Deep Learning for Computer Vision

Combining VAEs and GANs

Vineeth N Balasubramanian

Department of Computer Science and Engineering
Indian Institute of Technology, Hyderabad

Review: Questions

Questions

- Why does the encoder of a VAE map to a vector of means and a vector of standard deviations? Why does it not instead map to a vector of means and a covariance matrix?

Review: Questions

Questions

- Why does the encoder of a VAE map to a vector of means and a vector of standard deviations? Why does it not instead map to a vector of means and a covariance matrix? **We are explicitly learning a set of independent Gaussians, which makes the learning easier - and of course, it works!**

Review: Questions

Questions

- What about the decoder? If we assume a Mean Squared Error for the reconstruction loss, what is the covariance of the $p(x|z)$ Gaussian?

Review: Questions

Questions

- What about the decoder? If we assume a Mean Squared Error for the reconstruction loss, what is the covariance of the $p(x|z)$ Gaussian?
  Equivalent to modeling $p(x|z)$ as Gaussian with identity covariance; in this case, decoder output is mean $\mu(t)$ and, therefore, for an example x_i, you get the following reconstruction loss:

$$
\begin{aligned}
-\log(p(x_i|t_i)) &= -\log\left(\frac{1}{\sqrt{(2\pi)^k|I|}}\exp\left(-\frac{1}{2}(x_i-\mu(t_i))^\intercal I(x_i-\mu(t_i))\right)\right) \\
&= \frac{1}{2}||x_i-\mu(t_i)||^2 + \text{const.}
\end{aligned}
$$

This is MSE!

VAE vs GAN

VAE

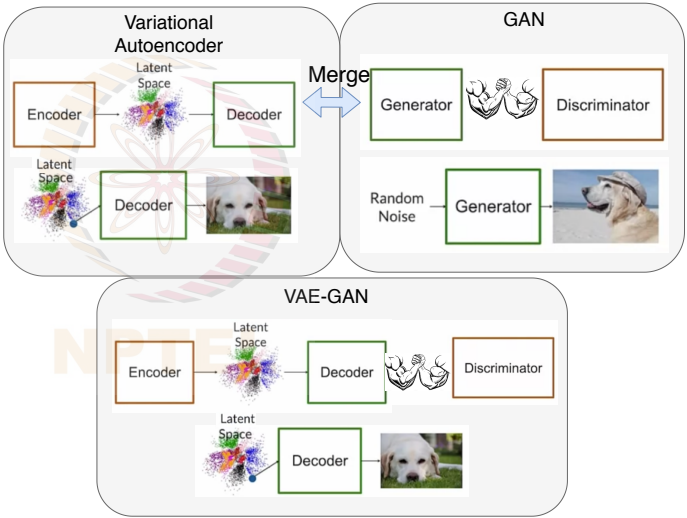
- **+ve** Learns an inference machine by mapping data to a latent space with distribution of choice enabling **fast/efficient inference**
- **-ve** Tends to distribute probability mass diffusely over data space resulting in blurred/**low quality image** samples

GAN

- **+ve:** Bypass inference and learn generative model that produces high quality samples without sacrificing sampling speed
- **-ve:** **Lacks** an effective **inference mechanism** preventing from reasoning about data at an abstract level

Can we combine VAEs and GANs?

- **Solution:** Bridge the gap between VAEs and GANs, learn models that generate high-quality samples along with an effective inference network



VAE Limitations

Let's recall the VAE objective:

$$\mathcal{L}_{VAE} = -\mathbb{E}_{q(z|x)}[\log p((x)|z)] + D_{KL}(q(z|x)||p(z))$$

VAE Limitations

Let's recall the VAE objective:

$$\begin{aligned}\mathcal{L}_{VAE} &= -\mathbb{E}_{q(z|x)}[\log p((x)|z)] + D_{KL}(q(z|x)||p(z)) \\ &= \mathcal{L}_{reconstruction} + \mathcal{L}_{prior}\end{aligned}$$

VAE Limitations

Let's recall the VAE objective:

$$\begin{aligned}\mathcal{L}_{VAE} &= -\mathbb{E}_{q(z|x)}[\log p((x)|z)] + D_{KL}(q(z|x)||p(z)) \\ &= \mathcal{L}_{reconstruction} + \mathcal{L}_{prior} \\ &= MSE(x, \tilde{x}) + D_{KL}(q(z|x)||p(z))\end{aligned}$$

