When Variational Auto-encoders meet Generative Adversarial Networks

JIANBO CHEN, BILLY FANG, CHENG JU *Departments of Statistics and Biostatistics, UC Berkeley* CS 294-129, FALL 2016

PROBLEMS

- Various generative model of images.
- Various applications including semi-supervised learning.

BACKGROUNDS AND OBJECTIVES

- Variational auto-encoders (VAE) use variational inference to learn intractable posterior distributions.
- VAE's drawbacks: Images tend to be blurry.
- Generative adversarial networks (GAN) learn a generative model based on game theory.
- GAN's drawbacks: Difficult to train; No explicit representation of the generative distribution $p_g(x)$.
- Intermediate Objectives: Investigate various types of VAE models.
- Ultimate Goal: Combine VAE and GAN to get 'better' generative models.

VARIATIONAL AUTO-ENCODER

Variational auto-encoder (VAE) [4]

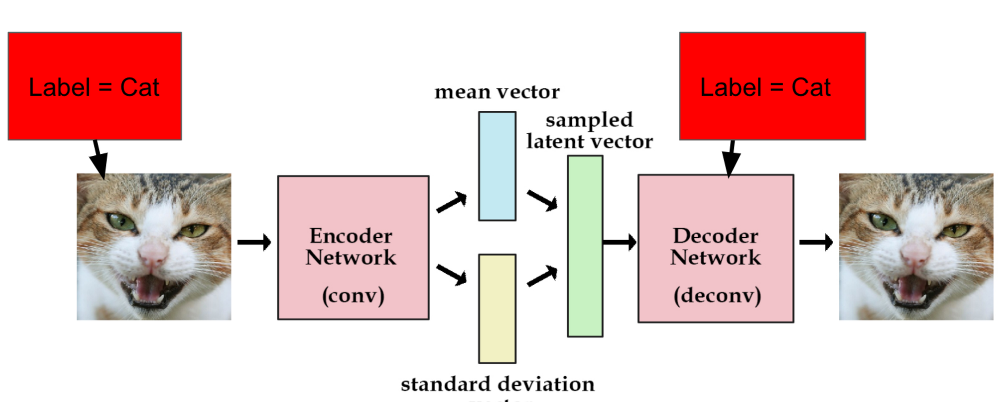
- Latent variable model: $z \sim \mathcal{N}(0, I)$, $x \mid z \sim f(x; z, \theta)$ (e.g. Bernoulli)
- Variational inference: maximize lower bound on log likelihood

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

Conditional VAE (CVAE) [8]

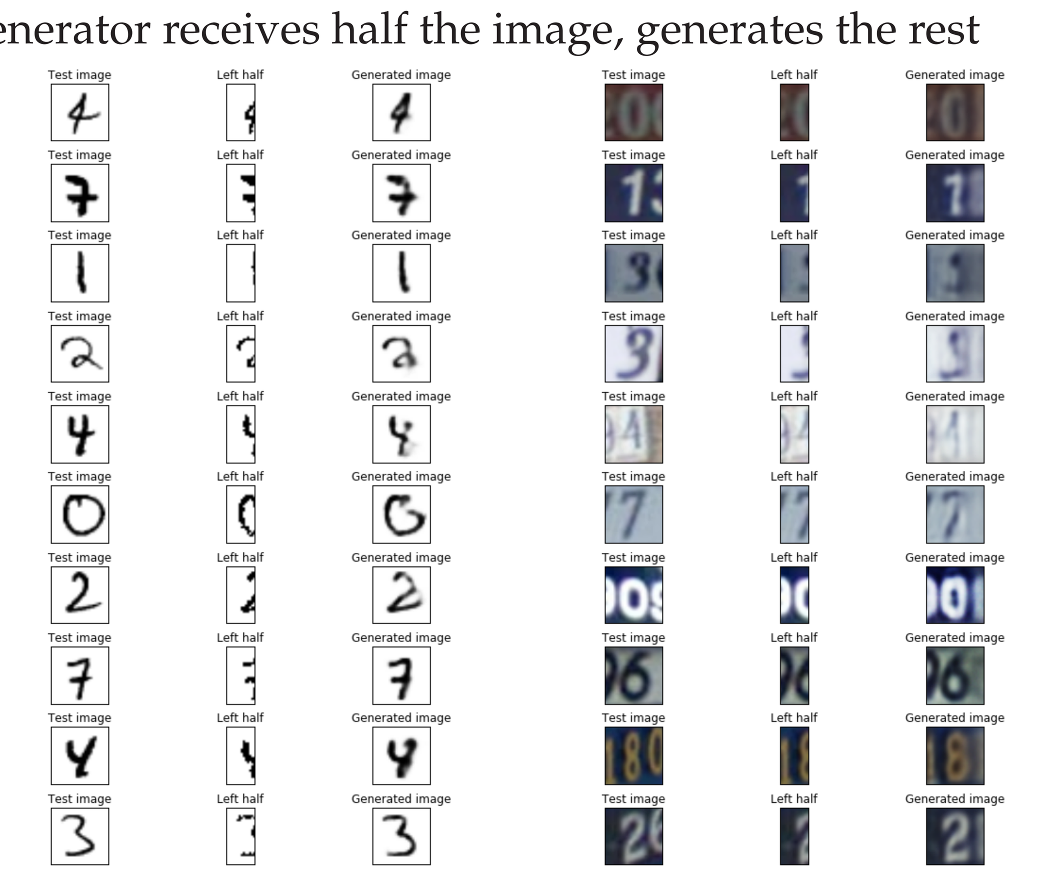
- Condition everything on side information y (label, partial image, etc.)

