Conditional Variational Auto-encoders for Sequential Data

Jianbo Chen, Billy Fang, Cheng Ju

Problem statement

Explore ways to generate sequential data

$$\begin{tikzcd} \text{VAE [1]} \arrow[r] \arrow[d] & \text{Conditional VAE [2],[3]} \arrow[d] \\ \text{DRAW [4]} \arrow[r] & \text{Conditional DRAW} \end{tikzcd}$$

Next step: Add dependencies between the latent random variables at neighboring time-steps [5] and derive its conditional form.

Dataset: MNIST digits from TensorFlow

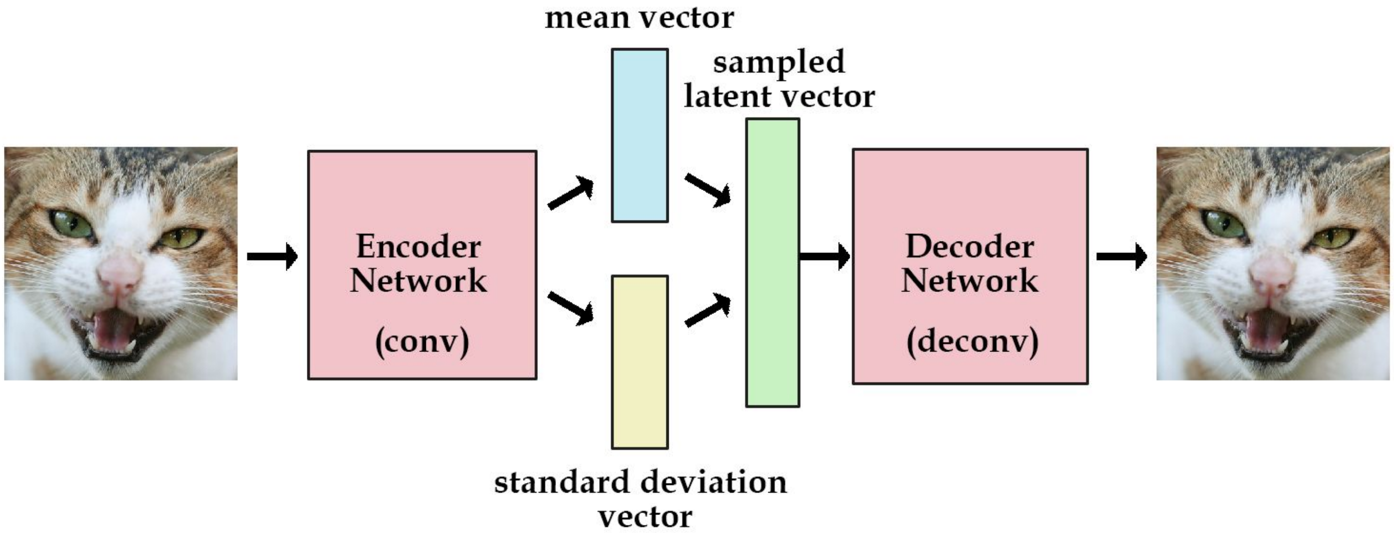
Idea behind Variational auto-encoder (VAE)

- Generative model:
  - Latent variable z ~ N(0,I)
  - x | z ~ N(f(z),I)
- Goal: Generate new examples x
- Maximize likelihood p(x). But p(x) is intractable to compute.
- Idea: Use Q(z | x)=N(m(x),Sigma(x)) to approximate p(z | x), which yields a lower bound on p(x):

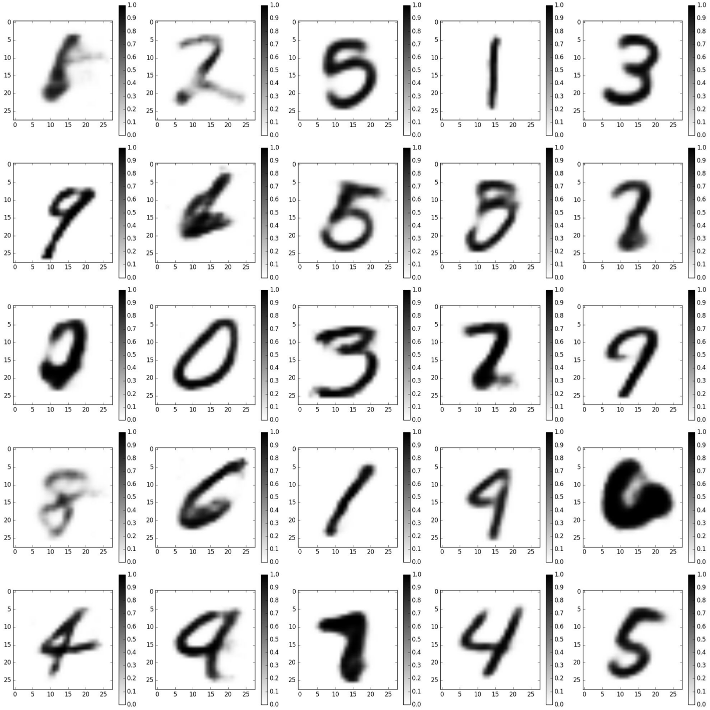
$$\log p(x) \geq \mathbb{E}_{z\sim Q(\cdot|x)}[\log p(x \mid z)] - \mathrm{KL}(Q(z \mid x)\|p(z)).$$

