909
Iteration 998, loss: 0.5926269888877869
Iteration 999, loss: 0.5379374027252197
End of epoch 6, loss: 1.7589681148529053
Validation loss: 1.1912322044372559
Iteration 0, loss: 0.6665000915527344
Iteration 1, loss: 0.5462044477462769
Iteration 2, loss: 0.5681434273719788
Iteration 3, loss: 0.5382514595985413
Iteration 4, loss: 0.6469284892082214
Iteration 5, loss: 0.6145490407943726
Iteration 6, loss: 0.5895761251449585
Iteration 7, loss: 0.5469764471054077
Iteration 8, loss: 0.5905635356903076
Iteration 9, loss: 0.565906822681427
Iteration 10, loss: 0.632039487361908
Iteration 11, loss: 0.6332642436027527
Iteration 12, loss: 0.669865071773529
Iteration 13, loss: 0.5474310517311096
Iteration 14, loss: 0.6287732720375061
Iteration 15, loss: 0.6013845801353455
Iteration 16, loss: 0.6191433668136597
```

```
Iteration 17, loss: 0.6411393880844116
Iteration 18, loss: 0.6649078726768494
Iteration 19, loss: 0.6173925995826721
Iteration 20, loss: 0.6467505097389221
Iteration 21, loss: 0.6631468534469604
Iteration 22, loss: 0.6291888952255249
Iteration 23, loss: 0.5497475862503052
Iteration 24, loss: 0.5927566289901733
Iteration 25, loss: 0.6493628025054932
Iteration 26, loss: 0.7061859965324402
Iteration 27, loss: 0.6799107193946838
Iteration 28, loss: 0.6068140268325806
Iteration 29, loss: 0.6181630492210388
Iteration 30, loss: 0.6145358681678772
Iteration 31, loss: 0.5703355073928833
Iteration 32, loss: 0.6114943027496338
Iteration 33, loss: 0.6379684209823608
Iteration 34, loss: 0.6692882776260376
Iteration 35, loss: 0.5864897966384888
Iteration 36, loss: 0.5897533893585205
Iteration 37, loss: 0.5257866382598877
Iteration 38, loss: 0.6136492490768433
Iteration 39, loss: 0.5911397337913513
Iteration 40, loss: 0.5905194282531738
Iteration 41, loss: 0.6472712159156799
Iteration 42, loss: 0.5430986881256104
Iteration 43, loss: 0.5616767406463623
Iteration 44, loss: 0.5867823362350464
Iteration 45, loss: 0.6741635799407959
Iteration 46, loss: 0.6364250183105469
Iteration 47, loss: 0.5873063802719116
Iteration 48, loss: 0.6330575942993164
Iteration 49, loss: 0.6577568650245667
Iteration 50, loss: 0.5959106683731079
Iteration 51, loss: 0.5896448493003845
Iteration 52, loss: 0.5892296433448792
Iteration 53, loss: 0.5686604380607605
Iteration 54, loss: 0.5675601959228516
Iteration 55, loss: 0.6781317591667175
Iteration 56, loss: 0.49863559007644653
Iteration 57, loss: 0.6366126537322998
Iteration 58, loss: 0.6902245879173279
Iteration 59, loss: 0.5382639765739441
Iteration 60, loss: 0.5660065412521362
Iteration 61, loss: 0.7081008553504944
Iteration 62, loss: 0.6707220673561096
Iteration 63, loss: 0.5987451076507568
Iteration 64, loss: 0.5644603967666626
Iteration 65, loss: 0.6303953528404236
Iteration 66, loss: 0.5499672293663025
Iteration 67, loss: 0.6474276185035706
Iteration 68, loss: 0.6033510565757751
Iteration 69, loss: 0.6534325480461121
Iteration 70, loss: 0.5596904158592224
Iteration 71, loss: 0.6506742238998413
Iteration 72, loss: 0.5627849102020264
Iteration 73, loss: 0.5620670914649963
Iteration 74, loss: 0.7031892538070679
Iteration 75, loss: 0.5946177244186401
Iteration 76, loss: 0.5539941191673279
Iteration 77, loss: 0.6164015531539917
Iteration 78, loss: 0.6183942556381226
Iteration 79, loss: 0.5214621424674988
Iteration 80, loss: 0.5846730470657349
Iteration 81, loss: 0.5897908210754395
Iteration 82, loss: 0.56578129529953
Iteration 83, loss: 0.6225101947784424
Iteration 84, loss: 0.6169043779373169
Iteration 85, loss: 0.5550137758255005
Iteration 86, loss: 0.5857762694358826
Iteration 87, loss: 0.5999957919120789
Iteration 88, loss: 0.6529505848884583
Iteration 89, loss: 0.6071507930755615
Iteration 90, loss: 0.5430310368537903
Iteration 91, loss: 0.6255862712860107
Iteration 92, loss: 0.522678792476654
Iteration 93, loss: 0.637019693851471
```

```
Iteration 94, loss: 0.5519654750823975
Iteration 95, loss: 0.57383131980896
Iteration 96, loss: 0.5432835817337036
Iteration 97, loss: 0.5362409353256226
Iteration 98, loss: 0.6347137689590454
Iteration 99, loss: 0.5778191685676575
Iteration 100, loss: 0.5075140595436096
Iteration 101, loss: 0.5632899403572083
Iteration 102, loss: 0.5535290241241455
Iteration 103, loss: 0.4942598044872284
Iteration 104, loss: 0.6977672576904297
Iteration 105, loss: 0.6209660172462463
Iteration 106, loss: 0.6033396124839783
Iteration 107, loss: 0.5138502717018127
Iteration 108, loss: 0.5753366947174072
Iteration 109, loss: 0.6587884426116943
Iteration 110, loss: 0.6365460157394409
Iteration 111, loss: 0.6780939102172852
Iteration 112, loss: 0.5492252707481384
Iteration 113, loss: 0.6767125725746155
Iteration 114, loss: 0.5959800481796265
Iteration 115, loss: 0.5974347591400146
Iteration 116, loss: 0.7157501578330994
Iteration 117, loss: 0.6689746379852295
Iteration 118, loss: 0.6050633192062378
Iteration 119, loss: 0.6031100153923035
Iteration 120, loss: 0.5334571599960327
Iteration 121, loss: 0.4840131998062134
Iteration 122, loss: 0.556280791759491
Iteration 123, loss: 0.57906574010849
Iteration 124, loss: 0.5547638535499573
Iteration 125, loss: 0.558188796043396
Iteration 126, loss: 0.5511943697929382
Iteration 127, loss: 0.5151695609092712
Iteration 128, loss: 0.5427678823471069
Iteration 129, loss: 0.5978320240974426
Iteration 130, loss: 0.5399001240730286
Iteration 131, loss: 0.5953906178474426
Iteration 132, loss: 0.6060569882392883
Iteration 133, loss: 0.5269914865493774
Iteration 134, loss: 0.5803970694541931
Iteration 135, loss: 0.5325458645820618
Iteration 136, loss: 0.5142765641212463
Iteration 137, loss: 0.5253081917762756
Iteration 138, loss: 0.5866838693618774
Iteration 139, loss: 0.5514191389083862
Iteration 140, loss: 0.5181262493133545
Iteration 141, loss: 0.5625912547111511
Iteration 142, loss: 0.618439257144928
Iteration 143, loss: 0.5201343297958374
Iteration 144, loss: 0.5023159384727478
Iteration 145, loss: 0.5679849982261658
Iteration 146, loss: 0.490822970867157
Iteration 147, loss: 0.5992709398269653
Iteration 148, loss: 0.6185916066169739
Iteration 149, loss: 0.6408127546310425
Iteration 150, loss: 0.5394671559333801
Iteration 151, loss: 0.6063522100448608
Iteration 152, loss: 0.5617595314979553
Iteration 153, loss: 0.6184964776039124
Iteration 154, loss: 0.6257454752922058
Iteration 155, loss: 0.5822387933731079
Iteration 156, loss: 0.5826345086097717
Iteration 157, loss: 0.517297625541687
Iteration 158, loss: 0.5070986151695251
Iteration 159, loss: 0.4587855637073517
Iteration 160, loss: 0.6057040095329285
Iteration 161, loss: 0.5416584014892578
Iteration 162, loss: 0.5757129192352295
Iteration 163, loss: 0.5612152218818665
Iteration 164, loss: 0.5229899883270264
Iteration 165, loss: 0.5405610203742981
Iteration 166, loss: 0.5859717726707458
Iteration 167, loss: 0.6102120280265808
Iteration 168, loss: 0.5451760292053223
Iteration 169, loss: 0.6179670095443726
Iteration 170, loss: 0.5598944425582886
```

