ICML 2018

JointGAN: Multi-Domain Joint Distribution Learning with Generative Adversarial Nets

Presenter: Shuyang Dai

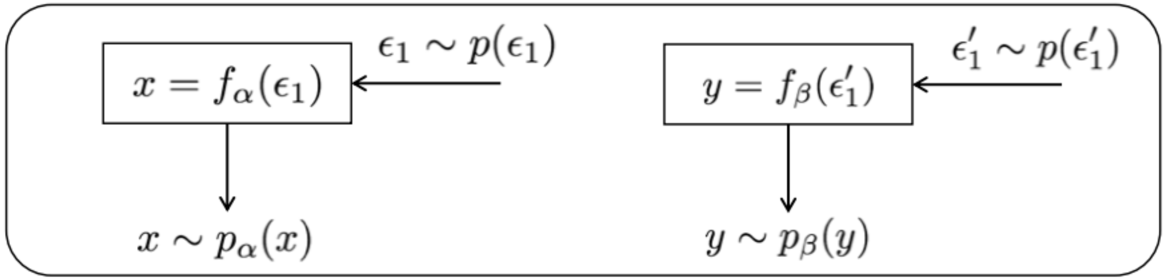
Yunchen Pu, Shuyang Dai, Zhe Gan, Weiyao Wang, Guoyin Wang, Yizhe Zhang, Ricardo Henao, Lawrence Carin

Facebook & Duke University & Microsoft Research

Joint Distribution Learning

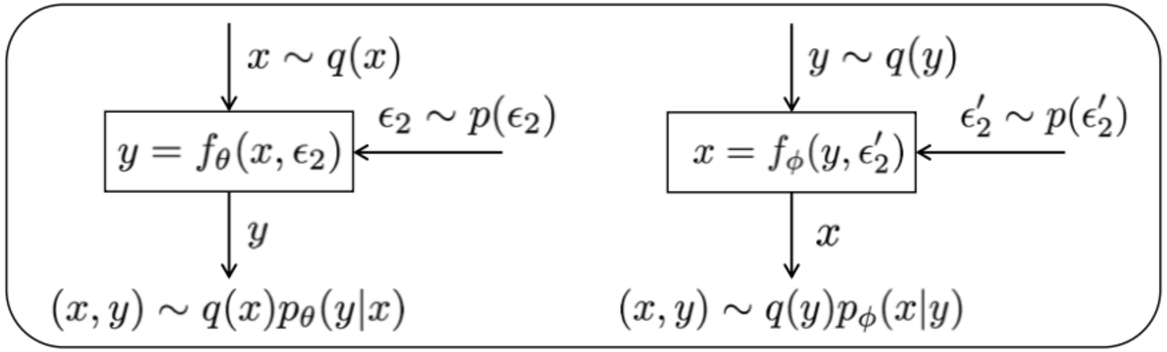
- Learning *marginals* with GAN:

$$\tilde{\boldsymbol{x}} = f_{\boldsymbol{\alpha}}(\boldsymbol{\epsilon}_1), \qquad \boldsymbol{\epsilon}_1 \sim p(\boldsymbol{\epsilon}_1)$$
$$\tilde{\boldsymbol{y}} = f_{\boldsymbol{\beta}}(\boldsymbol{\epsilon}_1'), \qquad \boldsymbol{\epsilon}_1' \sim p(\boldsymbol{\epsilon}_1')$$



- Learning *conditionals* with ALI, Cycle GAN, and Triangle GAN:

$$\tilde{\boldsymbol{y}} = f_{\boldsymbol{\theta}}(\boldsymbol{x}, \boldsymbol{\epsilon}_2), \quad \boldsymbol{x} \sim q(\boldsymbol{x}), \quad \boldsymbol{\epsilon}_2 \sim p(\boldsymbol{\epsilon}_2)$$
$$\tilde{\boldsymbol{x}} = f_{\boldsymbol{\phi}}(\boldsymbol{y}, \boldsymbol{\epsilon}_2'), \quad \boldsymbol{y} \sim q(\boldsymbol{y}), \quad \boldsymbol{\epsilon}_2' \sim p(\boldsymbol{\epsilon}_2')$$



Though joint distributions are matched,
only conditional distributions are learned.