- Learn both Q(z | x) and p(x | z) by maximizing lower bound on likelihood p(x).

VAE model


mean vector
sampled
latent vector
Encoder
Network
(conv)
Decoder
Network
(deconv)
standard deviation
vector

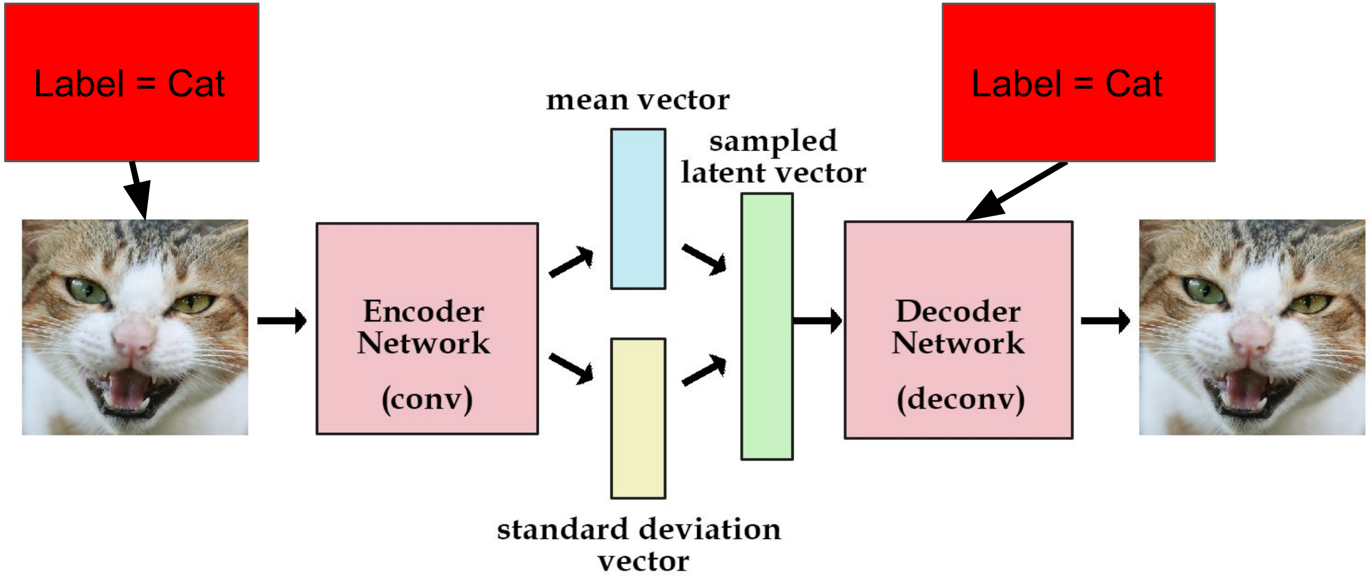
VAE-generated images



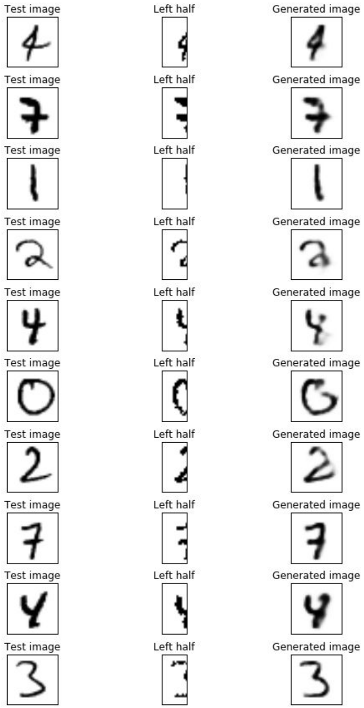
Conditional VAE (CVAE)

- Also condition on side information, e.g. partial images or labels

$$\log p(x \mid y) \geq \mathbb{E}_{z \sim Q(\cdot \mid y, x)}[\log p(x \mid y, z)] - \mathrm{KL}(Q(z \mid x, y) \| p(z \mid y)).$$