```
Iteration 171, loss: 0.6515529155731201
Iteration 172, loss: 0.6032933592796326
Iteration 173, loss: 0.5051769614219666
Iteration 174, loss: 0.538699209690094
Iteration 175, loss: 0.5710490942001343
Iteration 176, loss: 0.4715490937232971
Iteration 177, loss: 0.4770118296146393
Iteration 178, loss: 0.5543356537818909
Iteration 179, loss: 0.5139683485031128
Iteration 180, loss: 0.6010326147079468
Iteration 181, loss: 0.6003028750419617
Iteration 182, loss: 0.49797746539115906
Iteration 183, loss: 0.5883920192718506
Iteration 184, loss: 0.4944482147693634
Iteration 185, loss: 0.5540784597396851
Iteration 186, loss: 0.4932960271835327
Iteration 187, loss: 0.5822470188140869
Iteration 188, loss: 0.5672085881233215
Iteration 189, loss: 0.5790119767189026
Iteration 190, loss: 0.5656012296676636
Iteration 191, loss: 0.559664785861969
Iteration 192, loss: 0.5700756311416626
Iteration 193, loss: 0.5706323385238647
Iteration 194, loss: 0.5397428870201111
Iteration 195, loss: 0.5590371489524841
Iteration 196, loss: 0.6423826217651367
Iteration 197, loss: 0.5478471517562866
Iteration 198, loss: 0.5675585865974426
Iteration 199, loss: 0.5105149149894714
Iteration 200, loss: 0.4979204833507538
Iteration 201, loss: 0.5568066835403442
Iteration 202, loss: 0.5898823142051697
Iteration 203, loss: 0.5126275420188904
Iteration 204, loss: 0.5705310106277466
Iteration 205, loss: 0.6075810194015503
Iteration 206, loss: 0.608695387840271
Iteration 207, loss: 0.586946427822113
Iteration 208, loss: 0.5658473968505859
Iteration 209, loss: 0.49363434314727783
Iteration 210, loss: 0.5349817872047424
Iteration 211, loss: 0.619346559047699
Iteration 212, loss: 0.6035159826278687
Iteration 213, loss: 0.5249552726745605
Iteration 214, loss: 0.5949584245681763
Iteration 215, loss: 0.5322684049606323
Iteration 216, loss: 0.49463951587677
Iteration 217, loss: 0.5783490538597107
Iteration 218, loss: 0.5364951491355896
Iteration 219, loss: 0.5126954317092896
Iteration 220, loss: 0.5407660007476807
Iteration 221, loss: 0.5644316077232361
Iteration 222, loss: 0.48148074746131897
Iteration 223, loss: 0.5287088751792908
Iteration 224, loss: 0.5813982486724854
Iteration 225, loss: 0.6398972868919373
Iteration 226, loss: 0.4787994921207428
Iteration 227, loss: 0.5599289536476135
Iteration 228, loss: 0.5362964868545532
Iteration 229, loss: 0.5578166246414185
Iteration 230, loss: 0.5208192467689514
Iteration 231, loss: 0.49239861965179443
Iteration 232, loss: 0.5022550821304321
Iteration 233, loss: 0.5703920722007751
Iteration 234, loss: 0.5161157250404358
Iteration 235, loss: 0.527293860912323
Iteration 236, loss: 0.5175367593765259
Iteration 237, loss: 0.5119343996047974
Iteration 238, loss: 0.5252951979637146
Iteration 239, loss: 0.5267354846000671
Iteration 240, loss: 0.5302637815475464
Iteration 241, loss: 0.5011206269264221
Iteration 242, loss: 0.6000599265098572
Iteration 243, loss: 0.6010409593582153
Iteration 244, loss: 0.48315757513046265
Iteration 245, loss: 0.4481247365474701
Iteration 246, loss: 0.4617133140563965
Iteration 247, loss: 0.5138390064239502
```

