Variational Auto-encoders: Image representations for generation and classification

Jianbo Chen, Billy Fang, Cheng Ju
Team name: Chen/Ju/Fang

Outline: VAE models/applications that we explored

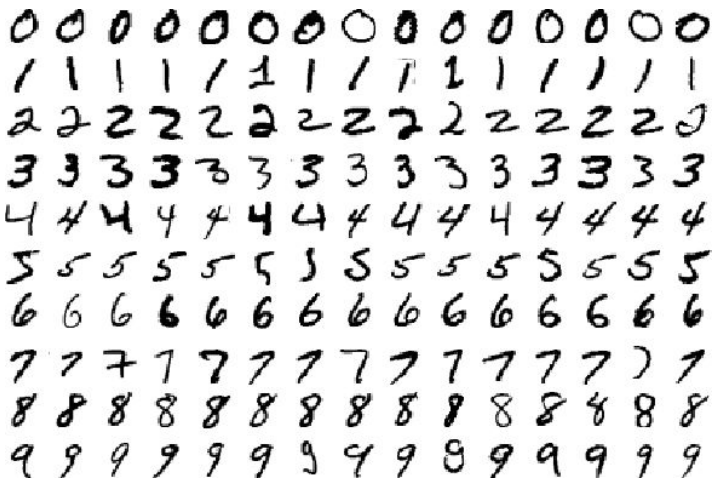
- Variational Auto-encoder (VAE): generation and “encoding”
- Conditional VAE (CVAE): leveraging side information / labels
- Semi-supervised learning (SSL) VAE: classification with limited labeling
- [C]VAE with Generative Adversarial Networks (GAN): better generation
- DRAW VAE (attention-based generation)
- Deep feature: ‘deeper’ reconstruction loss

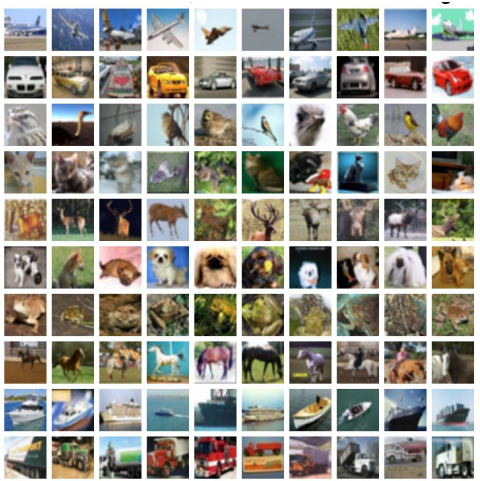
Problem statements:

- Generating “good” new examples of images from learned data (qualitative)
- Semi-supervised learning: classification accuracy

Data

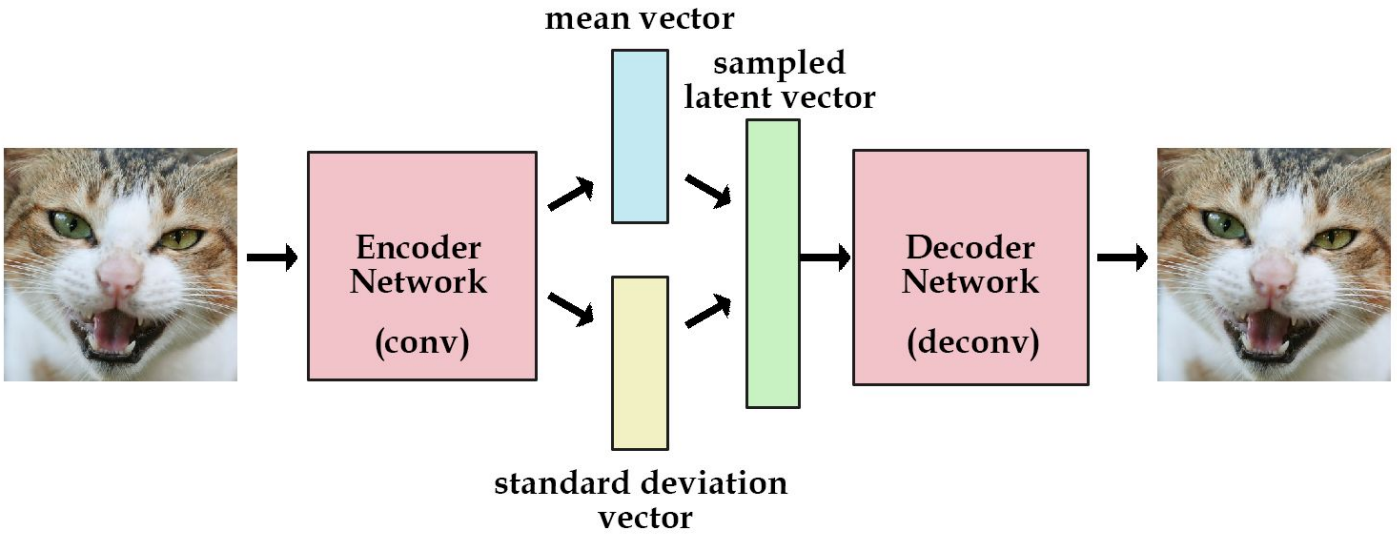
- MNIST handwritten digits (via TensorFlow)
- Street View House Numbers (SVHN)
- CIFAR-10 data