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



CVAE FOR IMAGE COMPLETION

Generator receives half the image, generates the rest



CVAE FOR STYLE TRANSFER

Transferring the style of the first column to all digits

ADDING GANs

- Add discriminator D to encourage sharpness
- Replace the pixel-wise VAE reconstruction loss by learned feature representations in GAN discriminator. [5]
- Objective for the generator (VAE) G:

$$\mathcal{L} = \mathcal{L}_{prior} + \mathcal{L}_{REC} + \mathcal{L}_{GAN},$$

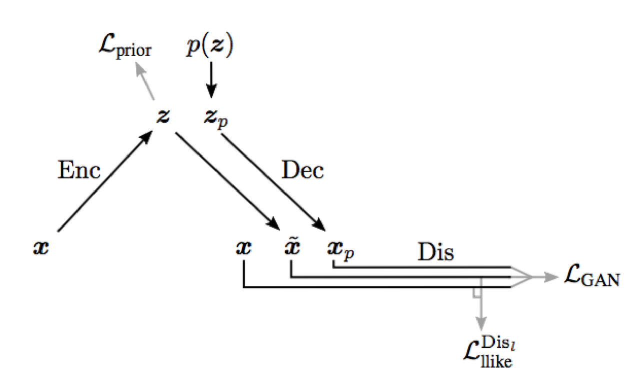
where

$$\mathcal{L}_{prior} = KL(p_z(z), q(z|x)),$$
$$\mathcal{L}_{REC} = -\mathbb{E}_{q(Z|X)}[p(Dis_l(X)|Z)],$$
$$\mathcal{L}_{GAN} = -\mathbb{E}_{z \sim p_z(z)}[\log D(G(z))].$$

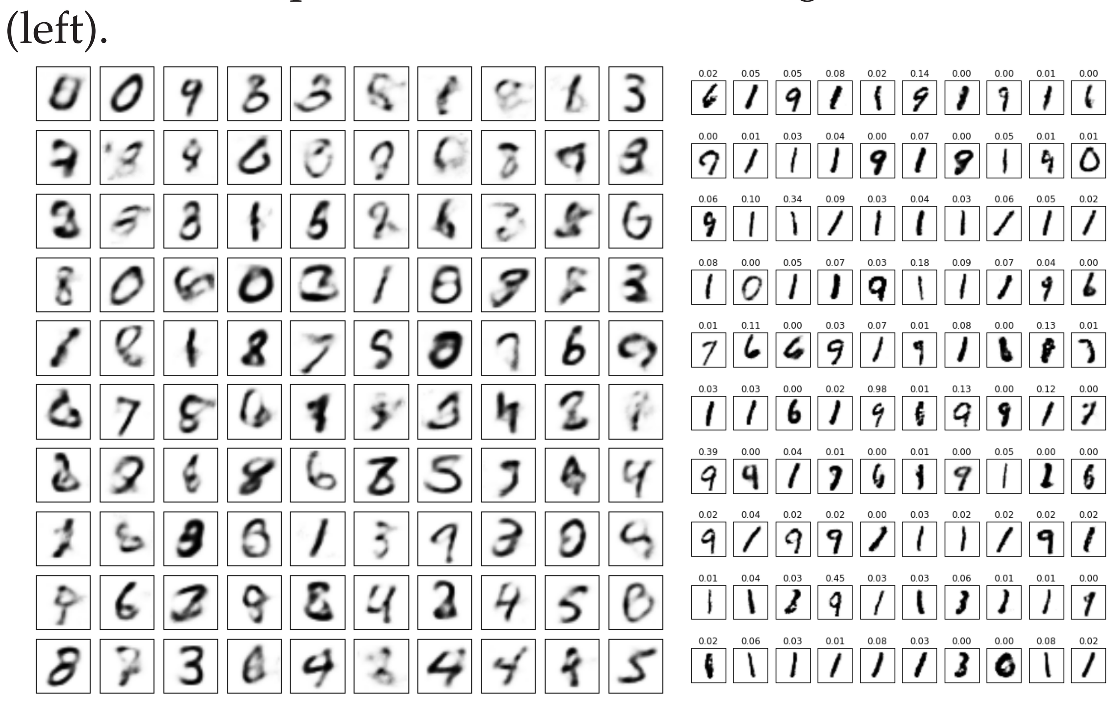
- Objective for the discriminator D:

$$-\mathbb{E}_{x \sim p_{data}}[\log D(x)] - \mathbb{E}_{z \sim p_z(z)}[\log(1 - D(G(z)))].$$

- In our implementations we use the DCGAN architecture [7]



Generated output of VAE with GAN (right) and without (left).