VAE Limitations

Let's recall the VAE objective:

$$\begin{aligned}\mathcal{L}_{VAE} &= -\mathbb{E}_{q(z|x)}[\log p((x)|z)] + D_{KL}(q(z|x)||p(z)) \\ &= \mathcal{L}_{reconstruction} + \mathcal{L}_{prior} \\ &= MSE(x, \tilde{x}) + D_{KL}(q(z|x)||p(z))\end{aligned}$$

Mean-Squared Error

VAE Limitations

Let's recall the VAE objective:

$$\begin{aligned}\mathcal{L}_{VAE} &= -\mathbb{E}_{q(z|x)}[\log p((x)|z)] + D_{KL}(q(z|x)||p(z)) \\ &= \mathcal{L}_{reconstruction} + \mathcal{L}_{prior} \\ &= MSE(x, \tilde{x}) + D_{KL}(q(z|x)||p(z))\end{aligned}$$

Mean-Squared Error

- Assumes signal fidelity is independent of temporal/spatial relationships $\rightarrow$ does not hold for images

VAE Limitations

Let's recall the VAE objective:

$$\begin{aligned}\mathcal{L}_{VAE} &= -\mathbb{E}_{q(z|x)}[\log p((x)|z)] + D_{KL}(q(z|x)||p(z)) \\ &= \mathcal{L}_{reconstruction} + \mathcal{L}_{prior} \\ &= MSE(x, \tilde{x}) + D_{KL}(q(z|x)||p(z))\end{aligned}$$

Mean-Squared Error

- Assumes signal fidelity is independent of temporal/spatial relationships $\rightarrow$ does not hold for images
- Element-wise metric unable to model human perception of image fidelity and quality $\rightarrow$ low image quality

VAE Limitations

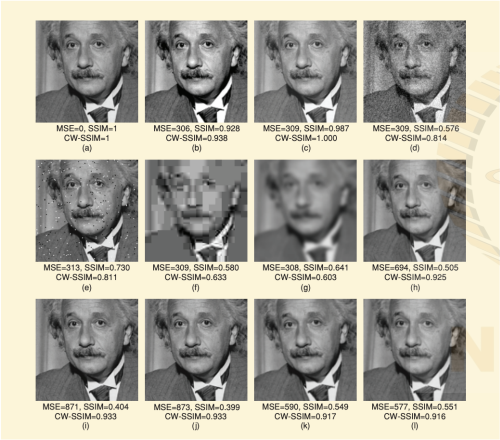
Let's recall the VAE objective:

$$\begin{aligned}\mathcal{L}_{VAE} &= -\mathbb{E}_{q(z|x)}[\log p((x)|z)] + D_{KL}(q(z|x)||p(z)) \\ &= \mathcal{L}_{reconstruction} + \mathcal{L}_{prior} \\ &= MSE(x, \tilde{x}) + D_{KL}(q(z|x)||p(z))\end{aligned}$$

Mean-Squared Error

- Assumes signal fidelity is independent of temporal/spatial relationships → does not hold for images
- Element-wise metric unable to model human perception of image fidelity and quality → low image quality
- Pixel-based loss metric does not respect semantic-preserving transforms, e.g. scaling/translation

MSE and Image Fidelity[1]



[FIG2] **Comparison of image fidelity measures for "Einstein" image altered with different types of distortions. (a) Reference image. (b) Mean contrast stretch. (c) Luminance shift. (d) Gaussian noise contamination. (e) Impulsive noise contamination. (f) JPEG compression. (g) Blurring. (h) Spatial scaling (zooming out). (i) Spatial shift (to the right). (j) Spatial shift (to the left). (k) Rotation (counter-clockwise). (l) Rotation (clockwise).**

- MSE values of all distorted images - images (b)–(g) relative to (a) - are nearly identical, even though images are perceptually different
- Images with small geometrical modifications - images (h)–(i) - present large MSE values relative to (a), yet show negligible change in perceived quality

[1] Wang and Bovik, Mean Squared Error: Love It or Leave It? IEEE Signal Processing Magazine, 2009

VAE Limitations

KL Divergence

- Focused on encouraging $q(z)$ to pick modes of $p(z) \rightarrow$ unable to match $q(z)$ to whole distribution of $p(z)$ well

VAE Limitations

KL Divergence

- Focused on encouraging $q(z)$ to pick modes of $p(z)$ $\rightarrow$ unable to match $q(z)$ to whole distribution of $p(z)$ well
- 'Spaces' or 'holes' in learned latent space $\rightarrow$ may fail to capture data manifold