VAE model

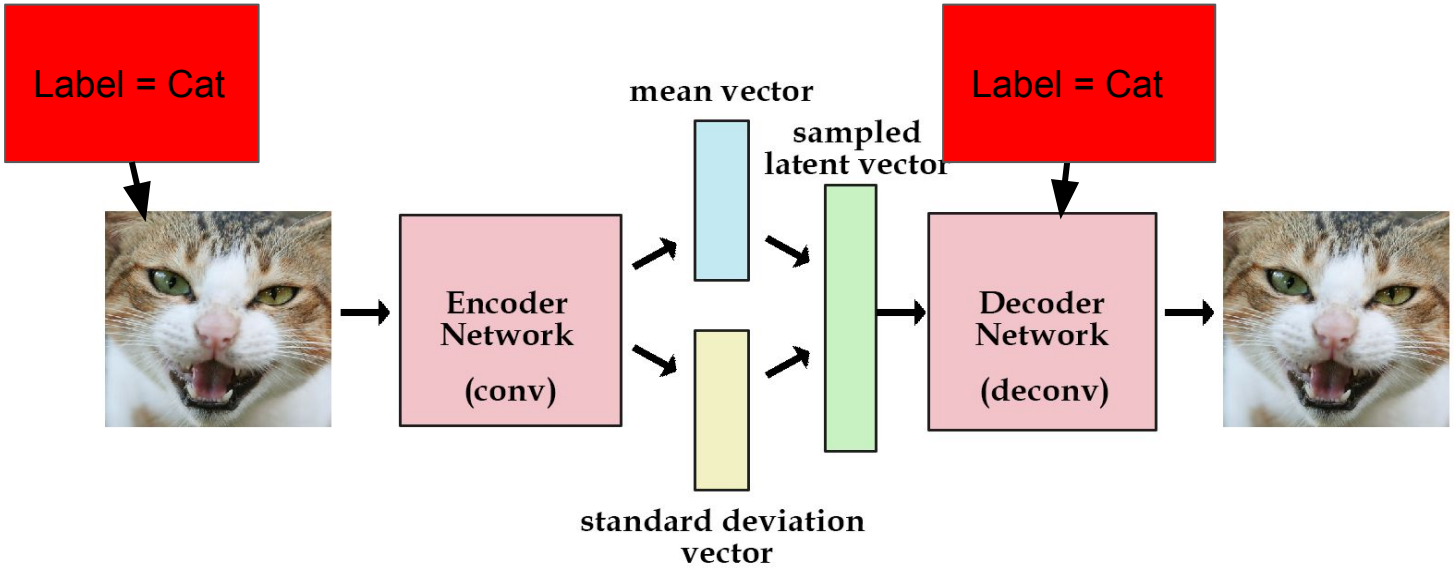
$$\log p(x) \geq \mathbb{E}_{z \sim Q(\cdot|x)}[\log p(x \mid z)] - \mathrm{KL}(Q(z \mid x) \| p(z)).$$