CVAE WITH CGAN

- Add a conditional GAN on top of a conditional VAE.
- A conditional GAN has a discriminator assigning probability to real images with different labels and fake images as a single class respectively. [6]

SEMI-SUPERVISED LEARNING (SSL) VAE

- Handle datasets with missing labels
- Models label distribution
- Labeled and unlabeled examples enter loss differently: see $\mathcal{L}(x, y)$ and $\mathcal{U}(x)$ below
- Use the encoder as a classifier.

Labeled and unlabeled loss, respectively:

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

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

Validation/test error on MNIST (55000 training examples)

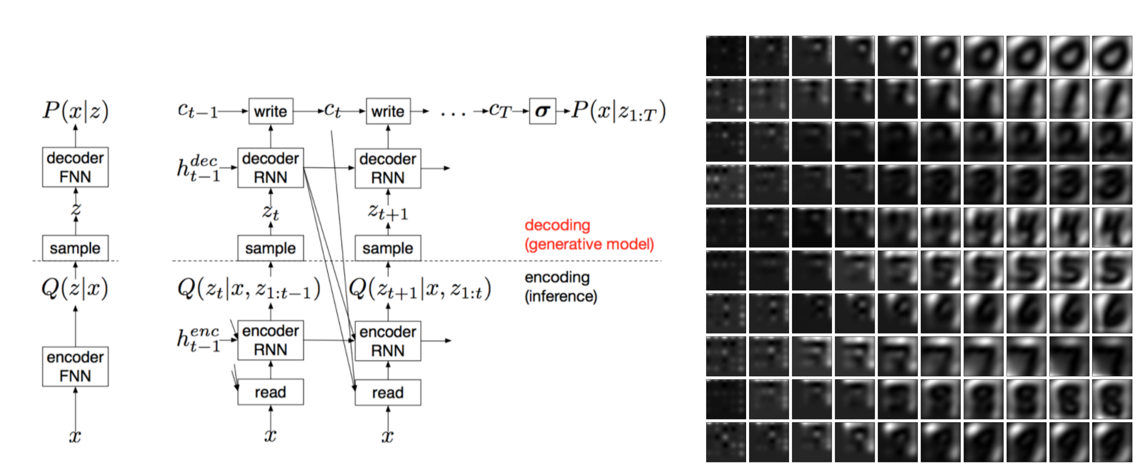
	1000 labeled	600 labeled
Fully connected	4.7%/ 5.1%	11.5%/12.0%
Convolutional	4.2%/4.8%	6.0%/6.2%
Kingma et al. [3]	2.4%	2.6%

SSL WITH GANs

- Add a conditional GAN on top of a SSL VAE.
- Both the discriminator of CGAN and the encoder of SSL VAE can be used as a classifier. We used the latter.
- Performance: Currently similar to SSL VAE.

DRAW

- Attention-based sequential generation
- RNN structure
- Generalization 1: Condition DRAW on labels (output appears below).
- Generalization 2: Taking DRAW as a black-box VAE, add GAN on top of it.



FUTURE DIRECTIONS

- Improve our GAN-based models for
  - VAE (generating more diverse digits)
  - SSL (better classification accuracy)
  - DRAW (better generation)
- Consider alternate classifiers for SSL with GAN: encoder, discriminator, or some ensemble?
- Investigate assessment of generative models.

REFERENCES

[1] Carl Doersch. Tutorial on variational autoencoders.

[2] Karol Gregor, Ivo Danihelka, Alex Graves, Danilo Jimenez Rezende, and Daan Wierstra. DRAW: A recurrent neural network for image generation.

[3] Diederik P Kingma, Shakir Mohamed, Danilo Jimenez Rezende, and Max Welling. Semi-supervised learning with deep generative models.

[4] Diederik P Kingma and Max Welling. Auto-encoding variational bayes.

[5] Anders Boesen Lindbo Larsen, Søren Kaae Sønderby, and Ole Winther. Autoencoding beyond pixels using a learned similarity metric.

[6] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets.

[7] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks.

[8] Kihyuk Sohn, Honglak Lee, and Xinchen Yan. Learning structured output representation using deep conditional generative models.