VAE Limitations

KL Divergence

- Focused on encouraging $q(z)$ to pick modes of $p(z)$ $\rightarrow$ unable to match $q(z)$ to whole distribution of $p(z)$ well
- 'Spaces' or 'holes' in learned latent space $\rightarrow$ may fail to capture data manifold
- May miss several local regions in data space $\rightarrow$ adversely effect generalization capability

VAE Limitations

KL Divergence

- Focused on encouraging $q(z)$ to pick modes of $p(z)$ $\rightarrow$ unable to match $q(z)$ to whole distribution of $p(z)$ well
- 'Spaces' or 'holes' in learned latent space $\rightarrow$ may fail to capture data manifold
- May miss several local regions in data space $\rightarrow$ adversely effect generalization capability

Form of Prior

- Requires access to exact functional form of prior

VAE Limitations

KL Divergence

- Focused on encouraging $q(z)$ to pick modes of $p(z)$ $\rightarrow$ unable to match $q(z)$ to whole distribution of $p(z)$ well
- 'Spaces' or 'holes' in learned latent space $\rightarrow$ may fail to capture data manifold
- May miss several local regions in data space $\rightarrow$ adversely effect generalization capability

Form of Prior

- Requires access to exact functional form of prior
- Difficult to optimize; not computable in closed form always for all priors $\rightarrow$ limits choice of priors that can be used

How to address?

VAE Limitations

KL Divergence

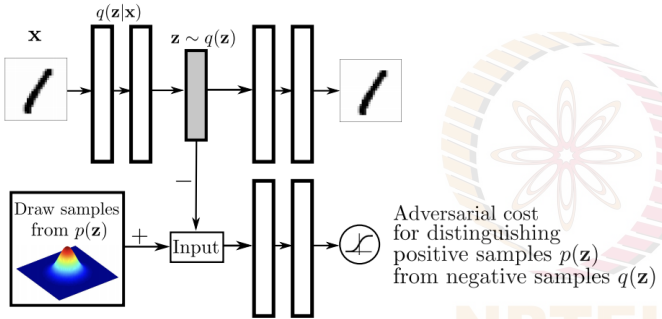
- Focused on encouraging $q(z)$ to pick modes of $p(z)$ $\rightarrow$ unable to match $q(z)$ to whole distribution of $p(z)$ well
- 'Spaces' or 'holes' in learned latent space $\rightarrow$ may fail to capture data manifold
- May miss several local regions in data space $\rightarrow$ adversely effect generalization capability

Form of Prior

- Requires access to exact functional form of prior
- Difficult to optimize; not computable in closed form always for all priors $\rightarrow$ limits choice of priors that can be used

How to address? Integrate with GANs to use its positives to overcome limitations!

Adversarial Autoencoder (AAE)[2]



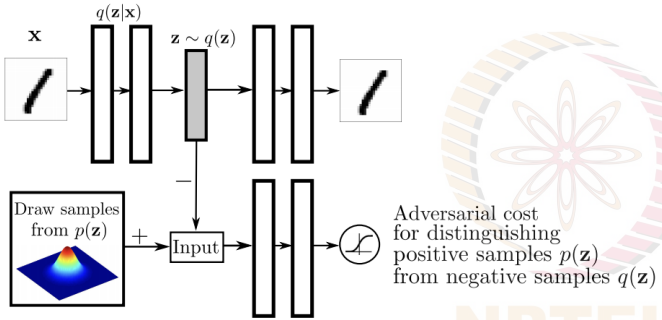
(Top) Standard VAE

(Bottom) Second network trained to discriminatively predict whether a sample arises from hidden code of autoencoder or input training distribution

- Aims to match aggregated posterior, $q(z)$, to an arbitrary prior, $p(z)$ via adversarial objective-based training

[2] Makhzani et al, Adversarial Autoencoders, ICLRW 2016

Adversarial Autoencoder (AAE)[2]



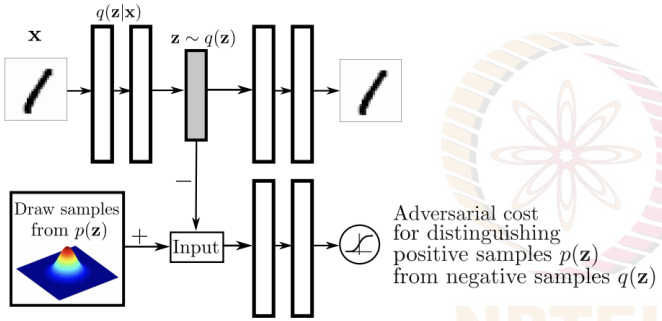
(Top) Standard VAE

(Bottom) Second network trained to discriminatively predict whether a sample arises from hidden code of autoencoder or input training distribution