```
Iteration 248, loss: 0.5299096703529358
Iteration 249, loss: 0.52947598695755
Iteration 250, loss: 0.5629012584686279
Iteration 251, loss: 0.5177149772644043
Iteration 252, loss: 0.5636367201805115
Iteration 253, loss: 0.499438613653183
Iteration 254, loss: 0.5534554123878479
Iteration 255, loss: 0.5131422877311707
Iteration 256, loss: 0.6427447199821472
Iteration 257, loss: 0.5382726192474365
Iteration 258, loss: 0.5972495675086975
Iteration 259, loss: 0.5276431441307068
Iteration 260, loss: 0.5093870759010315
Iteration 261, loss: 0.5535773634910583
Iteration 262, loss: 0.49427101016044617
Iteration 263, loss: 0.5072072148323059
Iteration 264, loss: 0.5481002330780029
Iteration 265, loss: 0.5660634636878967
Iteration 266, loss: 0.5692459344863892
Iteration 267, loss: 0.4935441017150879
Iteration 268, loss: 0.5320953130722046
Iteration 269, loss: 0.5496877431869507
Iteration 270, loss: 0.44785356521606445
Iteration 271, loss: 0.5197373032569885
Iteration 272, loss: 0.5652074217796326
Iteration 273, loss: 0.5424960255622864
Iteration 274, loss: 0.4838983118534088
Iteration 275, loss: 0.5362498760223389
Iteration 276, loss: 0.6382588148117065
Iteration 277, loss: 0.44100242853164673
Iteration 278, loss: 0.49675253033638
Iteration 279, loss: 0.5326586961746216
Iteration 280, loss: 0.5170150995254517
Iteration 281, loss: 0.499454140663147
Iteration 282, loss: 0.5015616416931152
Iteration 283, loss: 0.518037736415863
Iteration 284, loss: 0.5250194072723389
Iteration 285, loss: 0.5525103807449341
Iteration 286, loss: 0.6326991319656372
Iteration 287, loss: 0.5092136859893799
Iteration 288, loss: 0.5151699781417847
Iteration 289, loss: 0.5174272060394287
Iteration 290, loss: 0.5007020235061646
Iteration 291, loss: 0.5439106822013855
Iteration 292, loss: 0.5676199197769165
Iteration 293, loss: 0.5177472233772278
Iteration 294, loss: 0.5007203221321106
Iteration 295, loss: 0.5064064264297485
Iteration 296, loss: 0.538484513759613
Iteration 297, loss: 0.5927550792694092
Iteration 298, loss: 0.512265682220459
Iteration 299, loss: 0.5140582323074341
Iteration 300, loss: 0.4805278480052948
Iteration 301, loss: 0.5939115285873413
Iteration 302, loss: 0.494875431060791
Iteration 303, loss: 0.5340882539749146
Iteration 304, loss: 0.5151989459991455
Iteration 305, loss: 0.4673870801925659
Iteration 306, loss: 0.5741161704063416
Iteration 307, loss: 0.4837879538536072
Iteration 308, loss: 0.5481751561164856
Iteration 309, loss: 0.5550399422645569
Iteration 310, loss: 0.43667855858802795
Iteration 311, loss: 0.5520881414413452
Iteration 312, loss: 0.5754787921905518
Iteration 313, loss: 0.5182515978813171
Iteration 314, loss: 0.46308663487434387
Iteration 315, loss: 0.5628492832183838
Iteration 316, loss: 0.5751099586486816
Iteration 317, loss: 0.49791285395622253
Iteration 318, loss: 0.4432271420955658
Iteration 319, loss: 0.5024064779281616
Iteration 320, loss: 0.5372991561889648
Iteration 321, loss: 0.5044656991958618
Iteration 322, loss: 0.4982423186302185
Iteration 323, loss: 0.45612651109695435
Iteration 324, loss: 0.5766839385032654
```

```
Iteration 325, loss: 0.47564947605133057
Iteration 326, loss: 0.5218837261199951
Iteration 327, loss: 0.43969982862472534
Iteration 328, loss: 0.47625887393951416
Iteration 329, loss: 0.5513915419578552
Iteration 330, loss: 0.5249058604240417
Iteration 331, loss: 0.49737614393234253
Iteration 332, loss: 0.4425792098045349
Iteration 333, loss: 0.5232939124107361
Iteration 334, loss: 0.49125415086746216
Iteration 335, loss: 0.519015908241272
Iteration 336, loss: 0.5307953953742981
Iteration 337, loss: 0.5439947843551636
Iteration 338, loss: 0.5009881258010864
Iteration 339, loss: 0.5764271020889282
Iteration 340, loss: 0.48862549662590027
Iteration 341, loss: 0.5558931827545166
Iteration 342, loss: 0.4855753183364868
Iteration 343, loss: 0.4860617220401764
Iteration 344, loss: 0.5111207365989685
Iteration 345, loss: 0.5420048236846924
Iteration 346, loss: 0.506646990776062
Iteration 347, loss: 0.5129203200340271
Iteration 348, loss: 0.5082739591598511
Iteration 349, loss: 0.4712309241294861
Iteration 350, loss: 0.4382322132587433
Iteration 351, loss: 0.5038333535194397
Iteration 352, loss: 0.614671528339386
Iteration 353, loss: 0.5361002683639526
Iteration 354, loss: 0.5607672929763794
Iteration 355, loss: 0.4881361722946167
Iteration 356, loss: 0.5310043692588806
Iteration 357, loss: 0.5965496897697449
Iteration 358, loss: 0.4630119502544403
Iteration 359, loss: 0.5727478265762329
Iteration 360, loss: 0.5048577189445496
Iteration 361, loss: 0.5069353580474854
Iteration 362, loss: 0.5489486455917358
Iteration 363, loss: 0.4319227635860443
Iteration 364, loss: 0.4213278591632843
Iteration 365, loss: 0.5228357911109924
Iteration 366, loss: 0.5218896865844727
Iteration 367, loss: 0.4742611348628998
Iteration 368, loss: 0.5484575033187866
Iteration 369, loss: 0.4769482910633087
Iteration 370, loss: 0.4714743494987488
Iteration 371, loss: 0.45005881786346436
Iteration 372, loss: 0.514827311038971
Iteration 373, loss: 0.4344366490840912
Iteration 374, loss: 0.47005030512809753
Iteration 375, loss: 0.5244837403297424
Iteration 376, loss: 0.4874744117259979
Iteration 377, loss: 0.4473513960838318
Iteration 378, loss: 0.4856938123703003
Iteration 379, loss: 0.48039835691452026
Iteration 380, loss: 0.5310124158859253
Iteration 381, loss: 0.5063183307647705
Iteration 382, loss: 0.5330727696418762
Iteration 383, loss: 0.4752025008201599
Iteration 384, loss: 0.44737374782562256
Iteration 385, loss: 0.5960658192634583
Iteration 386, loss: 0.5341572165489197
Iteration 387, loss: 0.5328930616378784
Iteration 388, loss: 0.48610103130340576
Iteration 389, loss: 0.499947190284729
Iteration 390, loss: 0.5005716681480408
Iteration 391, loss: 0.5056277513504028
Iteration 392, loss: 0.5064346790313721
Iteration 393, loss: 0.5386257171630859
Iteration 394, loss: 0.4754786193370819
Iteration 395, loss: 0.46332284808158875
Iteration 396, loss: 0.5041200518608093
Iteration 397, loss: 0.46269533038139343
Iteration 398, loss: 0.4832817316055298
Iteration 399, loss: 0.4738452732563019
Iteration 400, loss: 0.5800175070762634
Iteration 401, loss: 0.5209293961524963
```