CVAE-generated images



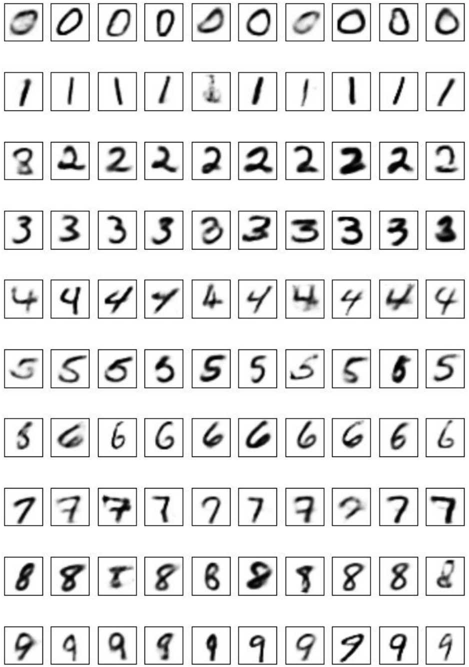
Semi-supervised VAE model

- Not all images are labeled
- Also models categorical distribution for labels
- Can be used for semi-supervised classification
- Labeled and unlabeled examples contribute to loss differently:

$$\log p(x, y) \geq \mathbb{E}_{z \sim Q(z|x,y)}[\log p(x \mid y, z) + \log p(y)] - \mathrm{KL}(Q(z \mid x, y) \| p(z)) =: -\mathcal{L}(x, y)$$

$$\log p(x) \geq \sum_y q(y \mid x)(-\mathcal{L}(x, y) + H(q(y \mid x)))$$

SSL VAE results



Labeled examples (out of 55000)	Validation error
10000	2.64%
5000	4.14%

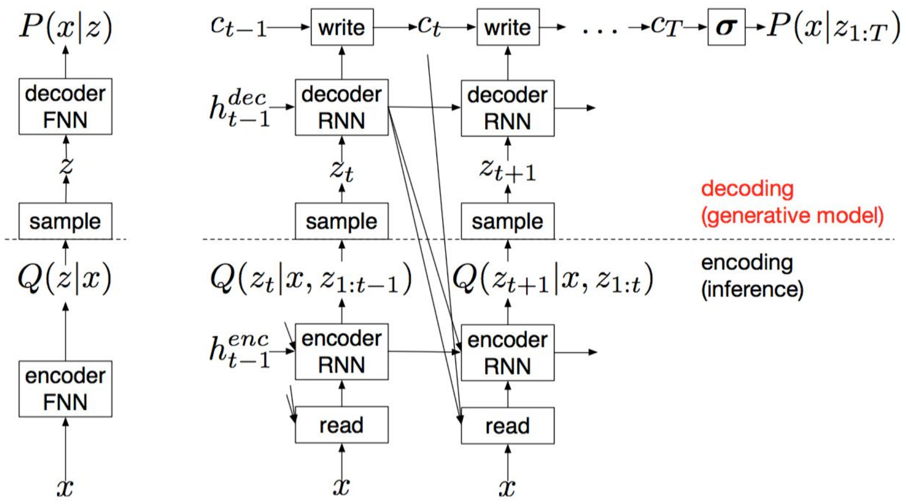
State of the art [3] does just as well with even fewer labeled examples

DRAW model

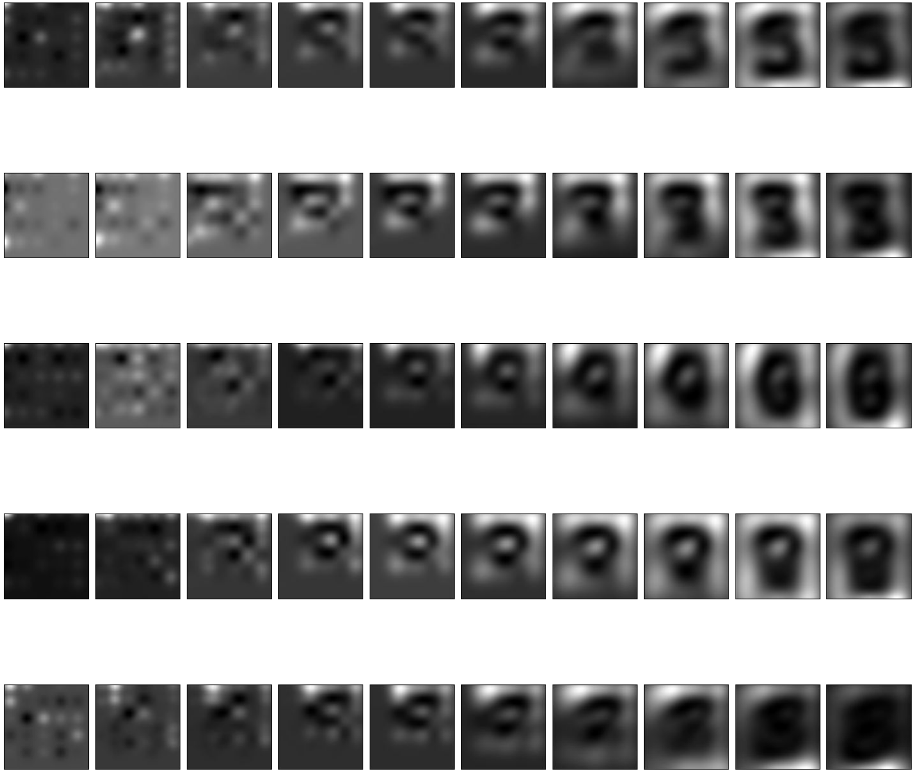
A recurrent version of variational autoencoder