- Aims to match aggregated posterior, $q(z)$, to an arbitrary prior, $p(z)$ via adversarial objective-based training
- Renders continuous learned latent space $\rightarrow$ captures data manifold well

[2] Makhzani et al, Adversarial Autoencoders, ICLRW 2016

Adversarial Autoencoder (AAE)[2]



(Top) Standard VAE

(Bottom) Second network trained to discriminatively predict whether a sample arises from hidden code of autoencoder or input training distribution

- Aims to match aggregated posterior, $q(z)$, to an arbitrary prior, $p(z)$ via adversarial objective-based training
- Renders continuous learned latent space $\rightarrow$ captures data manifold well
- Encoder converts data distribution to prior distribution, while decoder learns a deep generative model that maps imposed prior to data distribution

[2] Makhzani et al, Adversarial Autoencoders, ICLRW 2016

Training AAE

Objective:

$$\mathcal{L} = \mathbb{E}_x \underbrace{\left[\mathbb{E}_{q(z|x)}[-\log p(x|z)]\right]}_{\text{Reconstruction Error}} + \mathbb{E}_x \underbrace{\left[\mathrm{KL}(q(z|x)||p(z))\right]}_{\text{KL Regularizer}}$$

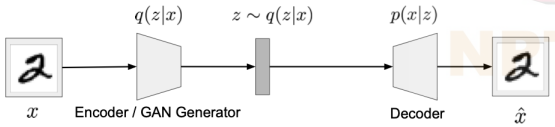
Replaced by adversarial loss in AAE

Training AAE

Objective:

$$\mathcal{L} = \mathbb{E}_x \left[\underbrace{\mathbb{E}_{q(z|x)}[-\log p(x|z)]}_{\text{Reconstruction Error}} \right] + \mathbb{E}_x \left[\underbrace{\mathrm{KL}(q(z|x)||p(z))}_{\text{KL Regularizer}} \right]$$

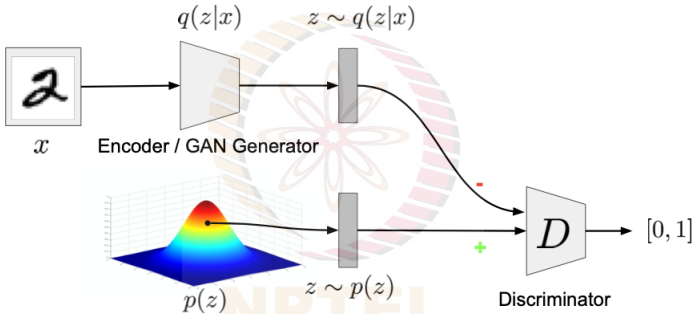
Replaced by adversarial loss in AAE



Reconstruction Phase:

- Introduce latent variable z with simple prior $p(z)$ (e.g. Gaussian)
- Sample $z \sim p(z)$, pass it through Generator $\hat{x} = G(z)$; where $\hat{x} \sim p_G$
- Introduce mechanism to ensure $p_G \approx p_{data}$

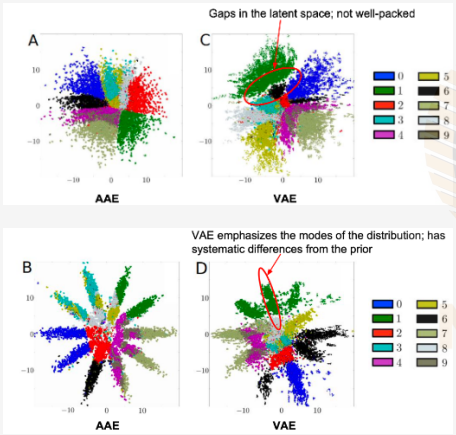
Training AAE



Regularization Phase

- Aims to match aggregated posterior, $q(z)$, to an arbitrary prior, $p(z)$ via adversarial objective-based training

