Summary of Recent (2018) Conditional Generative Models

Presentor: Ji Gao

Department of Computer Science, University of Virginia
https://qdata.github.io/deep2Read/

Outline

- Engel, J., Hoffman, M., & Roberts, A. (2017). Latent constraints: Learning to generate conditionally from unconditional generative models. arXiv preprint arXiv:1711.05772.

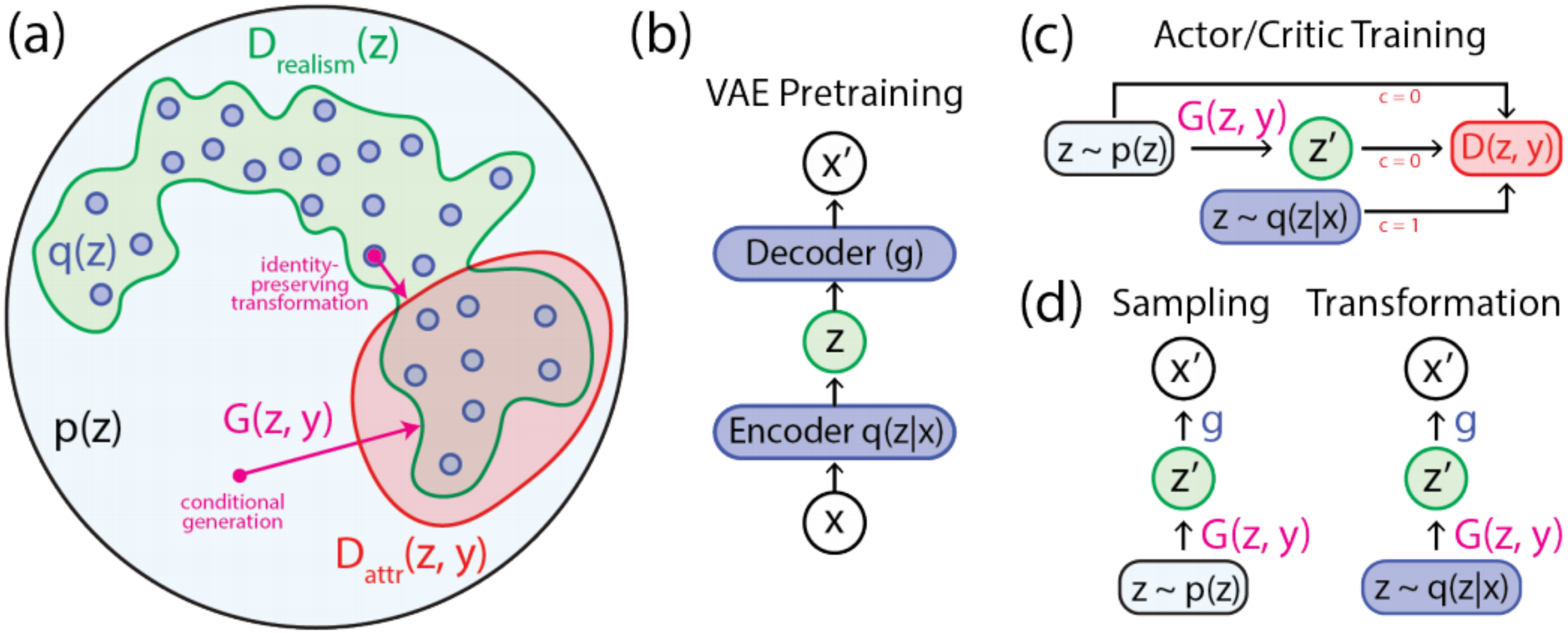
Latent Constraints: Learning to Generate Conditionally from Unconditional Generative Models

- Use additional training on generator model to sharpen it, make it conditioning on label
- Report to have better performance than CVAE and CGAN

Previous approaches

- CGAN and CVAE can generate samples conditioned on attribute information when available, but
- 1. they train with knowledge of the attribute labels for the whole training set, and it is not clear how to adapt them to new attributes without retraining from scratch
- 2. CGANs and CVAEs suffer from the same problems of mode-collapse and blurriness as their unconditional cousins

Process



(a)
$D_{realism}(z)$
q(z)
identity-preserving transformation
p(z)
G(z, y)
conditional generation
$D_{attr}(z, y)$
(b)
VAE Pretraining
x′
Decoder (g)
z
Encoder q(z|x)
x
(c)
Actor/Critic Training
z ~ p(z)
G(z, y)
z′
c = 0
D(z, y)
z ~ q(z|x)
c = 1
(d)
Sampling
Transformation
x′
g
z′
G(z, y)
z ~ p(z)
z ~ q(z|x)

Critic: Use CGAN on VAE

- Critic Loss [GAN]

$$\mathcal{L}_D(z) = \mathbb{E}_{z\sim q(z|x)}[\mathcal{L}_{c=1}(z)] + \mathbb{E}_{z\sim p(z)}[\mathcal{L}_{c=0}(z)] + \mathbb{E}_{z\sim G(p(z))}[\mathcal{L}_{c=0}(z)]$$