```
Iteration 402, loss: 0.4729577302932739
Iteration 403, loss: 0.4990099370479584
Iteration 404, loss: 0.5065807104110718
Iteration 405, loss: 0.5387891530990601
Iteration 406, loss: 0.4527890980243683
Iteration 407, loss: 0.4780719578266144
Iteration 408, loss: 0.48994606733322144
Iteration 409, loss: 0.5405473709106445
Iteration 410, loss: 0.5367697477340698
Iteration 411, loss: 0.4945034682750702
Iteration 412, loss: 0.5727501511573792
Iteration 413, loss: 0.46697866916656494
Iteration 414, loss: 0.514318585395813
Iteration 415, loss: 0.5167313814163208
Iteration 416, loss: 0.5157675743103027
Iteration 417, loss: 0.4419424831867218
Iteration 418, loss: 0.46532344818115234
Iteration 419, loss: 0.5379157066345215
Iteration 420, loss: 0.4511105716228485
Iteration 421, loss: 0.6121082305908203
Iteration 422, loss: 0.46130892634391785
Iteration 423, loss: 0.4066322147846222
Iteration 424, loss: 0.5324291586875916
Iteration 425, loss: 0.5534636378288269
Iteration 426, loss: 0.46485117077827454
Iteration 427, loss: 0.5073772668838501
Iteration 428, loss: 0.46696752309799194
Iteration 429, loss: 0.4808715581893921
Iteration 430, loss: 0.5150101184844971
Iteration 431, loss: 0.5365885496139526
Iteration 432, loss: 0.4367073178291321
Iteration 433, loss: 0.45268502831459045
Iteration 434, loss: 0.42411673069000244
Iteration 435, loss: 0.44967105984687805
Iteration 436, loss: 0.4656173586845398
Iteration 437, loss: 0.5013827681541443
Iteration 438, loss: 0.46309977769851685
Iteration 439, loss: 0.44370466470718384
Iteration 440, loss: 0.5038254261016846
Iteration 441, loss: 0.5254093408584595
Iteration 442, loss: 0.43819481134414673
Iteration 443, loss: 0.4992978870868683
Iteration 444, loss: 0.5064031481742859
Iteration 445, loss: 0.5206582546234131
Iteration 446, loss: 0.49334588646888733
Iteration 447, loss: 0.45695993304252625
Iteration 448, loss: 0.5856105089187622
Iteration 449, loss: 0.45609503984451294
Iteration 450, loss: 0.5029512643814087
Iteration 451, loss: 0.43483081459999084
Iteration 452, loss: 0.5198734998703003
Iteration 453, loss: 0.43002986907958984
Iteration 454, loss: 0.4549388587474823
Iteration 455, loss: 0.46836358308792114
Iteration 456, loss: 0.43033143877983093
Iteration 457, loss: 0.5748091340065002
Iteration 458, loss: 0.4367496371269226
Iteration 459, loss: 0.5142955183982849
Iteration 460, loss: 0.5943402051925659
Iteration 461, loss: 0.5324087738990784
Iteration 462, loss: 0.49666622281074524
Iteration 463, loss: 0.4745974838733673
Iteration 464, loss: 0.4437955319881439
Iteration 465, loss: 0.5359690189361572
Iteration 466, loss: 0.5369347929954529
Iteration 467, loss: 0.4742377996444702
Iteration 468, loss: 0.4962175190448761
Iteration 469, loss: 0.3648967146873474
Iteration 470, loss: 0.440378874540329
Iteration 471, loss: 0.5117113590240479
Iteration 472, loss: 0.4681377112865448
Iteration 473, loss: 0.46796199679374695
Iteration 474, loss: 0.4778057932853699
Iteration 475, loss: 0.48683732748031616
Iteration 476, loss: 0.5180161595344543
Iteration 477, loss: 0.5009137392044067
Iteration 478, loss: 0.4488855004310608
```

```
Iteration 479, loss: 0.5314197540283203
Iteration 480, loss: 0.5103327631950378
Iteration 481, loss: 0.551516592502594
Iteration 482, loss: 0.4895896911621094
Iteration 483, loss: 0.4575462341308594
Iteration 484, loss: 0.5176874399185181
Iteration 485, loss: 0.4576081931591034
Iteration 486, loss: 0.4811660051345825
Iteration 487, loss: 0.4667240381240845
Iteration 488, loss: 0.4824308753013611
Iteration 489, loss: 0.46739643812179565
Iteration 490, loss: 0.4566015899181366
Iteration 491, loss: 0.460296094417572
Iteration 492, loss: 0.5006201863288879
Iteration 493, loss: 0.522925853729248
Iteration 494, loss: 0.4926530122756958
Iteration 495, loss: 0.5348076224327087
Iteration 496, loss: 0.47689706087112427
Iteration 497, loss: 0.4555656611919403
Iteration 498, loss: 0.5127162337303162
Iteration 499, loss: 0.5190514922142029
Iteration 500, loss: 0.46004146337509155
Iteration 501, loss: 0.49850308895111084
Iteration 502, loss: 0.5474510788917542
Iteration 503, loss: 0.57948899269104
Iteration 504, loss: 0.4998190999031067
Iteration 505, loss: 0.49884191155433655
Iteration 506, loss: 0.5363461971282959
Iteration 507, loss: 0.4260466992855072
Iteration 508, loss: 0.5068140029907227
Iteration 509, loss: 0.4963063895702362
Iteration 510, loss: 0.5028018951416016
Iteration 511, loss: 0.5260578393936157
Iteration 512, loss: 0.5517080426216125
Iteration 513, loss: 0.5023217797279358
Iteration 514, loss: 0.48368149995803833
Iteration 515, loss: 0.43640217185020447
Iteration 516, loss: 0.42094945907592773
Iteration 517, loss: 0.4390764832496643
Iteration 518, loss: 0.43283236026763916
Iteration 519, loss: 0.5226492285728455
Iteration 520, loss: 0.4576597511768341
Iteration 521, loss: 0.4626428782939911
Iteration 522, loss: 0.4478260278701782
Iteration 523, loss: 0.4336366355419159
Iteration 524, loss: 0.5005069971084595
Iteration 525, loss: 0.5017322897911072
Iteration 526, loss: 0.4128393530845642
Iteration 527, loss: 0.46583038568496704
Iteration 528, loss: 0.44277483224868774
Iteration 529, loss: 0.5634096264839172
Iteration 530, loss: 0.4791496992111206
Iteration 531, loss: 0.49229151010513306
Iteration 532, loss: 0.45314231514930725
Iteration 533, loss: 0.44561001658439636
Iteration 534, loss: 0.48124316334724426
Iteration 535, loss: 0.5377395153045654
Iteration 536, loss: 0.5224063396453857
Iteration 537, loss: 0.4542461931705475
Iteration 538, loss: 0.46090900897979736
Iteration 539, loss: 0.45810553431510925
Iteration 540, loss: 0.5624277591705322
Iteration 541, loss: 0.4403349757194519
Iteration 542, loss: 0.4831533432006836
Iteration 543, loss: 0.48595502972602844
Iteration 544, loss: 0.5596581697463989
Iteration 545, loss: 0.47271251678466797
Iteration 546, loss: 0.50508713722229
Iteration 547, loss: 0.42498427629470825
Iteration 548, loss: 0.5143750905990601
Iteration 549, loss: 0.45913487672805786
Iteration 550, loss: 0.5562358498573303
Iteration 551, loss: 0.5460923910140991
Iteration 552, loss: 0.471020370721817
Iteration 553, loss: 0.41418230533599854
Iteration 554, loss: 0.48037534952163696
Iteration 555, loss: 0.4515359103679657
```