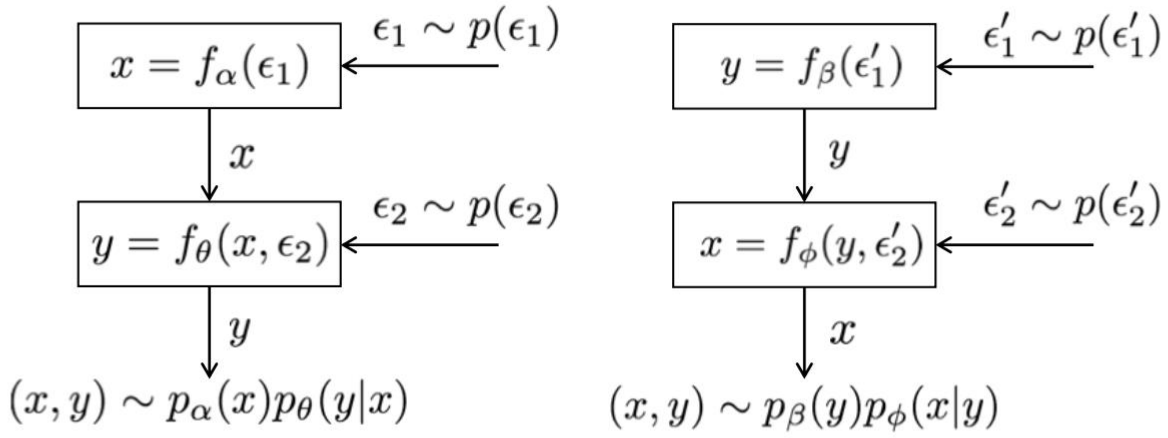
Is there a way to
fully learn the joint distribution ?

Joint Distribution Learning

- Potential advantages:
  - Synthesis of draws from any of the marginals
  - Synthesis of draws from the conditionals when other random variables are observed, *i.e.*, imputation
  - Also allows complete draws from the full joint distribution