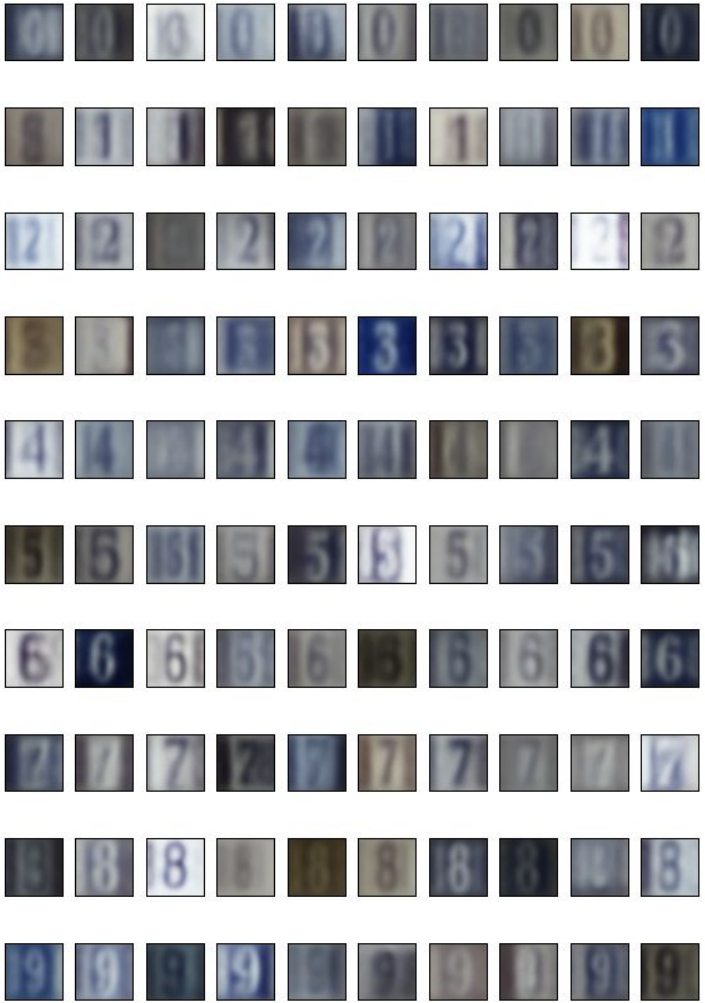
Conditional VAE (CVAE)

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

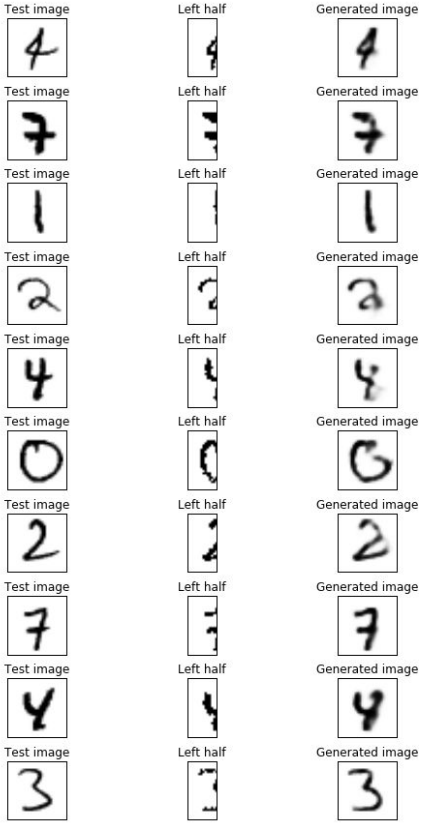
$$\log p(x \mid y) \geq \mathbb{E}_{z \sim Q(\cdot \mid y,x)}[\log p(x \mid y,z)] - \mathrm{KL}(Q(z \mid x,y) \| p(z \mid y)).$$