Experiment – Identity perserving

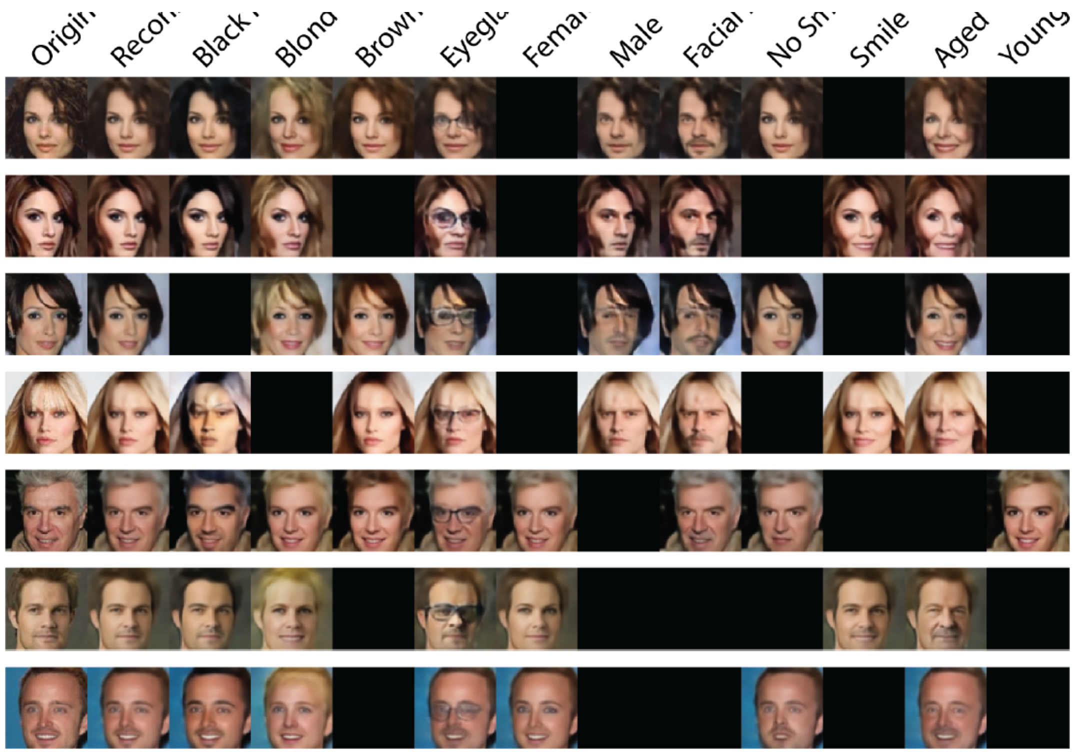
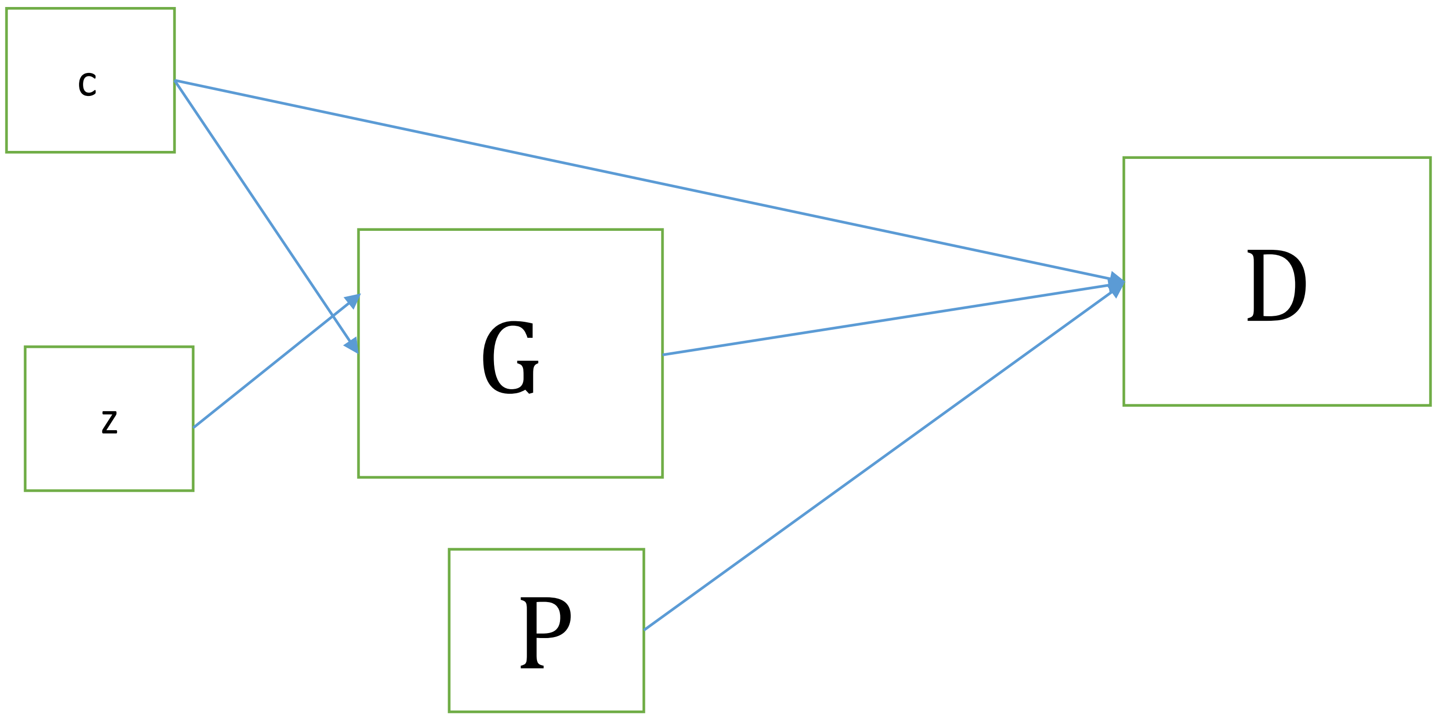