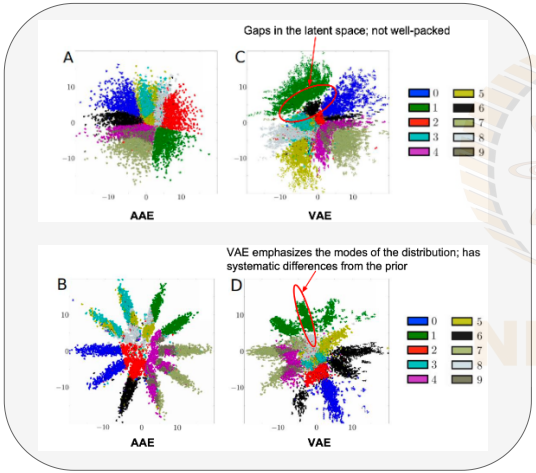
Learned Latent Space: AAE vs VAE[3]



- Latent space of test data using models trained on MNIST:
  - **Top:** Spherical 2-D Gaussian prior distribution
  - **Bottom:** Mixture of 10 2-D Gaussian

[3] Makhzani et al, Adversarial Autoencoders, ICLRW 2016

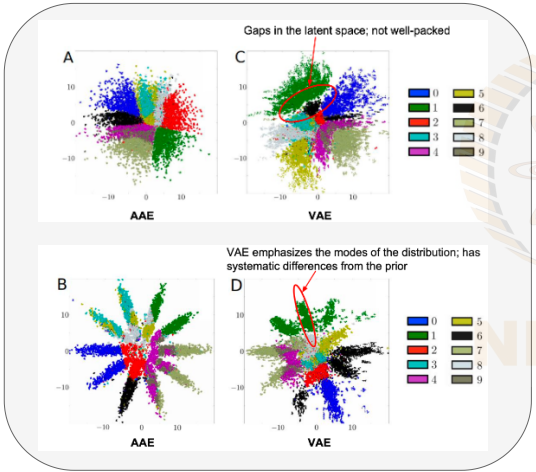
Learned Latent Space: AAE vs VAE[3]



- Latent space of test data using models trained on MNIST:
  - **Top:** Spherical 2-D Gaussian prior distribution
  - **Bottom:** Mixture of 10 2-D Gaussian
- Adversarial training to impose prior (AAE) renders a more continuous latent space relative to KL divergence-based (VAE) distribution alignment **(Top)**

[3] Makhzani et al, Adversarial Autoencoders, ICLRW 2016

Learned Latent Space: AAE vs VAE[3]



- Latent space of test data using models trained on MNIST:
  - **Top:** Spherical 2-D Gaussian prior distribution
  - **Bottom:** Mixture of 10 2-D Gaussian
- Adversarial training to impose prior (AAE) renders a more continuous latent space relative to KL divergence-based (VAE) distribution alignment **(Top)**
- AAE imposes multi-modal distributions better than VAE **(Bottom)**

[3] Makhzani et al, Adversarial Autoencoders, ICLRW 2016

Imposing Complex Priors

- **Left:** Latent space of AAE trained on MNIST dataset with Swiss roll distribution as prior $p(z)$
- **Right:** Samples generated by walking along the main Swiss roll axis

Imposing Complex Priors

- **Left:** Latent space of AAE trained on MNIST dataset with Swiss roll distribution as prior $p(z)$
- **Right:** Samples generated by walking along the main Swiss roll axis
- For AAE, only require sampling from prior distribution in order to induce $q(z)$ to match $p(z)$; exact functional form of prior is not required

VAE-GAN[4]

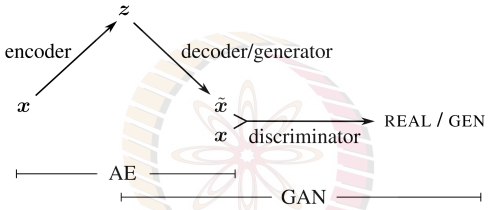


Figure 1. Overview of our network. We combine a VAE with a GAN by collapsing the decoder and the generator into one.

- Replace element-wise MSE in pixel space with feature-wise metric between discriminator's hidden representations

[4] Larsen et al, Autoencoding beyond Pixels using a Learned Similarity Metric, ICML 2016

VAE-GAN[4]

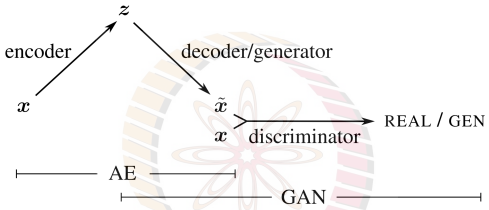


Figure 1. Overview of our network. We combine a VAE with a GAN by collapsing the decoder and the generator into one.