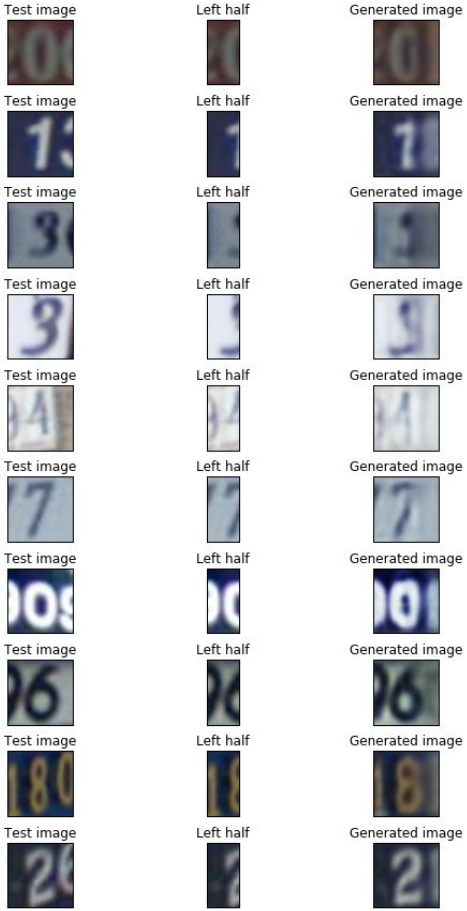


CVAE for digit generation

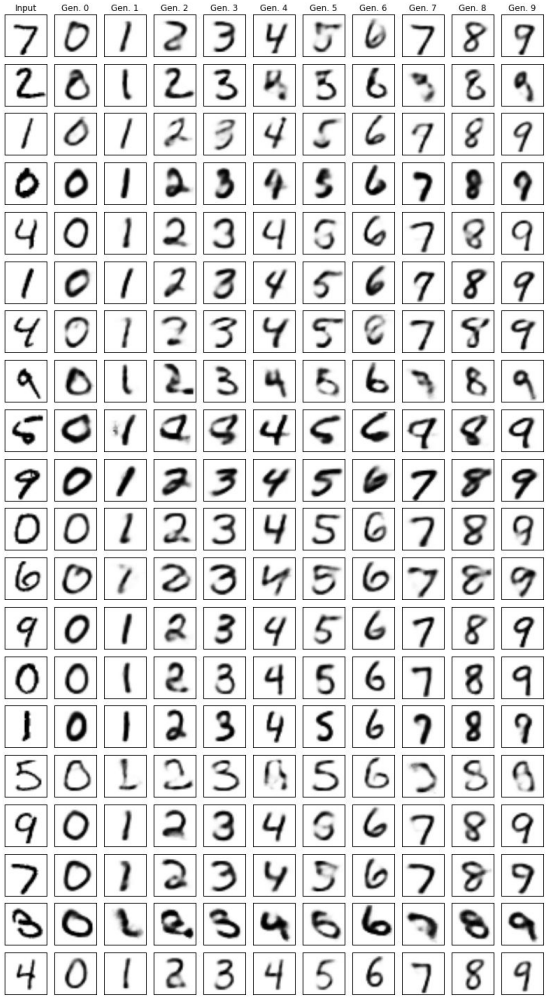


CVAE for image completion





CVAE for style mimicking



Semi-supervised VAE model

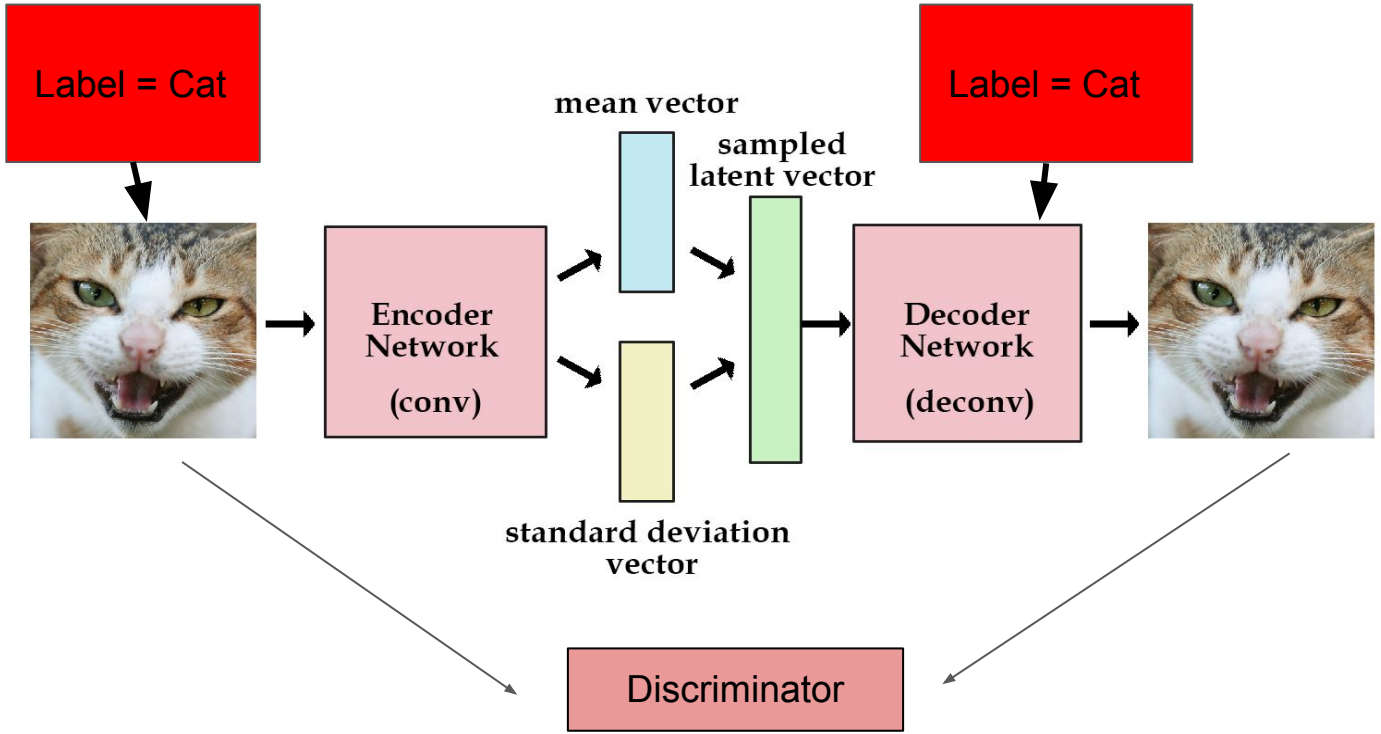
- Not all images are labeled
- Also models categorical distribution for labels
- Can be used for semi-supervised classification