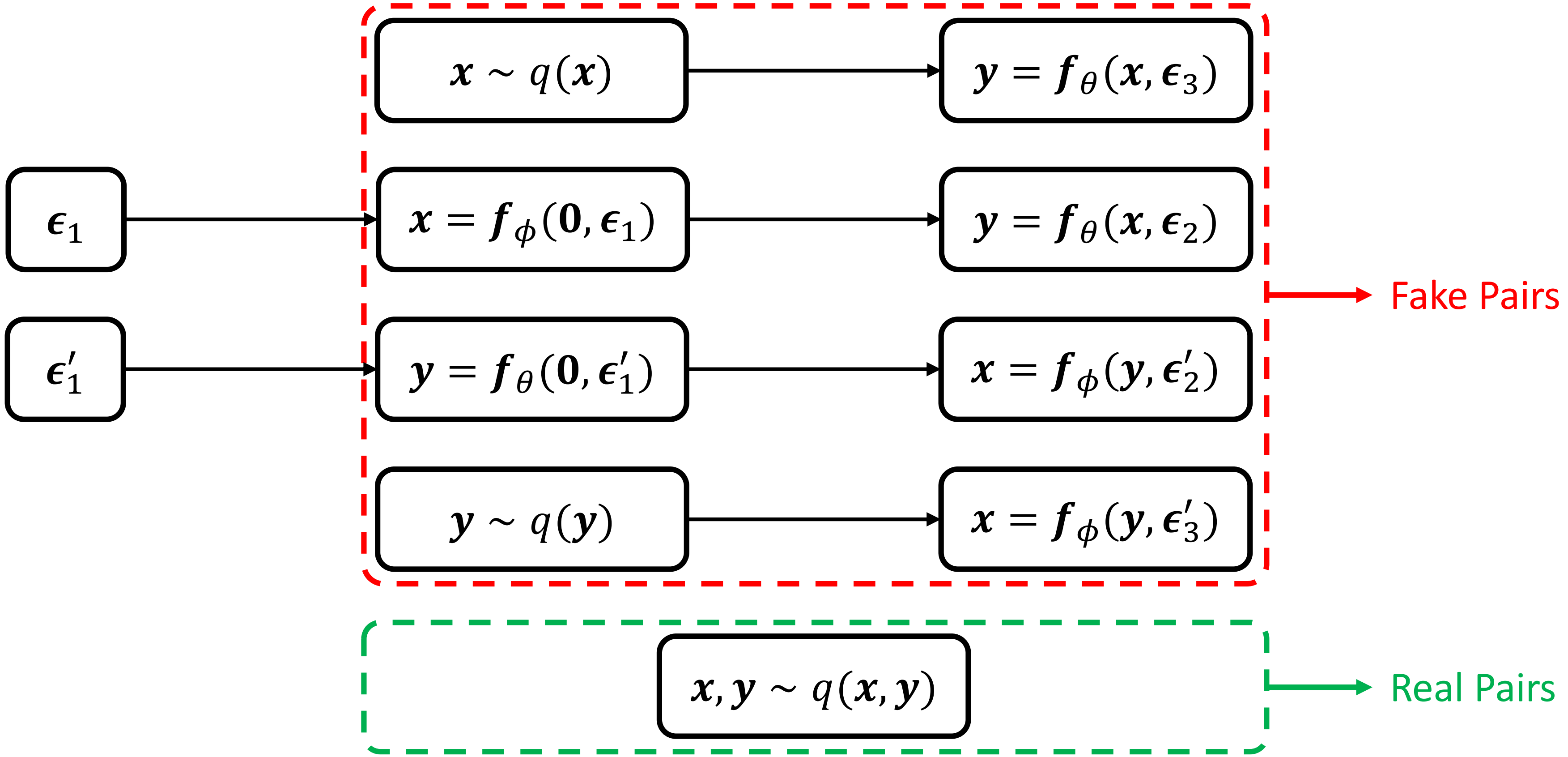
Joint Distribution Learning

- Learning joint distribution with the proposed JointGAN:



- To reduce the number of parameters, let $f_\alpha(\cdot) = f_\phi(\mathbf{0},\ \cdot\ )$ and $f_\beta(\cdot) = f_\theta(\mathbf{0},\ \cdot\ )$, where $\mathbf{0}$ is an all-zero tensor.

Training JointGAN



Training JointGAN

- A single 5-way critic (discriminator):

$$\begin{aligned} p_1(\boldsymbol{x},\boldsymbol{y}) &= q(\boldsymbol{x})p_{\boldsymbol{\theta}}(\boldsymbol{y}|\boldsymbol{x}), \quad & p_2(\boldsymbol{x},\boldsymbol{y}) &= q(\boldsymbol{y})p_{\boldsymbol{\phi}}(\boldsymbol{x}|\boldsymbol{y}) \\ p_3(\boldsymbol{x},\boldsymbol{y}) &= p_{\boldsymbol{\alpha}}(\boldsymbol{x})p_{\boldsymbol{\theta}}(\boldsymbol{y}|\boldsymbol{x}), \quad & p_4(\boldsymbol{x},\boldsymbol{y}) &= p_{\boldsymbol{\beta}}(\boldsymbol{y})p_{\boldsymbol{\phi}}(\boldsymbol{x}|\boldsymbol{y}) \\ p_5(\boldsymbol{x},\boldsymbol{y}) &= q(\boldsymbol{x},\boldsymbol{y}) \end{aligned} \tag{3}$$

- The minimax objective for JointGAN:

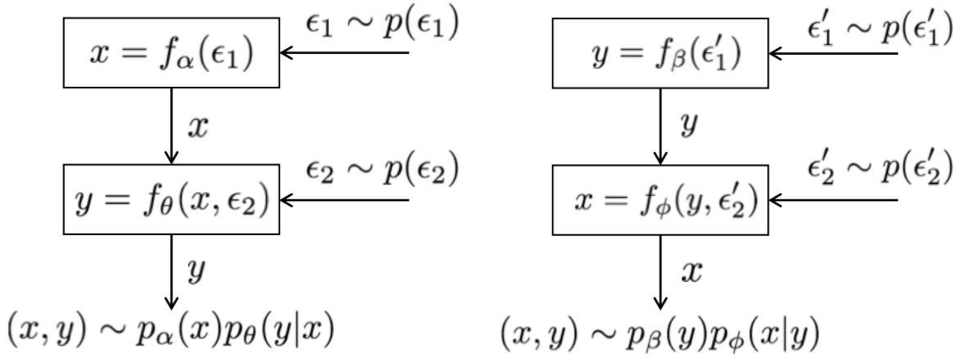
$$\begin{aligned} &\min_{\boldsymbol{\theta},\boldsymbol{\phi}} \max_{\boldsymbol{\omega}} \ \mathcal{L}_{\text{JointGAN}}(\boldsymbol{\theta},\boldsymbol{\phi},\boldsymbol{\omega}) \\ &\qquad = \textstyle\sum_{k=1}^{5} \mathbb{E}_{p_k(\boldsymbol{x},\boldsymbol{y})}[\log g_{\boldsymbol{\omega}}(\boldsymbol{x},\boldsymbol{y})[k]] \end{aligned} \tag{4}$$

