A Recurrent Neural Network For Image Generation

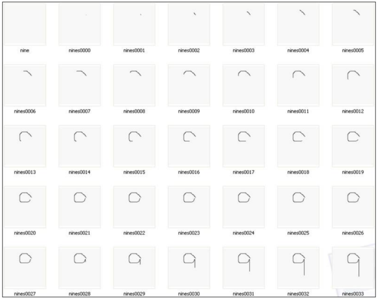
Mine Melodi Caliskan

CmpE597 - Project Proposal

March 18, 2018

Problem Statement

- Most approaches to automatic image generation, aim to generate entire scenes at once.
- The goal of this project is to reconstruct images step by step.



Deep Recurrent Attentive Writer (DRAW)

- DRAW reconstructs images, in which parts of a scene are created independently from others, and approximate sketches are successively refined.
- To model the phenomenon of working first on one part of the image, and then on another, an attention mechanism is used to restrict both the input region observed by the encoder, and the output region modified by the decoder.

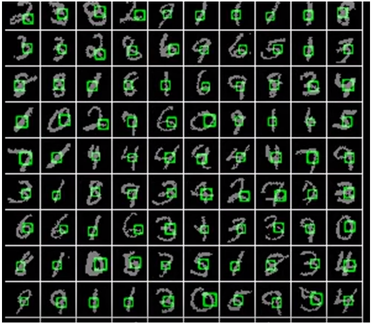
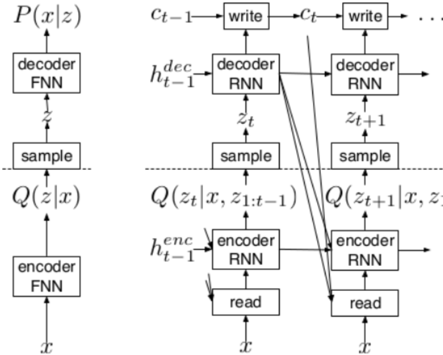


Figure 1: Reading MNIST

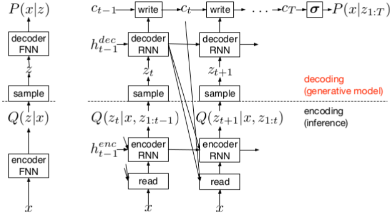
Figure 2: Generating MNIST

Basic recurrent architecture of DRAW

- The encoder-decoder-pair follows the design of variational autoencoders.
- The latent layer follows an n-dimensional gaussian distribution.
- The decoder receives a sample drawn from that gaussian distribution



Basic recurrent architecture of DRAW



- While the encoder reads from the input image, the decoder writes to an image canvas (where "write" is an addition, not a replacement of the old values).
- The model works in a fixed number of timesteps. At each timestep the encoder performs a read operation and the decoder a write operation.
- Both the encoder and the decoder receive the previous output of the encoder

Attention

The basic attention mechanism. (gx, gy) is the read patch center. delta is the strides. On the right: Patches with different sizes/strides and standard deviations/blurriness.

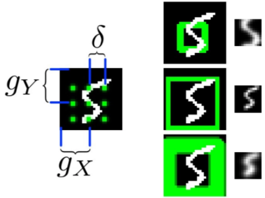


Figure 3. **Left:** A 3×3 grid of filters superimposed on an image. The stride (δ) and centre location (g_X, g_Y) are indicated. **Right:** Three $N \times N$ patches extracted from the image ($N = 12$). The green rectangles on the left indicate the boundary and precision (σ) of the patches, while the patches themselves are shown to the right. The top patch has a small δ and high σ, giving a zoomed-in but blurry view of the centre of the digit; the middle patch has large δ and low σ, effectively downsampling the whole image; and the bottom patch has high δ and σ.

Attention

- The selective read attention works on image patches of varying sizes. The result size is always NxN.
- The mechanism has the following parameters:
  - gx: x-axis coordinate of the center of the patch
  - gy: y-axis coordinate of the center of the patch

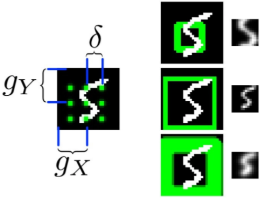


Figure 3. **Left:** A 3×3 grid of filters superimposed on an image. The stride (δ) and centre location (g_X, g_Y) are indicated. **Right:** Three $N \times N$ patches extracted from the image ($N = 12$). The green rectangles on the left indicate the boundary and precision (σ) of the patches, while the patches themselves are shown to the right. The top patch has a small δ and high σ, giving a zoomed-in but blurry view of the centre of the digit; the middle patch has large δ and low σ, effectively downsampling the whole image; and the bottom patch has high δ and σ.