```
Iteration 556, loss: 0.4415849447250366
Iteration 557, loss: 0.5180873870849609
Iteration 558, loss: 0.49544066190719604
Iteration 559, loss: 0.5130970478057861
Iteration 560, loss: 0.49786078929901123
Iteration 561, loss: 0.46748220920562744
Iteration 562, loss: 0.47417962551116943
Iteration 563, loss: 0.5572566390037537
Iteration 564, loss: 0.5313489437103271
Iteration 565, loss: 0.5034261345863342
Iteration 566, loss: 0.49985137581825256
Iteration 567, loss: 0.47077083587646484
Iteration 568, loss: 0.4513128399848938
Iteration 569, loss: 0.5468873977661133
Iteration 570, loss: 0.46761348843574524
Iteration 571, loss: 0.469775527715683
Iteration 572, loss: 0.4309632182121277
Iteration 573, loss: 0.4127328097820282
Iteration 574, loss: 0.41095593571662903
Iteration 575, loss: 0.49140724539756775
Iteration 576, loss: 0.4372660517692566
Iteration 577, loss: 0.4826372563838959
Iteration 578, loss: 0.49205252528190613
Iteration 579, loss: 0.4986810088157654
Iteration 580, loss: 0.46256452798843384
Iteration 581, loss: 0.47272637486457825
Iteration 582, loss: 0.46471163630485535
Iteration 583, loss: 0.49295946955680847
Iteration 584, loss: 0.46045252680778503
Iteration 585, loss: 0.4384380877017975
Iteration 586, loss: 0.4761814475059509
Iteration 587, loss: 0.5196762681007385
Iteration 588, loss: 0.4082944095134735
Iteration 589, loss: 0.562664270401001
Iteration 590, loss: 0.44146937131881714
Iteration 591, loss: 0.4451931416988373
Iteration 592, loss: 0.5638116598129272
Iteration 593, loss: 0.4668465256690979
Iteration 594, loss: 0.4378790855407715
Iteration 595, loss: 0.43066903948783875
Iteration 596, loss: 0.5240601301193237
Iteration 597, loss: 0.4704431891441345
Iteration 598, loss: 0.5595895051956177
Iteration 599, loss: 0.45854300260543823
Iteration 600, loss: 0.561322808265686
Iteration 601, loss: 0.4563794732093811
Iteration 602, loss: 0.48281505703926086
Iteration 603, loss: 0.47734153270721436
Iteration 604, loss: 0.37856972217559814
Iteration 605, loss: 0.42255374789237976
Iteration 606, loss: 0.39000898599624634
Iteration 607, loss: 0.43166834115982056
Iteration 608, loss: 0.424675315618515
Iteration 609, loss: 0.48792383074760437
Iteration 610, loss: 0.5467600226402283
Iteration 611, loss: 0.3876274824142456
Iteration 612, loss: 0.4239962697029114
Iteration 613, loss: 0.4425959885120392
Iteration 614, loss: 0.45432284474372864
Iteration 615, loss: 0.5306940078735352
Iteration 616, loss: 0.5329943299293518
Iteration 617, loss: 0.4367283582687378
Iteration 618, loss: 0.4498133659362793
Iteration 619, loss: 0.46252724528312683
Iteration 620, loss: 0.45820552110671997
Iteration 621, loss: 0.4768962562084198
Iteration 622, loss: 0.44655147194862366
Iteration 623, loss: 0.48336169123649597
Iteration 624, loss: 0.5230897665023804
Iteration 625, loss: 0.41584256291389465
Iteration 626, loss: 0.4938866198062897
Iteration 627, loss: 0.49654847383499146
Iteration 628, loss: 0.4970387816429138
Iteration 629, loss: 0.39577361941337585
Iteration 630, loss: 0.4797130227088928
Iteration 631, loss: 0.4332394599914551
Iteration 632, loss: 0.46040791273117065
```

```
Iteration 633, loss: 0.47353506088256836
Iteration 634, loss: 0.4706834554672241
Iteration 635, loss: 0.44337478280067444
Iteration 636, loss: 0.44540658593177795
Iteration 637, loss: 0.46098312735557556
Iteration 638, loss: 0.5508003234863281
Iteration 639, loss: 0.47532227635383606
Iteration 640, loss: 0.4579496681690216
Iteration 641, loss: 0.578161358833313
Iteration 642, loss: 0.4592892825603485
Iteration 643, loss: 0.46707403659820557
Iteration 644, loss: 0.4830702543258667
Iteration 645, loss: 0.46697738766670227
Iteration 646, loss: 0.4629068374633789
Iteration 647, loss: 0.3941839933395386
Iteration 648, loss: 0.41819053888320923
Iteration 649, loss: 0.4318258464336395
Iteration 650, loss: 0.4073052406311035
Iteration 651, loss: 0.5089078545570374
Iteration 652, loss: 0.45972639322280884
Iteration 653, loss: 0.5230918526649475
Iteration 654, loss: 0.4596474766731262
Iteration 655, loss: 0.509145975112915
Iteration 656, loss: 0.4675593674182892
Iteration 657, loss: 0.47878405451774597
Iteration 658, loss: 0.5284796357154846
Iteration 659, loss: 0.47787362337112427
Iteration 660, loss: 0.4238538146018982
Iteration 661, loss: 0.4509246051311493
Iteration 662, loss: 0.5102038979530334
Iteration 663, loss: 0.409471333026886
Iteration 664, loss: 0.49638479948043823
Iteration 665, loss: 0.5919950008392334
Iteration 666, loss: 0.43395155668258667
Iteration 667, loss: 0.4422321617603302
Iteration 668, loss: 0.41774553060531616
Iteration 669, loss: 0.4922780990600586
Iteration 670, loss: 0.4417787790298462
Iteration 671, loss: 0.4274269640445709
Iteration 672, loss: 0.5044046640396118
Iteration 673, loss: 0.5007553696632385
Iteration 674, loss: 0.4313114583492279
Iteration 675, loss: 0.5060082674026489
Iteration 676, loss: 0.43399474024772644
Iteration 677, loss: 0.4165359139442444
Iteration 678, loss: 0.48568564653396606
Iteration 679, loss: 0.5009620189666748
Iteration 680, loss: 0.5397014021873474
Iteration 681, loss: 0.46424606442451477
Iteration 682, loss: 0.43324825167655945
Iteration 683, loss: 0.47985151410102844
Iteration 684, loss: 0.42895013093948364
Iteration 685, loss: 0.48194414377212524
Iteration 686, loss: 0.43903636932373047
Iteration 687, loss: 0.47275686264038086
Iteration 688, loss: 0.5103411674499512
Iteration 689, loss: 0.43094733357429504
Iteration 690, loss: 0.4403078854084015
Iteration 691, loss: 0.4740965664386749
Iteration 692, loss: 0.44762250781059265
Iteration 693, loss: 0.49657365679740906
Iteration 694, loss: 0.4644930958747864
Iteration 695, loss: 0.47746676206588745
Iteration 696, loss: 0.5970909595489502
Iteration 697, loss: 0.44526612758636475
Iteration 698, loss: 0.4821227490901947
Iteration 699, loss: 0.5026988983154297
Iteration 700, loss: 0.5670831203460693
Iteration 701, loss: 0.47385916113853455
Iteration 702, loss: 0.5152519941329956
Iteration 703, loss: 0.43825483322143555
Iteration 704, loss: 0.4554094076156616
Iteration 705, loss: 0.47186899185180664
Iteration 706, loss: 0.4334597885608673
Iteration 707, loss: 0.4402013421058655
Iteration 708, loss: 0.4315657913684845
Iteration 709, loss: 0.4782262444496155
```