where the critic $g_\omega(\boldsymbol{x},\boldsymbol{y}) \in \Delta^4$ has softmax on the top layer:

$$\textstyle\sum_{k=1}^{5} g_{\boldsymbol{\omega}}(\boldsymbol{x},\boldsymbol{y})[k] = 1 \ \text{ and } \ g_{\boldsymbol{\omega}}(\boldsymbol{x},\boldsymbol{y})[k] \in (0,1) \tag{5}$$

JointGAN: Computational Cost

- The capacity of our model is increased without introducing additional computational cost
- The parameters of the generators are tied together:
  - $f_\alpha(\cdot) = f_\phi(\mathbf{0},\ \cdot\,)$ and $f_\beta(\cdot) = f_\theta(\mathbf{0},\ \cdot\,)$
  - *Almost no additional parameters when compared with ALI*



- A single 5-way critic (discriminator):
  - Also, *almost no additional parameters when compared with ALI*

Experiment: generated image pairs

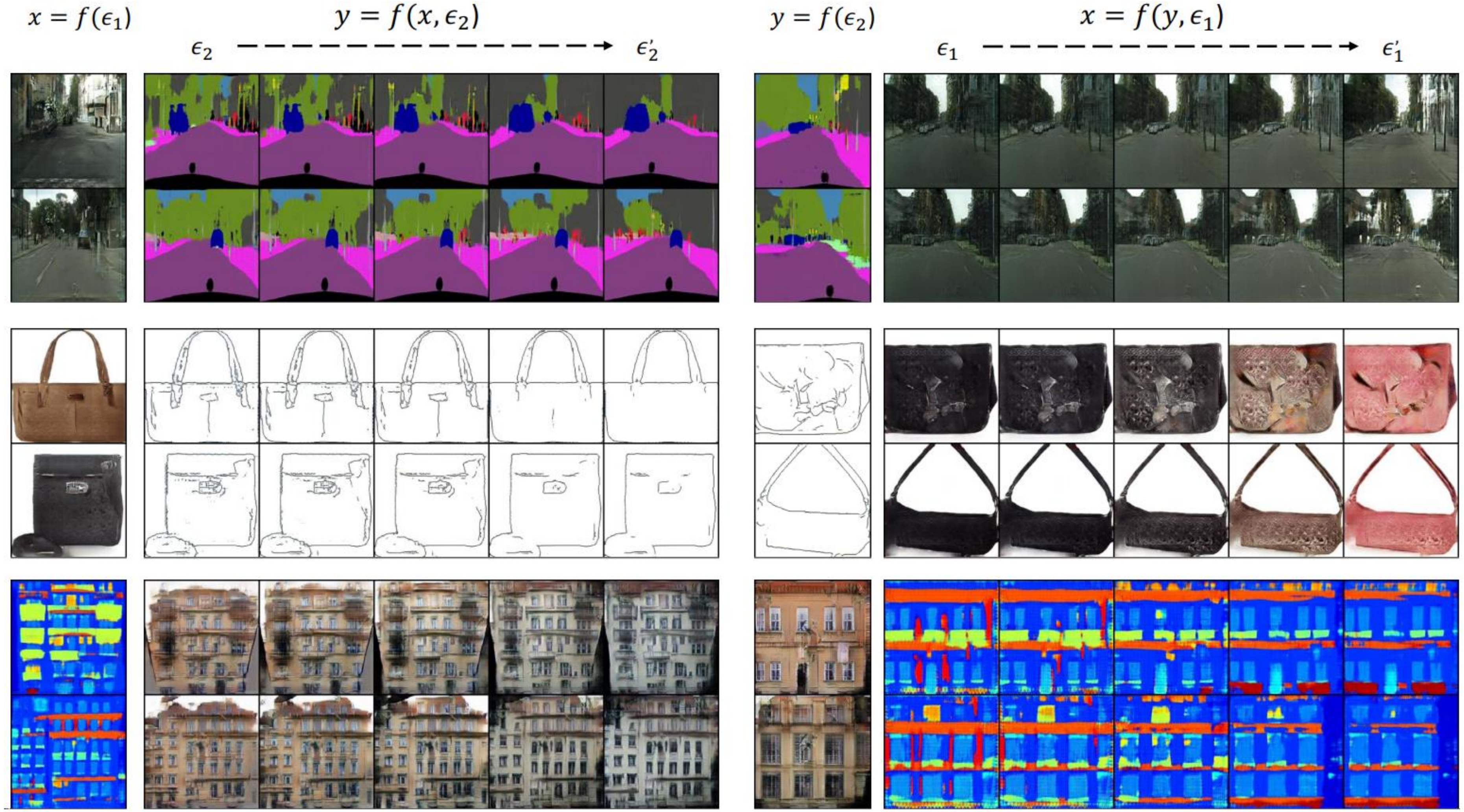
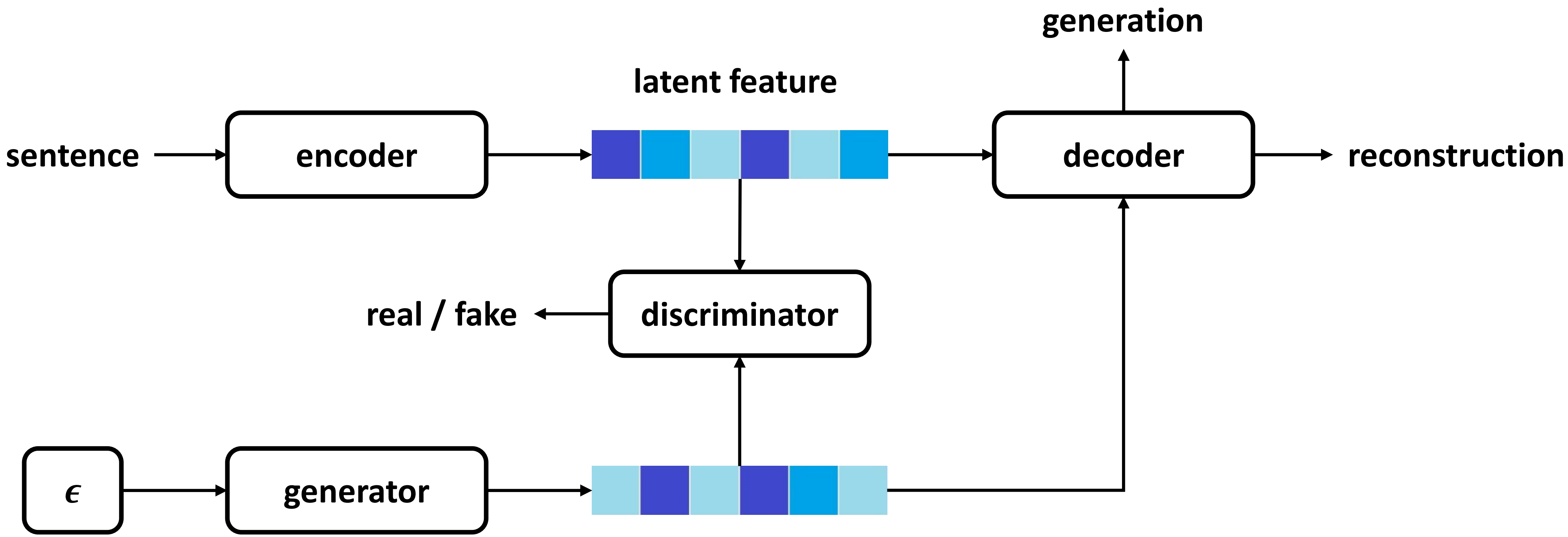


Figure 1. Generated paired samples from models trained on paired data.

Experiment: generated image-caption pairs

Figure 2. Generated paired samples of caption features and images. Left block: from generated images to caption features. Right block: from generated caption features to images.