- Labeled and unlabeled examples contribute to loss differently:

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

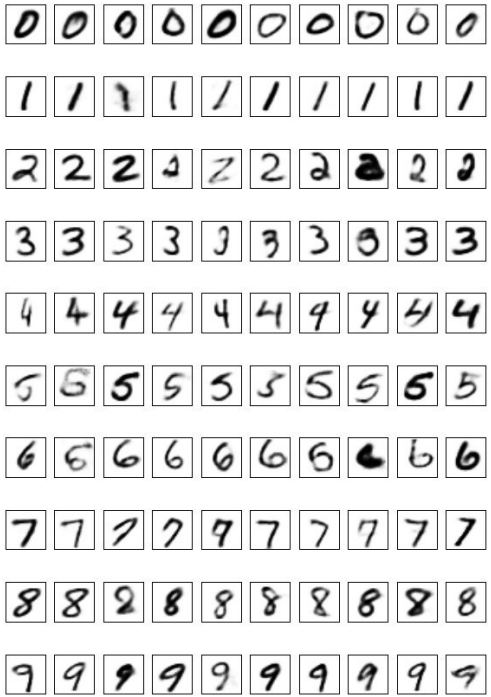
$$\log p(x) \geq \sum_{y} q(y \mid x)(-\mathcal{L}(x, y) + H(q(y \mid x)))$$

Validation / Test error			
Labeled MNIST examples (out of 55000)	Fully connected	Convolutional	Kingma et al.
1000	4.7% / 5.1%	4.2% / 4.8%	2.4%
600	11.5% / 12%	7.0% / 7.2%	2.6%