Attention

The mechanism has the following parameters:

- delta: Strides. The higher the strides value, the larger the read image patch.
- sigma: Standard deviation. The higher the sigma value, the more blurry the extracted patch will be.
- gamma: Intensity-Multiplier. Will be used on the result.
- All of these parameters are generated using a linear transformation

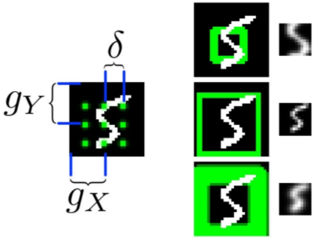
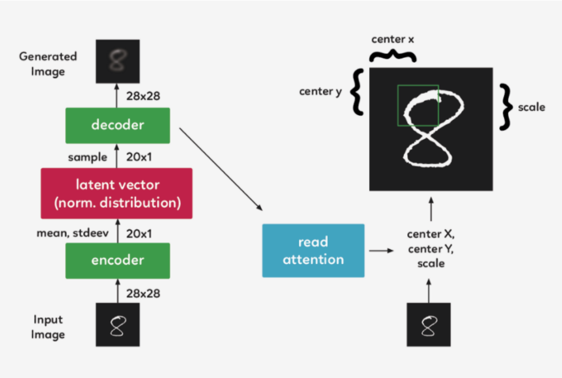


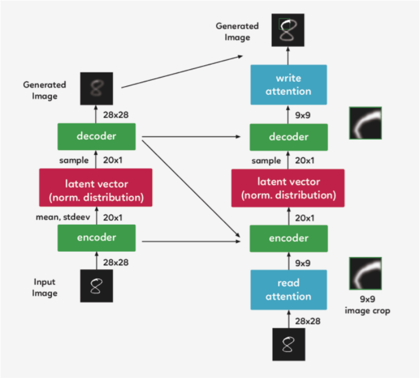
Figure 3. **Left:** A 3×3 grid of filters superimposed on an image. The stride (δ) and centre location (g_X, g_Y) are indicated. **Right:** Three $N \times N$ patches extracted from the image ($N = 12$). The green rectangles on the left indicate the boundary and precision (σ) of the patches, while the patches themselves are shown to the right. The top patch has a small δ and high σ, giving a zoomed-in but blurry view of the centre of the digit; the middle patch has large δ and low σ, effectively downsampling the whole image; and the bottom patch has high δ and σ.

Attention



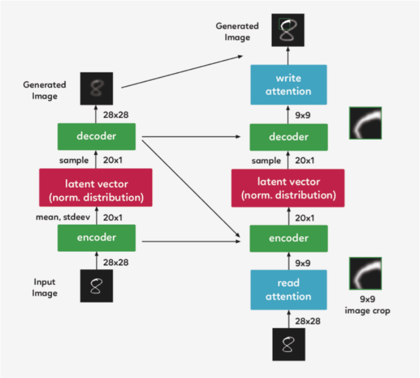
- The mechanism places a grid of NxN gaussians on the image. The grid is centered at (gx, gy). The gaussians are delta pixels apart from each other and have a standard deviation of sigma.
- Each gaussian is applied to the image, the center pixel is read and added to the result

Attention



- The previous time step's decoder hidden state is used. Using a simple fully-connected layer, the hidden state can be mapped to three parameters that represent the square crop: center x, center y, and the scale.

Attention



- The second part of the attention gate, the “write” attention, which have the same setup as the “read” section, except that the “write” attention gate uses the current decoder instead of the previous timestep’s decoder.

An Extension of DRAW

Deep Convolutional and Recurrent writer (DCRW).
The contributions of this extension are:

- The RNN in encoder has been replaced by CNN as Convolutional Neural Networks are considered to be state of the art in image processing in deep learning
- After replacing RNN with CNN the attention mechanism have been introduced into the Convolutional neural network.

References

- Karol Gregor, Ivo Danihelka, Alex Graves, Danilo Jimenez Rezende, Daan Wierstra "DRAW: A Recurrent Neural Network For Image Generation", 2015, Google DeepMind, arXiv:1502.04623
- Sadaf Gulshad, Jong-Hwan Kim. "Deep Convolutional and Recurrent Writer.", 2017 International Joint Conference on Neural Networks (IJCNN), DOI: 10.1109/IJCNN.2017.7966206
- Cup of Data, "DRAW: A Recurrent Neural Network For Image Generation." `http://blog.cupofdata.com/draw-a-recurrent-neural-network-for-image-generation-7`