Handle Discrete Text Data



Extension I: No Paired Data

- When paired draws from $p_5(\boldsymbol{x}, \boldsymbol{y}) = q(\boldsymbol{x}, \boldsymbol{y})$ are not available, cycle consistency is imposed:

$$\begin{aligned} & \min_{\boldsymbol{\theta},\boldsymbol{\phi}} \max_{\boldsymbol{\omega}} \ \mathcal{L}_{\text{JointGAN}}(\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\omega}) \\ & \quad = \textstyle\sum_{k=1}^{4} \mathbb{E}_{p_k(\boldsymbol{x},\boldsymbol{y})}[\log g'_{\boldsymbol{\omega}}(\boldsymbol{x}, \boldsymbol{y})[k]] + R_{\boldsymbol{\theta},\boldsymbol{\phi}}(\boldsymbol{x}, \boldsymbol{y}) \end{aligned} \quad (6)$$

where $R_{\theta,\phi}(\boldsymbol{x}, \boldsymbol{y})$ is a regularization term:

$$\begin{aligned} R_{\boldsymbol{\theta},\boldsymbol{\phi}}(\boldsymbol{x}, \boldsymbol{y}) = & \ \mathbb{E}_{\boldsymbol{x}\sim q(\boldsymbol{x}), \boldsymbol{y}\sim p_{\boldsymbol{\theta}}(\boldsymbol{y}|\boldsymbol{x}), \hat{\boldsymbol{x}}\sim p_{\boldsymbol{\phi}}(\boldsymbol{x}|\boldsymbol{y})} ||\boldsymbol{x} - \hat{\boldsymbol{x}}|| \\ & + \mathbb{E}_{\boldsymbol{y}\sim q(\boldsymbol{y}), \boldsymbol{x}\sim p_{\boldsymbol{\phi}}(\boldsymbol{x}|\boldsymbol{y}), \hat{\boldsymbol{y}}\sim p_{\boldsymbol{\theta}}(\boldsymbol{y}|\boldsymbol{x})} ||\boldsymbol{y} - \hat{\boldsymbol{y}}|| \end{aligned} \quad (7)$$

Extension I: No Paired Data

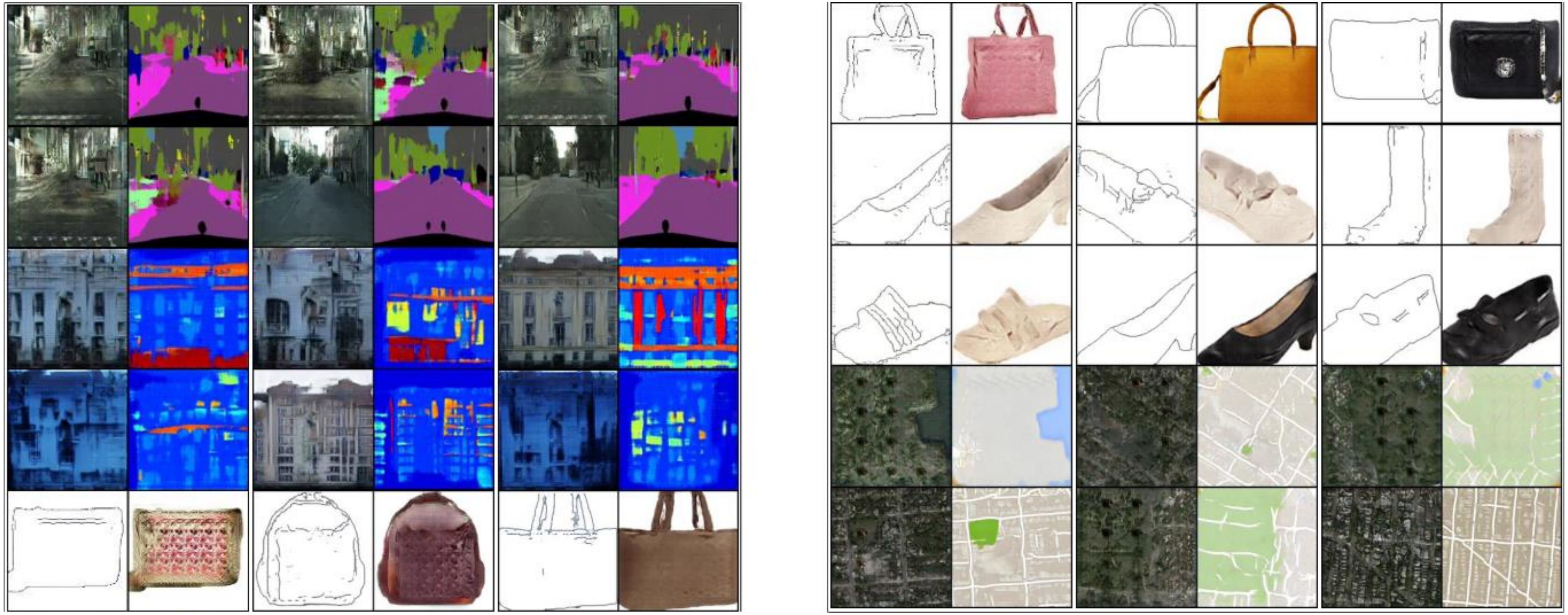


Figure 3. Generated paired samples from models trained on unpaired data.