- Replace element-wise MSE in pixel space with feature-wise metric between discriminator's hidden representations
- Combines advantage of GAN as high-quality generative model and VAE as a method that produces an encoding of data into latent space z

[4] Larsen et al, Autoencoding beyond Pixels using a Learned Similarity Metric, ICML 2016

Loss Formulation

Let $Dis_l(x)$ denote hidden representation of l^{th} layer of discriminator; then:

$$p(Dis_l(x)|z) = \mathcal{N}(Dis_l(x)|Dis_l(\tilde{x})|, I)$$

where $\tilde{x} \sim Dec(z)$ is a sample from decoder

Loss Formulation

Let $Dis_l(x)$ denote hidden representation of l^{th} layer of discriminator; then:

$$p(Dis_l(x)|z) = \mathcal{N}(Dis_l(x)|Dis_l(\tilde{x})|, I)$$

where $\tilde{x} \sim Dec(z)$ is a sample from decoder

$$\mathcal{L}^{Dis_l}_{recon-content} = -\mathbb{E}_{q(z|x)}[\log p(Dis_l(x)|z)]$$

Loss Formulation

Let $Dis_l(x)$ denote hidden representation of l^{th} layer of discriminator; then:

$$p(Dis_l(x)|z) = \mathcal{N}(Dis_l(x)|Dis_l(\tilde{x})|, I)$$

where $\tilde{x} \sim Dec(z)$ is a sample from decoder

$$\mathcal{L}^{Dis_l}_{recon-content} = -\mathbb{E}_{q(z|x)}[\log p(Dis_l(x)|z)]$$

$$\mathcal{L}^{GAN}_{recon-style} = \log(Dis(x)) + \log(1 - Dis(Gen(z)))$$

Loss Formulation

Let $Dis_l(x)$ denote hidden representation of l^{th} layer of discriminator; then:

$$p(Dis_l(x)|z) = \mathcal{N}(Dis_l(x)|Dis_l(\tilde{x})|, I)$$

where $\tilde{x} \sim Dec(z)$ is a sample from decoder

$$\mathcal{L}^{Dis_l}_{recon-content} = -\mathbb{E}_{q(z|x)}[\log p(Dis_l(x)|z)]$$

$$\mathcal{L}^{GAN}_{recon-style} = \log(Dis(x)) + \log(1 - Dis(Gen(z)))$$

$$\mathcal{L}_{prior} = D_{KL}(q(z|x)||p(z))$$

Loss Formulation

Let $Dis_l(x)$ denote hidden representation of l^{th} layer of discriminator; then:

$$p(Dis_l(x)|z) = \mathcal{N}(Dis_l(x)|Dis_l(\tilde{x})|, I)$$

where $\tilde{x} \sim Dec(z)$ is a sample from decoder

$$\mathcal{L}^{Dis_l}_{recon-content} = -\mathbb{E}_{q(z|x)}[\log p(Dis_l(x)|z)]$$

$$\mathcal{L}^{GAN}_{recon-style} = \log(Dis(x)) + \log(1 - Dis(Gen(z)))$$

$$\mathcal{L}_{prior} = D_{KL}(q(z|x)||p(z))$$

$$\mathcal{L}_{\textbf{total}} = \mathcal{L}^{Dis_l}_{recon-content} + \mathcal{L}^{GAN}_{recon-style} + \mathcal{L}_{prior}$$

Training Algorithm

Algorithm 1 Training the VAE/GAN model

$\boldsymbol{\theta}_{\text{Enc}}, \boldsymbol{\theta}_{\text{Dec}}, \boldsymbol{\theta}_{\text{Dis}} \leftarrow$ initialize network parameters

repeat

$\boldsymbol{X} \leftarrow$ random mini-batch from dataset

$\boldsymbol{Z} \leftarrow \text{Enc}(\boldsymbol{X})$

$\mathcal{L}_{\text{prior}} \leftarrow D_{\text{KL}}(q(\boldsymbol{Z}|\boldsymbol{X})||p(\boldsymbol{Z}))$

$\tilde{\boldsymbol{X}} \leftarrow \text{Dec}(\boldsymbol{Z})$

$\mathcal{L}_{\text{llike}}^{\text{Dis}_l} \leftarrow -\mathbb{E}_{q(\boldsymbol{Z}|\boldsymbol{X})}\left[p(\text{Dis}_l(\boldsymbol{X})|\boldsymbol{Z})\right]$