Spatial Attention Mechanism mimics the foveation of the human eye

Sequential VAE framework that allows for the iterative construction of complex images



DRAW-generated images



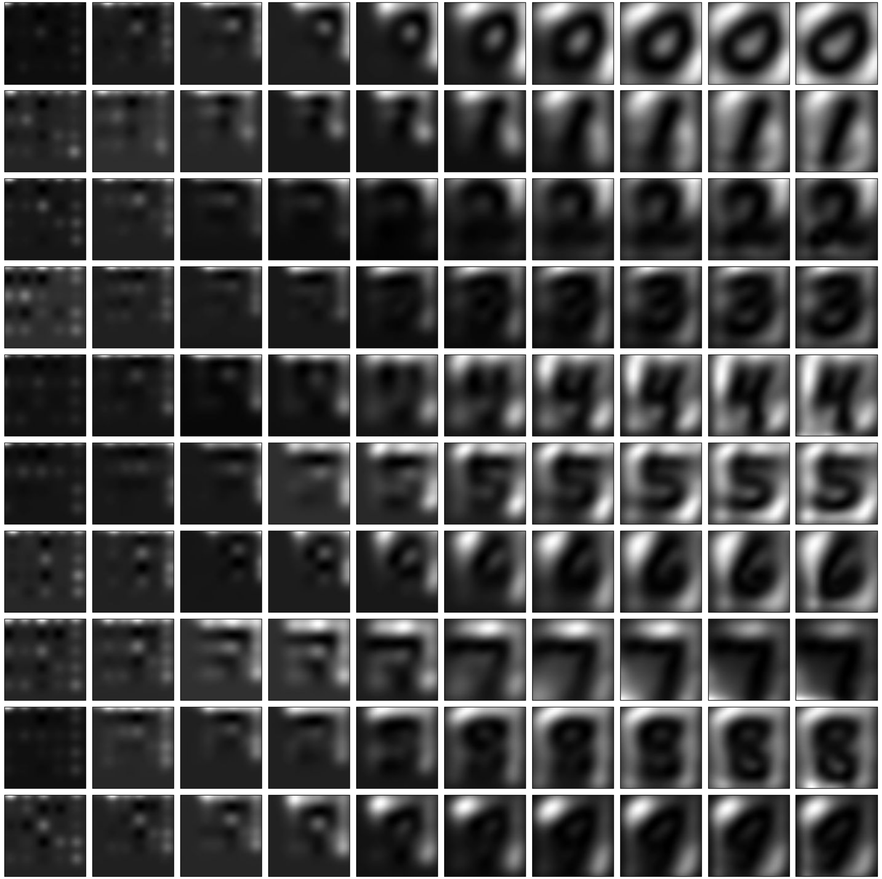
DRAW networks combine a novel spatial attention mechanism that mimics the foveation of the human eye.

Image in column t is generated by LSTM in time t.

Each row shows how DRAW sequentially generates an image

Conditional DRAW

- Add label information to decoder and encoder to mimic the structure of the CVAE
- Input: random normal variable z; an integer between 0 and 9.
- Figure on the right: each row shows how DRAW sequentially generates an image, given a digit as input



Looking forward

- We hope to move toward generating sequential data, such as speech and handwriting.
- Add dependencies between the latent random variables at neighboring time-steps [5] and derive its conditional form.

Tools

- Tensorflow
- GeForce GTX 770

References

[1] Kingma, Welling. Auto-Encoding Variational Bayes

[2] Sohn, Yan, Lee. Learning Structured Output Representation using Deep Conditional Generative Models

[3] Kingma, Rezende, Mohamed, Welling. Semi-Supervised Learning with Deep Generative Models

[4] Gregor, Danihelka, Graves, Rezende, Wierstra. DRAW: A Recurrent Neural Network For Image Generation

[5] Chung, Kastner, Dinh, Goel, Courville, Bengio. A Recurrent Latent Variable Model for Sequential Data