```
Iteration 710, loss: 0.41718485951423645
Iteration 711, loss: 0.46867263317108154
Iteration 712, loss: 0.39466947317123413
Iteration 713, loss: 0.476556658744812
Iteration 714, loss: 0.4237561821937561
Iteration 715, loss: 0.45303410291671753
Iteration 716, loss: 0.43639785051345825
Iteration 717, loss: 0.4499928057193756
Iteration 718, loss: 0.4618817865848541
Iteration 719, loss: 0.4507053792476654
Iteration 720, loss: 0.4481567144393921
Iteration 721, loss: 0.4451250433921814
Iteration 722, loss: 0.43442612886428833
Iteration 723, loss: 0.519943356513977
Iteration 724, loss: 0.518090546131134
Iteration 725, loss: 0.44066718220710754
Iteration 726, loss: 0.4583241045475006
Iteration 727, loss: 0.5193432569503784
Iteration 728, loss: 0.4544009864330292
Iteration 729, loss: 0.4600769579410553
Iteration 730, loss: 0.5162679553031921
Iteration 731, loss: 0.48490235209465027
Iteration 732, loss: 0.4600362181663513
Iteration 733, loss: 0.47571662068367004
Iteration 734, loss: 0.47049376368522644
Iteration 735, loss: 0.4198992848396301
Iteration 736, loss: 0.48035743832588196
Iteration 737, loss: 0.48168039321899414
Iteration 738, loss: 0.5433246493339539
Iteration 739, loss: 0.46412262320518494
Iteration 740, loss: 0.5309469103813171
Iteration 741, loss: 0.5226169228553772
Iteration 742, loss: 0.5068941712379456
Iteration 743, loss: 0.4978451430797577
Iteration 744, loss: 0.4391196370124817
Iteration 745, loss: 0.5137466788291931
Iteration 746, loss: 0.5043619275093079
Iteration 747, loss: 0.5599902868270874
Iteration 748, loss: 0.5040391683578491
Iteration 749, loss: 0.4495948553085327
Iteration 750, loss: 0.4983457922935486
Iteration 751, loss: 0.4421243369579315
Iteration 752, loss: 0.5266970992088318
Iteration 753, loss: 0.5193880796432495
Iteration 754, loss: 0.425741970539093
Iteration 755, loss: 0.51084965467453
Iteration 756, loss: 0.5035163164138794
Iteration 757, loss: 0.45231086015701294
Iteration 758, loss: 0.45038557052612305
Iteration 759, loss: 0.4563136100769043
Iteration 760, loss: 0.4274163842201233
Iteration 761, loss: 0.4319327473640442
Iteration 762, loss: 0.43246152997016907
Iteration 763, loss: 0.48416340351104736
Iteration 764, loss: 0.4425591826438904
Iteration 765, loss: 0.41287878155708313
Iteration 766, loss: 0.38945239782333374
Iteration 767, loss: 0.4530540108680725
Iteration 768, loss: 0.4437200725078583
Iteration 769, loss: 0.4051138758659363
Iteration 770, loss: 0.44042131304740906
Iteration 771, loss: 0.4290321469306946
Iteration 772, loss: 0.45971137285232544
Iteration 773, loss: 0.41939470171928406
Iteration 774, loss: 0.4738575220108032
Iteration 775, loss: 0.4858933985233307
Iteration 776, loss: 0.40748754143714905
Iteration 777, loss: 0.43827420473098755
Iteration 778, loss: 0.4636385142803192
Iteration 779, loss: 0.43128058314323425
Iteration 780, loss: 0.5283559560775757
Iteration 781, loss: 0.41093015670776367
Iteration 782, loss: 0.40602588653564453
Iteration 783, loss: 0.4423830807209015
Iteration 784, loss: 0.41326218843460083
Iteration 785, loss: 0.4022022783756256
Iteration 786, loss: 0.4192413091659546
```

```
Iteration 787, loss: 0.45003554224967957
Iteration 788, loss: 0.47266390919685364
Iteration 789, loss: 0.50911545753479
Iteration 790, loss: 0.42251336574554443
Iteration 791, loss: 0.44270017743110657
Iteration 792, loss: 0.38357898592948914
Iteration 793, loss: 0.4237861931324005
Iteration 794, loss: 0.37611058354377747
Iteration 795, loss: 0.450855553150177
Iteration 796, loss: 0.5114054083824158
Iteration 797, loss: 0.4222618639469147
Iteration 798, loss: 0.42353785037994385
Iteration 799, loss: 0.45665809512138367
Iteration 800, loss: 0.4165031611919403
Iteration 801, loss: 0.45036283135414124
Iteration 802, loss: 0.46866485476493835
Iteration 803, loss: 0.396745890378952
Iteration 804, loss: 0.4737938344478607
Iteration 805, loss: 0.45122987031936646
Iteration 806, loss: 0.44000786542892456
Iteration 807, loss: 0.4005368649959564
Iteration 808, loss: 0.47155624628067017
Iteration 809, loss: 0.4855249524116516
Iteration 810, loss: 0.4888263940811157
Iteration 811, loss: 0.3931560218334198
Iteration 812, loss: 0.4505190849304199
Iteration 813, loss: 0.39020779728889465
Iteration 814, loss: 0.47439196705818176
Iteration 815, loss: 0.46601879596710205
Iteration 816, loss: 0.3863501250743866
Iteration 817, loss: 0.47030550241470337
Iteration 818, loss: 0.4793294370174408
Iteration 819, loss: 0.5095840692520142
Iteration 820, loss: 0.44736334681510925
Iteration 821, loss: 0.5421078205108643
Iteration 822, loss: 0.4511737823486328
Iteration 823, loss: 0.40640416741371155
Iteration 824, loss: 0.4815562069416046
Iteration 825, loss: 0.4113156497478485
Iteration 826, loss: 0.4633575677871704
Iteration 827, loss: 0.4563363790512085
Iteration 828, loss: 0.4073048233985901
Iteration 829, loss: 0.466789186000824
Iteration 830, loss: 0.4389910399913788
Iteration 831, loss: 0.5536973476409912
Iteration 832, loss: 0.38051557540893555
Iteration 833, loss: 0.4290030896663666
Iteration 834, loss: 0.4181872010231018
Iteration 835, loss: 0.5440652966499329
Iteration 836, loss: 0.4409351646900177
Iteration 837, loss: 0.42292946577072144
Iteration 838, loss: 0.42648208141326904
Iteration 839, loss: 0.41455304622650146
Iteration 840, loss: 0.42466339468955994
Iteration 841, loss: 0.4537380337715149
Iteration 842, loss: 0.4269311726093292
Iteration 843, loss: 0.4259868860244751
Iteration 844, loss: 0.38573864102363586
Iteration 845, loss: 0.4343963861465454
Iteration 846, loss: 0.44829124212265015
Iteration 847, loss: 0.44861161708831787
Iteration 848, loss: 0.5421838164329529
Iteration 849, loss: 0.5207926034927368
Iteration 850, loss: 0.35398194193840027
Iteration 851, loss: 0.43467971682548523
Iteration 852, loss: 0.4251871407032013
Iteration 853, loss: 0.4490538239479065
Iteration 854, loss: 0.38761234283447266
Iteration 855, loss: 0.3826557397842407
Iteration 856, loss: 0.43512514233558917
Iteration 857, loss: 0.4911326766014099
Iteration 858, loss: 0.4911806285381317
Iteration 859, loss: 0.43983447551727295
Iteration 860, loss: 0.44404715299606323
Iteration 861, loss: 0.43993741273880005
Iteration 862, loss: 0.4700457453727722
Iteration 863, loss: 0.4710559844970703
```