$\boldsymbol{Z}_p \leftarrow$ samples from prior $\mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$

$\boldsymbol{X}_p \leftarrow \text{Dec}(\boldsymbol{Z}_p)$

$\mathcal{L}_{\text{GAN}} \leftarrow \log(\text{Dis}(\boldsymbol{X})) + \log(1 - \text{Dis}(\tilde{\boldsymbol{X}})) + \log(1 - \text{Dis}(\boldsymbol{X}_p))$

// Update parameters according to gradients

$\boldsymbol{\theta}_{\text{Enc}} \overset{+}{\leftarrow} -\nabla_{\boldsymbol{\theta}_{\text{Enc}}}(\mathcal{L}_{\text{prior}} + \mathcal{L}_{\text{llike}}^{\text{Dis}_l})$

$\boldsymbol{\theta}_{\text{Dec}} \overset{+}{\leftarrow} -\nabla_{\boldsymbol{\theta}_{\text{Dec}}}(\gamma\mathcal{L}_{\text{llike}}^{\text{Dis}_l} - \mathcal{L}_{\text{GAN}})$

$\boldsymbol{\theta}_{\text{Dis}} \overset{+}{\leftarrow} -\nabla_{\boldsymbol{\theta}_{\text{Dis}}}\mathcal{L}_{\text{GAN}}$

until deadline

- Dis should not try to minimize $L_{recon-content}^{Dis_l}$ as this would collapse the discriminator to 0

Training Algorithm

Algorithm 1 Training the VAE/GAN model

$\boldsymbol{\theta}_{\text{Enc}}, \boldsymbol{\theta}_{\text{Dec}}, \boldsymbol{\theta}_{\text{Dis}} \leftarrow$ initialize network parameters

repeat

$\boldsymbol{X} \leftarrow$ random mini-batch from dataset

$\boldsymbol{Z} \leftarrow \text{Enc}(\boldsymbol{X})$

$\mathcal{L}_{\text{prior}} \leftarrow D_{\text{KL}}(q(\boldsymbol{Z}|\boldsymbol{X})\|p(\boldsymbol{Z}))$

$\tilde{\boldsymbol{X}} \leftarrow \text{Dec}(\boldsymbol{Z})$

$\mathcal{L}_{\text{llike}}^{\text{Dis}_l} \leftarrow -\mathbb{E}_{q(\boldsymbol{Z}|\boldsymbol{X})}\left[p(\text{Dis}_l(\boldsymbol{X})|\boldsymbol{Z})\right]$

$\boldsymbol{Z}_p \leftarrow$ samples from prior $\mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$

$\boldsymbol{X}_p \leftarrow \text{Dec}(\boldsymbol{Z}_p)$

$\mathcal{L}_{\text{GAN}} \leftarrow \log(\text{Dis}(\boldsymbol{X})) + \log(1 - \text{Dis}(\tilde{\boldsymbol{X}})) + \log(1 - \text{Dis}(\boldsymbol{X}_p))$

// Update parameters according to gradients

$\boldsymbol{\theta}_{\text{Enc}} \overset{+}{\leftarrow} -\nabla_{\boldsymbol{\theta}_{\text{Enc}}}(\mathcal{L}_{\text{prior}} + \mathcal{L}_{\text{llike}}^{\text{Dis}_l})$

$\boldsymbol{\theta}_{\text{Dec}} \overset{+}{\leftarrow} -\nabla_{\boldsymbol{\theta}_{\text{Dec}}}(\gamma\mathcal{L}_{\text{llike}}^{\text{Dis}_l} - \mathcal{L}_{\text{GAN}})$

$\boldsymbol{\theta}_{\text{Dis}} \overset{+}{\leftarrow} -\nabla_{\boldsymbol{\theta}_{\text{Dis}}}\mathcal{L}_{\text{GAN}}$

until deadline

- Dis should not try to minimize $L_{recon-content}^{Dis_l}$ as this would collapse the discriminator to 0
- Use samples $\mathbf{X_p}$ from prior directly in GAN loss in addition to $\tilde{\mathbf{X}}$

Training Algorithm

Algorithm 1 Training the VAE/GAN model

$\boldsymbol{\theta}_{\text{Enc}}, \boldsymbol{\theta}_{\text{Dec}}, \boldsymbol{\theta}_{\text{Dis}} \leftarrow$ initialize network parameters

repeat

$\boldsymbol{X} \leftarrow$ random mini-batch from dataset

$\boldsymbol{Z} \leftarrow \text{Enc}(\boldsymbol{X})$

$\mathcal{L}_{\text{prior}} \leftarrow D_{\text{KL}}(q(\boldsymbol{Z}|\boldsymbol{X})||p(\boldsymbol{Z}))$