Extension II: Multiple Domains

- Consider two specific forms for joint random variables $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$:

$$\begin{aligned} p(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) &= p_{\boldsymbol{\alpha}}(\boldsymbol{x}) p_{\boldsymbol{\nu}}(\boldsymbol{y}|\boldsymbol{x}) p_{\boldsymbol{\gamma}}(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{y}) \\ &= p_{\boldsymbol{\beta}}(\boldsymbol{z}) p_{\boldsymbol{\psi}}(\boldsymbol{y}|\boldsymbol{z}) p_{\boldsymbol{\eta}}(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{z}) \end{aligned} \tag{8}$$

- Assume only empirical draws from $q(\boldsymbol{x}, \boldsymbol{y})$ and $q(\boldsymbol{y}, \boldsymbol{z})$ are available. Let the critic be a 6-way softmax classifier:

$$\begin{aligned} p_{\boldsymbol{\alpha}}(\boldsymbol{x}) p_{\boldsymbol{\nu}}(\boldsymbol{y}|\boldsymbol{x}) p_{\boldsymbol{\gamma}}(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{y}), &\qquad p_{\boldsymbol{\beta}}(\boldsymbol{z}) p_{\boldsymbol{\psi}}(\boldsymbol{z}|\boldsymbol{y}) p_{\boldsymbol{\eta}}(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{z}) \\ q(\boldsymbol{x}) p_{\boldsymbol{\nu}}(\boldsymbol{y}|\boldsymbol{x}) p_{\boldsymbol{\gamma}}(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{y}), &\qquad q(\boldsymbol{x}) p_{\boldsymbol{\psi}}(\boldsymbol{z}|\boldsymbol{y}) p_{\boldsymbol{\eta}}(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{z}) \\ q(\boldsymbol{x}, \boldsymbol{y}) p_{\boldsymbol{\gamma}}(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{y}), &\qquad q(\boldsymbol{y}, \boldsymbol{z}) p_{\boldsymbol{\eta}}(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{z}) . \end{aligned} \tag{9}$$