(C)VAEGAN: adding an adversarial discriminator



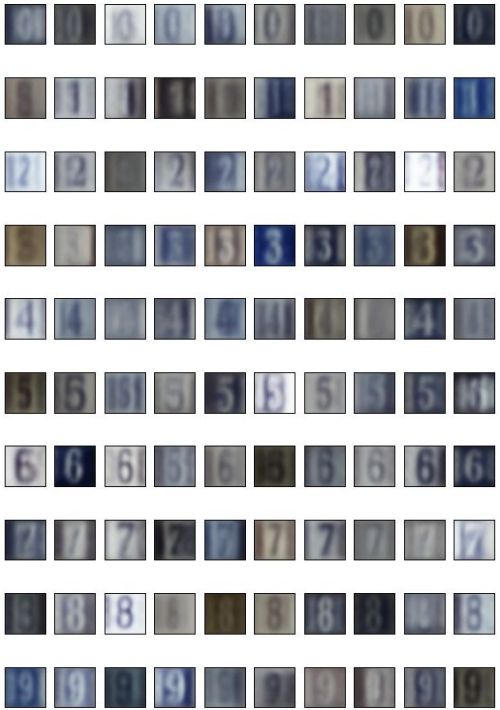
CVAEGAN: sharper images



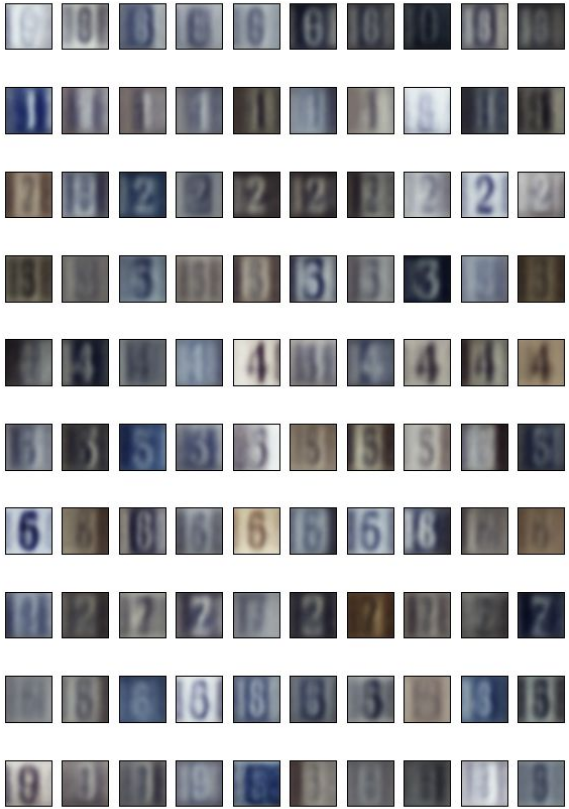
CVAE

CVAEGAN

CVAEGAN: sharper images?



CVAE



CVAEGAN

CVAEGAN: sharper images?