$\tilde{\boldsymbol{X}} \leftarrow \text{Dec}(\boldsymbol{Z})$

$\mathcal{L}_{\text{llike}}^{\text{Dis}_l} \leftarrow -\mathbb{E}_{q(\boldsymbol{Z}|\boldsymbol{X})}\left[p(\text{Dis}_l(\boldsymbol{X})|\boldsymbol{Z})\right]$

$\boldsymbol{Z}_p \leftarrow$ samples from prior $\mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$

$\boldsymbol{X}_p \leftarrow \text{Dec}(\boldsymbol{Z}_p)$

$\mathcal{L}_{\text{GAN}} \leftarrow \log(\text{Dis}(\boldsymbol{X})) + \log(1 - \text{Dis}(\tilde{\boldsymbol{X}})) + \log(1 - \text{Dis}(\boldsymbol{X}_p))$

// Update parameters according to gradients

$\boldsymbol{\theta}_{\text{Enc}} \overset{+}{\leftarrow} -\nabla_{\boldsymbol{\theta}_{\text{Enc}}}(\mathcal{L}_{\text{prior}} + \mathcal{L}_{\text{llike}}^{\text{Dis}_l})$

$\boldsymbol{\theta}_{\text{Dec}} \overset{+}{\leftarrow} -\nabla_{\boldsymbol{\theta}_{\text{Dec}}}(\gamma\mathcal{L}_{\text{llike}}^{\text{Dis}_l} - \mathcal{L}_{\text{GAN}})$

$\boldsymbol{\theta}_{\text{Dis}} \overset{+}{\leftarrow} -\nabla_{\boldsymbol{\theta}_{\text{Dis}}}\mathcal{L}_{\text{GAN}}$

until deadline

- Dis should not try to minimize $L_{recon-content}^{Dis_l}$ as this would collapse the discriminator to 0
- Use samples $\mathbf{X_p}$ from prior directly in GAN loss in addition to $\tilde{\mathbf{X}}$
- Weight $L_{recon-content}^{Dis_l}$ and $L_{recon-style}^{GAN}$ to weight ability to reconstruct vs fooling discriminator

Comparing Generated Samples

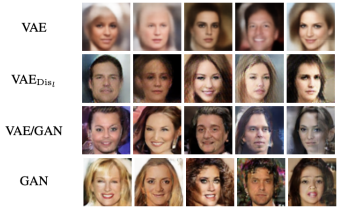


Figure 3. Samples from different generative models.

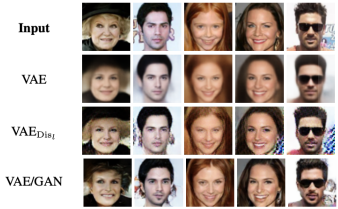
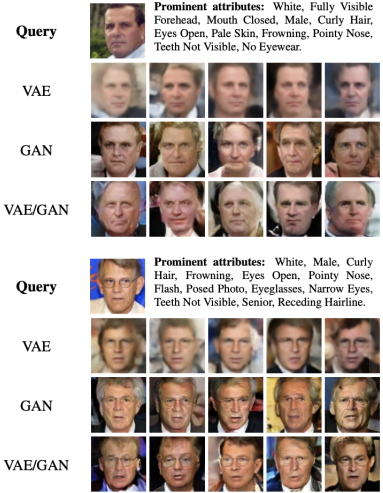


Figure 4. Reconstructions from different autoencoders.

- Draw samples from $p(z)$ and propagate these through decoder to generate new images
- VAE draws frontal part well, but off-center gets blurry
- VAE^l_{Dis} produces sharper images even off-center

Conditional Generation



- Concatenate face attribute vector (LFW image dataset) to vector representation of input in Enc, Dec and Dis while training
- Use trained model to generate faces conditioned on held-out test attributes
- Compared to an ordinary VAE, the VAE/GAN model yields significantly better images visually

Homework

Readings

- A wizard's guide to Adversarial Autoencoders: Part 1, Autoencoder
- A wizard's guide to Adversarial Autoencoders: Part 2, Exploring latent space with Adversarial Autoencoders
- What The Heck Are VAE-GANs?
- (YouTube video) VAE-GAN Explained