Figure 5: Identity-preserving transformations with optimization. Two separate critics are trained, one for attributes and one for the realism constraint. Starting at the latent points corresponding to the data reconstructions, we then perform gradient ascent in latent space on a weighted combination of critic values (1.0 attribute, 0.1 marginal posterior), stopping when a threshold value is passed for both critics. Images remain semantically close to the original because the pixel-wise likelihood of VAE training encourages identity-preserving reconstructions, and the dynamics of gradient ascent are naturally limited to finding solutions close in latent space. Panels are black for attributes of the original image, as the procedure just returns the original point in latent space.

Conditional Image Synthesis with Auxiliary Classifier GANs

- $X_{fake} = G(c; z)$



Conditional Image Synthesis with Auxiliary Classifier GANs

$$L_S = E[\log P(S = real \mid X_{real})] + E[\log P(S = fake \mid X_{fake})] \quad (2)$$

$$L_C = E[\log P(C = c \mid X_{real})] + E[\log P(C = c \mid X_{fake})] \quad (3)$$

- D maximize $L_S + L_c$
- G maximize $L_C - L_S$

Result: Diversity

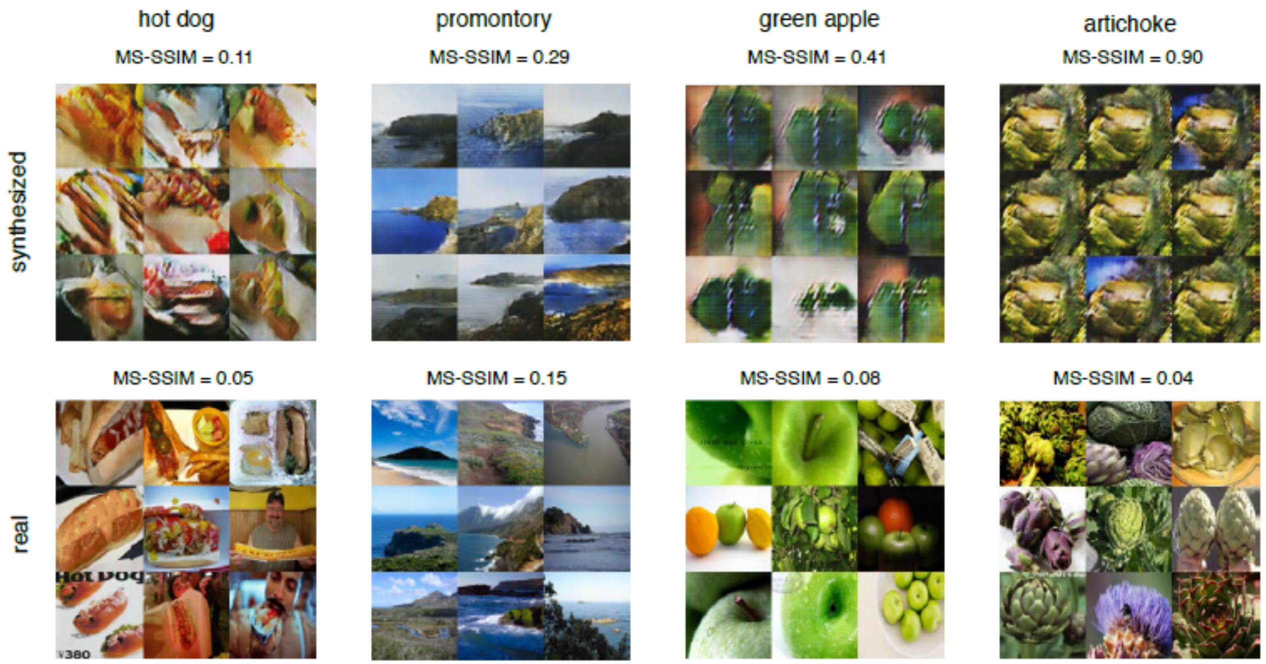


Figure 3. Examples of different MS-SSIM scores. The top and bottom rows contain AC-GAN samples and training data, respectively.