Extension II: Multiple Domains

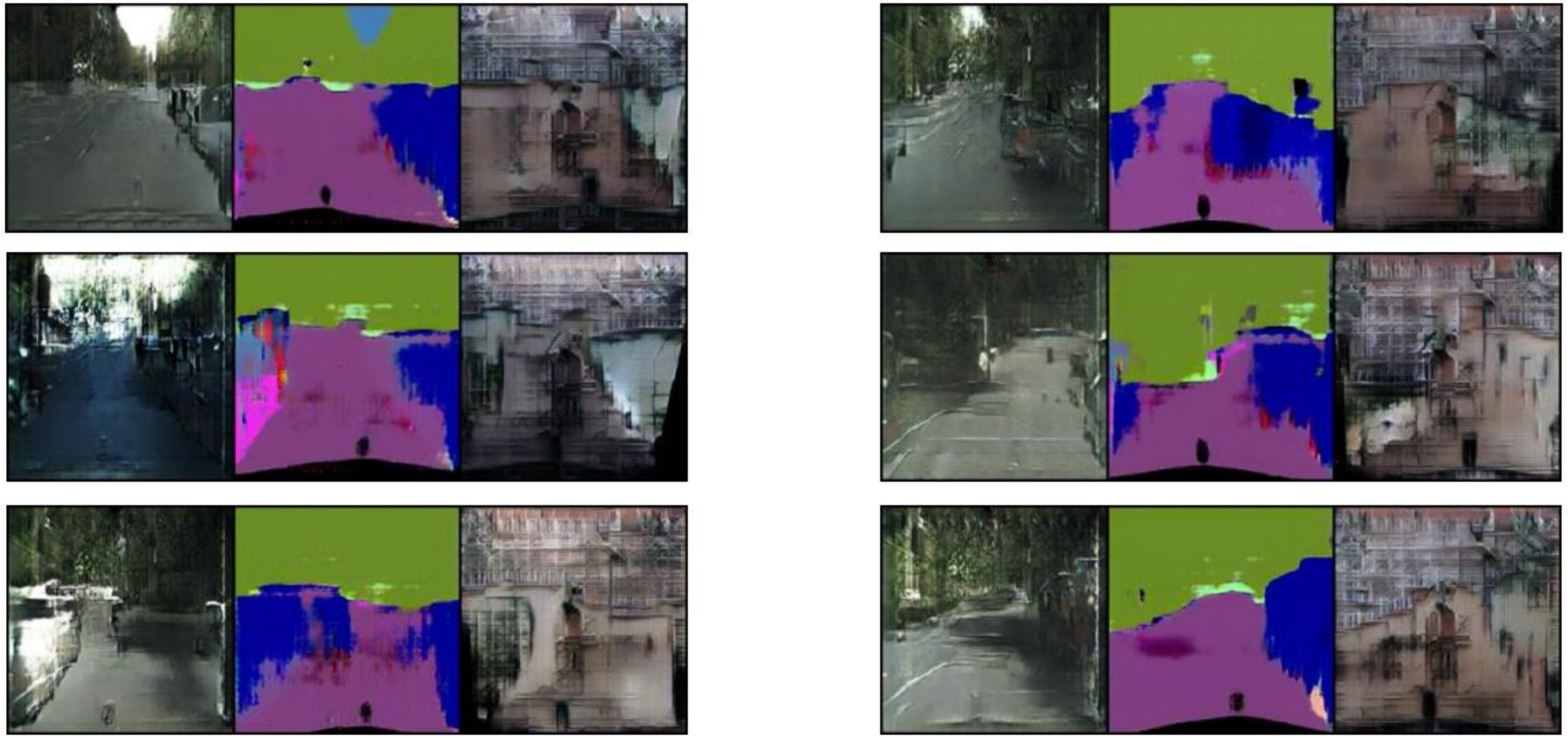


Figure 4. Generated paired samples from models trained on facades↔labels↔cityscapes. Left: sequentially generated from left to right. Right: generated from right to left.

Quantitative results

- Compare with a two-step baseline:
  - WGAN-GP is employed to model the two marginals
  - Pix2pix and CycleGAN are utilized to model the conditionals for the case w/wo paired empirical draws

Table 1. Human evaluation results on the quality of generated pairs.

Method	Realism	Relevance
Trained with paired data		
WGAN-GP + Pix2pix wins	2.32%	3.1%
JointGAN wins	17.93%	36.32%
Not distinguishable	79.75%	60.58%
Trained with unpaired data		
WGAN-GP + CycleGAN wins	0.13%	1.31%
JointGAN wins	81.55%	40.87%
Not distinguishable	18.32%	57.82%

Quantitative results

- We further define *Relevance Score*
  - Used to evaluate the quality and relevance of two generated images
  - Calculated as the cosine similarity between two images that are embedded into a shared latent space, which are learned via training a ranking model

Table 2. Relevance scores of the generated pairs on the five two-domain image datasets.

	edges↔shoes	edges↔handbags	labels↔cityscapes	labels↔facades	maps↔satellites
True pairs	0.684	0.672	0.591	0.529	0.514
Random pairs	0.008	0.005	0.012	0.011	0.054
Other pairs	0.113	0.139	0.092	0.076	0.081
WGAN-GP + Pix2pix	0.352	0.343	0.301	0.288	0.125
JointGAN (paired)	0.488	0.489	0.377	0.364	0.328
WGAN-GP + CycleGAN	0.203	0.195	0.201	0.139	0.091
JointGAN (unpaired)	0.452	0.461	0.339	0.341	0.299

Summary

- We propose JointGAN for multi-domain joint distribution learning
- JointGAN provides freedom to draw samples from various marginalized or conditional distributions
- Besides image pairs, we also provided generated image-caption pairs

Come to our poster for details
@ Hall B #109

Yunchen Pu, Shuyang Dai, Zhe Gan, Weiyao Wang, Guoyin Wang, Yizhe Zhang, Ricardo Henao, Lawrence Carin

Facebook & Duke University & Microsoft Research