CVAE

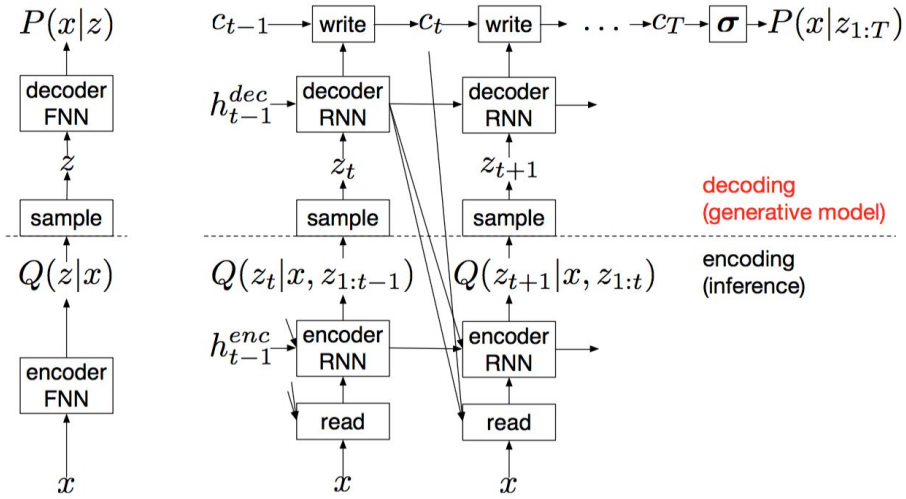
CVAEGAN

DRAW model

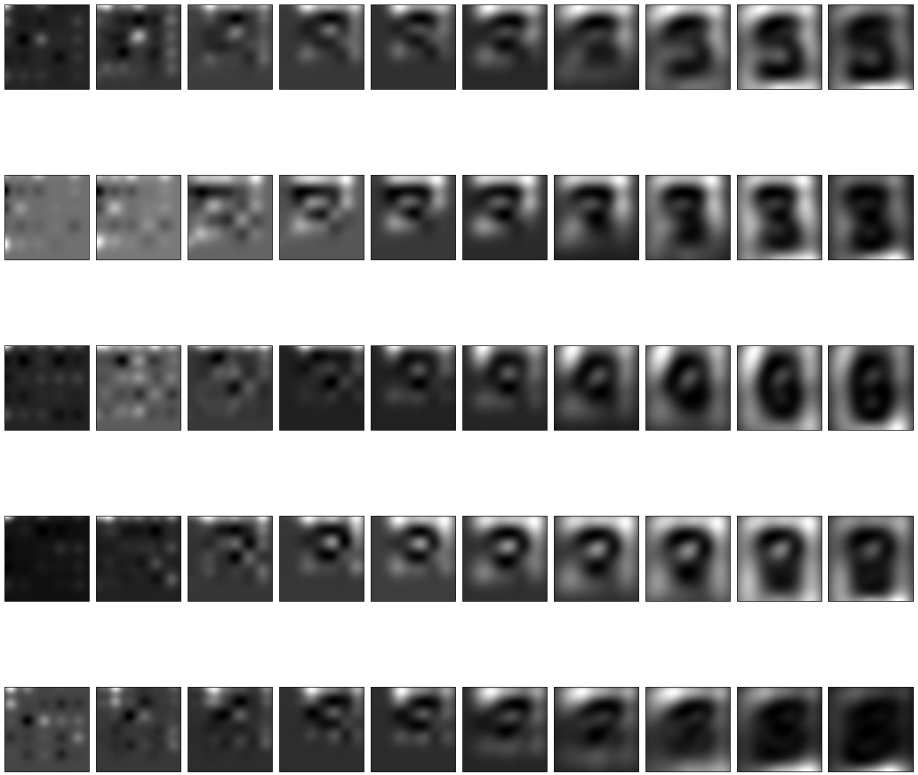
A recurrent version of variational autoencoder

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



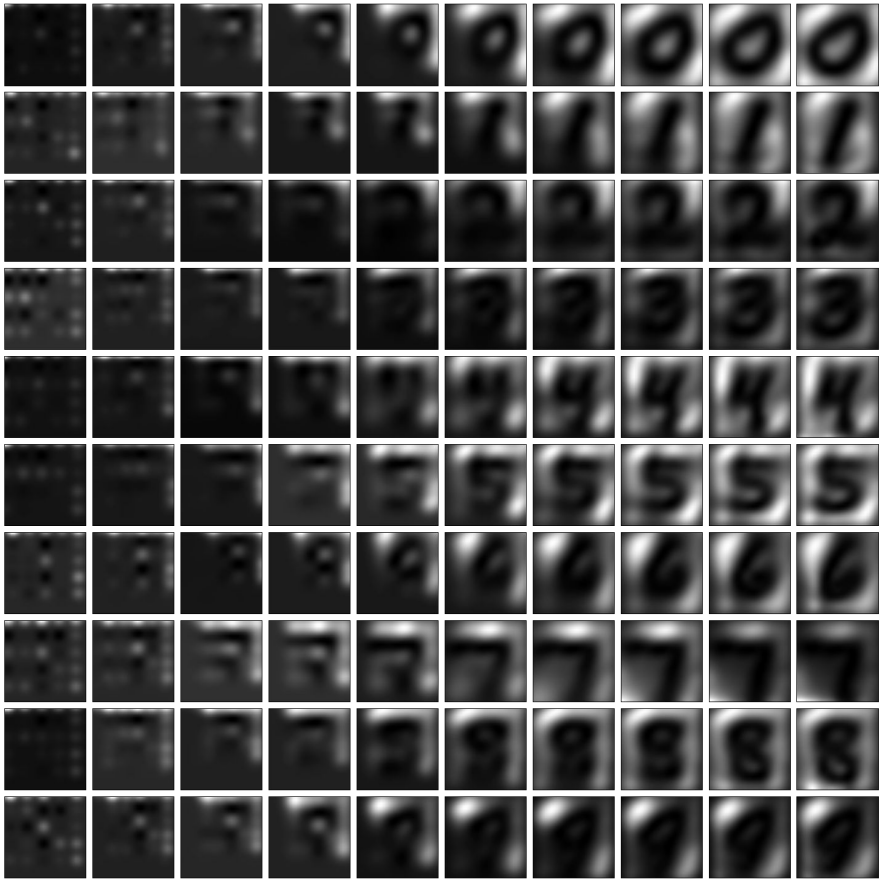
DRAW networks combine a novel spatial attention mechanism that mimics the foveation of the human eye.