```
Iteration 863, loss: 0.4710559844970703
Iteration 864, loss: 0.4528883695602417
Iteration 865, loss: 0.4473179578781128
Iteration 866, loss: 0.4387912452220917
Iteration 867, loss: 0.41039082407951355
Iteration 868, loss: 0.43675845861434937
Iteration 869, loss: 0.37202173471450806
Iteration 870, loss: 0.40200966596603394
Iteration 871, loss: 0.4330611526966095
Iteration 872, loss: 0.43553730845451355
Iteration 873, loss: 0.42006462812423706
Iteration 874, loss: 0.4500080943107605
Iteration 875, loss: 0.41395267844200134
Iteration 876, loss: 0.39581865072250366
Iteration 877, loss: 0.39973127841949463
Iteration 878, loss: 0.5434184670448303
Iteration 879, loss: 0.42232653498649597
Iteration 880, loss: 0.4159753918647766
Iteration 881, loss: 0.4046385884284973
Iteration 882, loss: 0.3958177864551544
Iteration 883, loss: 0.376852810382843
Iteration 884, loss: 0.40905603766441345
Iteration 885, loss: 0.4410563111305237
Iteration 886, loss: 0.42614108324050903
Iteration 887, loss: 0.4080106317996979
Iteration 888, loss: 0.3860209584236145
Iteration 889, loss: 0.41548824310302734
Iteration 890, loss: 0.4247642159461975
Iteration 891, loss: 0.4393032193183899
Iteration 892, loss: 0.4083010256290436
Iteration 893, loss: 0.4181317090988159
Iteration 894, loss: 0.48688599467277527
Iteration 895, loss: 0.3947857618331909
Iteration 896, loss: 0.39997848868370056
Iteration 897, loss: 0.41435083746910095
Iteration 898, loss: 0.4585551619529724
Iteration 899, loss: 0.36391326785087585
Iteration 900, loss: 0.4346119165420532
Iteration 901, loss: 0.40695542097091675
Iteration 902, loss: 0.38561245799064636
Iteration 903, loss: 0.34395188093185425
Iteration 904, loss: 0.483974814414978
Iteration 905, loss: 0.45933279395103455
Iteration 906, loss: 0.4167375862598419
Iteration 907, loss: 0.41299107670783997
Iteration 908, loss: 0.4685550332069397
Iteration 909, loss: 0.4083004593849182
Iteration 910, loss: 0.4446195065975189
Iteration 911, loss: 0.41355615854263306
Iteration 912, loss: 0.41116729378700256
Iteration 913, loss: 0.3663681745529175
Iteration 914, loss: 0.3874569237232208
Iteration 915, loss: 0.42584228515625
Iteration 916, loss: 0.397993266582489
Iteration 917, loss: 0.33673346042633057
Iteration 918, loss: 0.4415537714958191
Iteration 919, loss: 0.40949827432632446
Iteration 920, loss: 0.3825960159301758
Iteration 921, loss: 0.4064168334007263
Iteration 922, loss: 0.45628249645233154
Iteration 923, loss: 0.45038461685180664
Iteration 924, loss: 0.43037131428718567
Iteration 925, loss: 0.45520246028900146
Iteration 926, loss: 0.4065370559692383
Iteration 927, loss: 0.38350769877433777
Iteration 928, loss: 0.4702770411968231
Iteration 929, loss: 0.3453662693500519
Iteration 930, loss: 0.4525356590747833
Iteration 931, loss: 0.4402886927127838
Iteration 932, loss: 0.3562065660953522
Iteration 933, loss: 0.43700382113456726
Iteration 934, loss: 0.49446916580200195
Iteration 935, loss: 0.4238187372684479
Iteration 936, loss: 0.4591507613658905
Iteration 937, loss: 0.4521297216415405
Iteration 938, loss: 0.4089770019054413
Iteration 939, loss: 0.46351078152656555
Iteration 940, loss: 0.3873854875564575
```

```
Iteration 940, loss: 0.3873854875564575
Iteration 941, loss: 0.4694371819496155
Iteration 942, loss: 0.3744162917137146
Iteration 943, loss: 0.45482203364372253
Iteration 944, loss: 0.5299216508865356
Iteration 945, loss: 0.5228284001350403
Iteration 946, loss: 0.44800570607185364
Iteration 947, loss: 0.4100286662578583
Iteration 948, loss: 0.41317880153656006
Iteration 949, loss: 0.3549821972846985
Iteration 950, loss: 0.4222787022590637
Iteration 951, loss: 0.41734907031059265
Iteration 952, loss: 0.40172815322875977
Iteration 953, loss: 0.37745407223701477
Iteration 954, loss: 0.41142672300338745
Iteration 955, loss: 0.42366883158683777
Iteration 956, loss: 0.373458594083786
Iteration 957, loss: 0.4426725208759308
Iteration 958, loss: 0.38154375553131104
Iteration 959, loss: 0.4017448127269745
Iteration 960, loss: 0.4293513298034668
Iteration 961, loss: 0.34749382734298706
Iteration 962, loss: 0.3842560350894928
Iteration 963, loss: 0.43093040585517883
Iteration 964, loss: 0.44513261318206787
Iteration 965, loss: 0.47686073184013367
Iteration 966, loss: 0.4921889305114746
Iteration 967, loss: 0.34905657172203064
Iteration 968, loss: 0.3661159873008728
Iteration 969, loss: 0.43498244881629944
Iteration 970, loss: 0.44794410467147827
Iteration 971, loss: 0.3830909729003906
Iteration 972, loss: 0.3997810184955597
Iteration 973, loss: 0.41291528940200806
Iteration 974, loss: 0.3983769714832306
Iteration 975, loss: 0.4112418591976166
Iteration 976, loss: 0.44491884112358093
Iteration 977, loss: 0.4187615215778351
Iteration 978, loss: 0.37693876028060913
Iteration 979, loss: 0.4267154037952423
Iteration 980, loss: 0.36929208040237427
Iteration 981, loss: 0.38091257214546204
Iteration 982, loss: 0.43088021874427795
Iteration 983, loss: 0.39700719714164734
Iteration 984, loss: 0.41779422760009766
Iteration 985, loss: 0.4202190041542053
Iteration 986, loss: 0.4089076817035675
Iteration 987, loss: 0.4145519435405731
Iteration 988, loss: 0.31634217500686646
Iteration 989, loss: 0.398227721452713
Iteration 990, loss: 0.37161943316459656
Iteration 991, loss: 0.3986757695674896
Iteration 992, loss: 0.40357935428619385
Iteration 993, loss: 0.3743937909603119
Iteration 994, loss: 0.3253327012062073
Iteration 995, loss: 0.37630051374435425
Iteration 996, loss: 0.4159518778324127
Iteration 997, loss: 0.55936598777771
Iteration 998, loss: 0.35163599252700806
Iteration 999, loss: 0.35530751943588257
End of epoch 7, loss: 1.4969019889831543
Validation loss: 1.0703672170639038
```

In [43]:

```python
# Save the complete models

decoder.save('models/decoder')
encoder.save('models/encoder')

# decoder = tf.keras.models.load_model('models/decoder')
# encoder = tf.keras.models.load_model('models/encoder')
```

```
WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/tensorflow/python/training/tracking/trac
king.py:111: Model.state_updates (from tensorflow.python.keras.engine.training) is deprecated and will
```

```
king.py:111: Model.state_updates (from tensorflow.python.keras.engine.training) is deprecated and will
be removed in a future version.
Instructions for updating:
This property should not be used in TensorFlow 2.0, as updates are applied automatically.
```


```
WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/tensorflow/python/training/tracking/trac
king.py:111: Model.state_updates (from tensorflow.python.keras.engine.training) is deprecated and will
be removed in a future version.
Instructions for updating:
This property should not be used in TensorFlow 2.0, as updates are applied automatically.
```


```
WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/tensorflow/python/training/tracking/trac
king.py:111: Layer.updates (from tensorflow.python.keras.engine.base_layer) is deprecated and will be r
emoved in a future version.
Instructions for updating:
This property should not be used in TensorFlow 2.0, as updates are applied automatically.
```


```
WARNING:tensorflow:From /usr/local/lib/python3.6/dist-packages/tensorflow/python/training/tracking/trac
king.py:111: Layer.updates (from tensorflow.python.keras.engine.base_layer) is deprecated and will be r
emoved in a future version.
Instructions for updating:
This property should not be used in TensorFlow 2.0, as updates are applied automatically.
```


```
INFO:tensorflow:Assets written to: models/decoder/assets
```


```
INFO:tensorflow:Assets written to: models/decoder/assets
```


```
INFO:tensorflow:Assets written to: models/encoder/assets
```

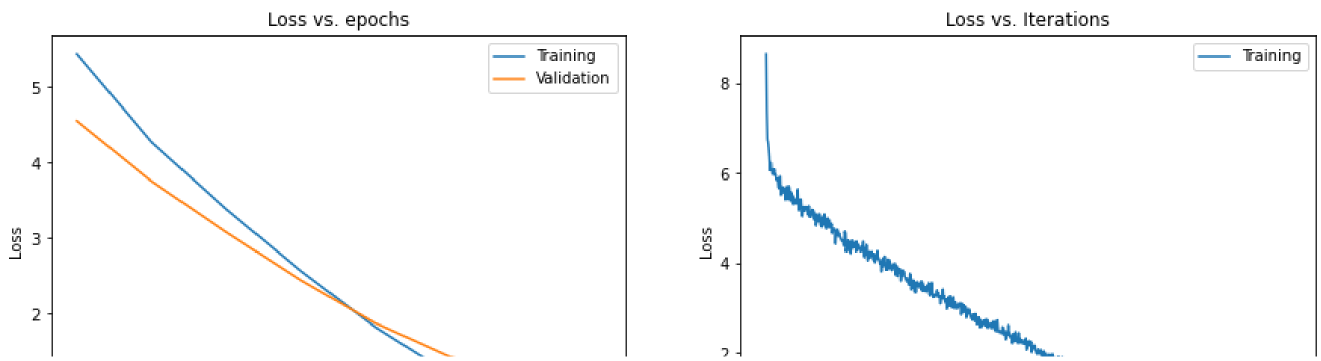

```
INFO:tensorflow:Assets written to: models/encoder/assets
```

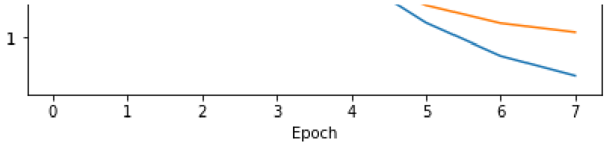

In [44]:

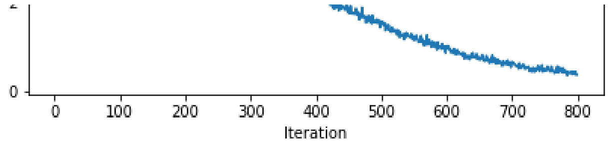
```python
# Plot the learning curves
import matplotlib.pyplot as plt
plt.figure(figsize=(15,5))

plt.subplot(121)
plt.plot(epoch_history['loss'])
plt.plot(epoch_history['val_loss'])
plt.title('Loss vs. epochs')
plt.ylabel('Loss')
plt.xlabel('Epoch')
plt.legend(['Training', 'Validation'], loc='upper right')

plt.subplot(122)
plt.plot(batch_history['loss'])
plt.title('Loss vs. Iterations')
plt.ylabel('Loss')
plt.xlabel('Iteration')
plt.legend(['Training'], loc='upper right')
plt.show()
```







7. Use the model to translate

Now it's time to put your model into practice! You should run your translation for five randomly sampled English sentences from the dataset. For each sentence, the process is as follows:

- Preprocess and embed the English sentence according to the model requirements.
- Pass the embedded sentence through the encoder to get the encoder hidden and cell states.
- Starting with the special `"<start>"` token, use this token and the final encoder hidden and cell states to get the one-step prediction from the decoder, as well as the decoder's updated hidden and cell states.
- Create a loop to get the next step prediction and updated hidden and cell states from the decoder, using the most recent hidden and cell states. Terminate the loop when the `"<end>"` token is emitted, or when the sentence has reached a maximum length.
- Decode the output token sequence into German text and print the English text and the model's German translation.

In [45]:

```
# Sample a random English sentence to translate

for _ in range(5):
    english_sentence = english[np.random.choice(len(english))]
    print("English input: \t\t", re.sub(r" ([?.!,])", r"\1 ", english_sentence))

    max_translation_length = 10

    german_translation = ["<start>"]
    ger_token = german_tokenizer.texts_to_sequences(german_translation)
    ger_token = np.array(ger_token)

    embedded_sentence = embedding_layer(tf.constant(english_sentence.split(' ')))

    hidden, cell = encoder(embedded_sentence[np.newaxis, ...])

    while True:
        decoder_output, hidden, cell = decoder(ger_token, hidden_state=hidden, cell_state=cell)
        ger_token = np.argmax(np.squeeze(decoder_output))
        if ger_token == 0:
            break
        ger_word = german_tokenizer.index_word[ger_token]
        ger_token = ger_token[np.newaxis, np.newaxis, ...]
        if ger_word == "<end>":
            break
        german_translation.append(ger_word)
        if len(german_translation) >= max_translation_length:
            break

    sentence = ' '.join(german_translation[1:])
    print("Model translation:\t", re.sub(r" ([?.!,])", r"\1 ", sentence))
    print('')
```

```
English input:     tom can't sing.
Model translation:  tom kann nicht zaehlen.

English input:     how perceptive!
Model translation:  wie aufregend!

English input:     could i ask why?
Model translation:  darf ich fragen,  warum?

English input:     what's it worth?
Model translation:  was ist das nun?

English input:     i'll be glad to.
Model translation:  ich werde brav sein.
```

In [ ]:

```

```

In [ ]:

```

```