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



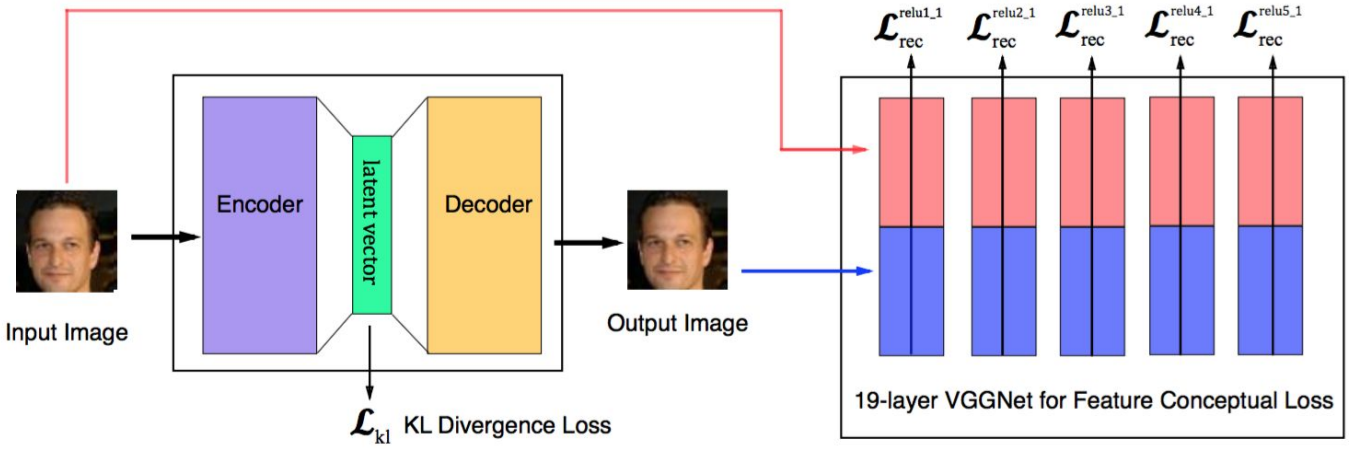
Deep Feature Consistent Variational Autoencoder

Feature Perceptual Loss:

$$\mathcal{L}_{rec}^{l} = \frac{1}{2C^{l}W^{l}H^{l}} \sum_{c=1}^{C^{l}} \sum_{w=1}^{W^{l}} \sum_{h=1}^{H^{l}} (\Phi(x)_{c,w,h}^{l} - \Phi(\bar{x})_{c,w,h}^{l})^{2}$$

1. Pre-train a Deep Network for prediction task (e.g. VGG net).
2. Build VAE with Feature Perceptual Loss: instead of the reconstruction loss, we put the original and generated image into the pre-trained net, and compute the loss for the feature maps (deep features).
3. Note: need to fix the weight for the pre-trained when training VAE.

Network Structure for Deep Feature VAE



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

[6] Hou, Xianxu, et al. "Deep Feature Consistent Variational Autoencoder." arXiv preprint arXiv:1610.00291 (2016).