
Appendix for “JointGAN: Multi-Domain Joint Distribution Learning with Generative Adversarial Nets”

**Yunchen Pu¹ Shuyang Dai² Zhe Gan³ Weiyao Wang² Guoyin Wang² Yizhe Zhang³ Ricardo Henao²
Lawrence Carin²**

A. Proof of Proposition 1

We first consider a general optimization problem as the following

$$\min \mathcal{L}(f_1, \dots, f_K) = \min \sum_{k=1}^K p_k(\mathbf{x}, \mathbf{y}) \log \frac{f_k}{\sum_{i=1}^K f_i}. \quad (1)$$

For any f_k , if we fix all other variables, the local optimal f_k^* is given by:

$$\frac{\partial \mathcal{L}}{\partial f_k} = \frac{p_k(\mathbf{x}, \mathbf{y})}{f_k} - \sum_{j=1}^K \frac{p_j(\mathbf{x}, \mathbf{y})}{\sum_{i=1}^K f_i} = 0 \quad (2)$$

$$\Leftrightarrow f_k = \left(\sum_{i=1, i \neq k}^K f_i \right) \frac{p_k(\mathbf{x}, \mathbf{y})}{\sum_{j=1, j \neq k}^K p_j(\mathbf{x}, \mathbf{y})}. \quad (3)$$

The K equations below

$$f_k = \left(\sum_{i=1, i \neq k}^K f_i \right) \frac{p_k(\mathbf{x}, \mathbf{y})}{\sum_{j=1, j \neq k}^K p_j(\mathbf{x}, \mathbf{y})}, \quad \text{for } k = 1, \dots, K, \quad (4)$$

have a global solution with

$$f_k = C p_k(\mathbf{x}, \mathbf{y}), \quad \text{for } k = 1, \dots, K, \quad (5)$$

where $C \neq 0$ is a constant. Let $\hat{f}_k = \frac{f_k}{\sum_{i=1}^K f_i}$, the global optimal of (1) is achieved at $\hat{f}_k = \frac{p_k(\mathbf{x}, \mathbf{y})}{\sum_{j=1}^K p_j(\mathbf{x}, \mathbf{y})}$.

Let $K = 5$ and $\hat{f}_k = g_{\omega}(\mathbf{x}, \mathbf{y})[k]$. This indicates that with fixed (θ, ϕ) , the optimal critic g_{ω} in (10) in the main paper is achieved at

$$g_{\omega^*}(\mathbf{x}, \mathbf{y})[k] = \frac{p_k(\mathbf{x}, \mathbf{y})}{\sum_{j=1}^5 p_j(\mathbf{x}, \mathbf{y})}. \quad (6)$$

With optimal g_{ω^*} , the objective (10) in the main paper can be expressed as

$$\mathcal{L} = \sum_{k=1}^5 \mathbb{E}_{p_k(\mathbf{x}, \mathbf{y})} \log \frac{p_k(\mathbf{x}, \mathbf{y})}{\sum_{j=1}^5 p_j(\mathbf{x}, \mathbf{y})} \quad (7)$$

$$= -5 \log 5 + \sum_{k=1}^5 \text{KL}\left(p_k(\mathbf{x}, \mathbf{y}) \middle\| \frac{\sum_{j=1}^5 p_j(\mathbf{x}, \mathbf{y})}{5}\right). \quad (8)$$

The global minimum of (8) is achieved at $p_1(\mathbf{x}, \mathbf{y}) = p_2(\mathbf{x}, \mathbf{y}) = p_3(\mathbf{x}, \mathbf{y}) = p_4(\mathbf{x}, \mathbf{y}) = p_5(\mathbf{x}, \mathbf{y})$.

B. Calculating the relevance score

We use relevance score to evaluate the quality and relevance of two generated samples. The relevance score is calculated by the cosine similarity between random variables \mathbf{x} and \mathbf{y} :

$$R(\mathbf{x}, \mathbf{y}) = \frac{\mathbf{f}(\mathbf{x})^\top \mathbf{g}(\mathbf{y})}{\|\mathbf{f}(\mathbf{x})\| \cdot \|\mathbf{g}(\mathbf{y})\|}, \quad (9)$$

where $f(\cdot)$ and $g(\cdot)$ are two feature extractors (typically implemented as CNNs) that embed the two high-dimensional data \mathbf{x} and \mathbf{y} into a shared low-dimensional latent space. After the two feature extractors are trained on the paired dataset $\{\mathbf{x}_i, \mathbf{y}_i\}_{i=1,N}$, we can use the cosine similarity to evaluate the relevance of two synthesized joint samples $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{y}}$.

To learn the feature extractors, a ranking model (Huang et al., 2013) is trained as below. First, we consider using the samples of \mathbf{x} to query samples of \mathbf{y} . Given that we have the relevance scores between the query \mathbf{x} and each of the target sample \mathbf{y}_j : $R(\mathbf{x}, \mathbf{y}_j)$, we define the posterior probability of the correct candidate given the query by the following softmax function

$$P(\mathbf{y}^+ | \mathbf{x}) = \frac{\exp(\gamma R(\mathbf{x}, \mathbf{y}^+))}{\sum_{\mathbf{y}' \in \mathcal{Y}} \exp(\gamma R(\mathbf{x}, \mathbf{y}'))}, \quad (10)$$

where \mathbf{y}^+ denotes the correct target sample (the positive sign denotes that it is a positive sample), γ is a tuning hyper-parameter in the softmax function (to be tuned empirically on a validation set). \mathcal{Y} denotes the set of candidate samples to be ranked, which includes the positive sample \mathbf{y}^+ and J randomly selected incorrect (negative) candidates $\{\mathbf{y}_j^-; j = 1, \dots, J\}$.

Similarly, we also consider using samples of \mathbf{y} to query samples of \mathbf{x} . A corresponding posterior is defined as follow:

$$P(\mathbf{x}^+ | \mathbf{y}) = \frac{\exp(\gamma R(\mathbf{y}, \mathbf{x}^+))}{\sum_{\mathbf{x}' \in \mathcal{X}} \exp(\gamma R(\mathbf{y}, \mathbf{x}'))}. \quad (11)$$

The model parameters are learned to maximize the likelihood of the correct candidates given the queries across the training set, in both the above directions. That is, we minimize the following loss function

$$L(\boldsymbol{\theta}) = -\log \prod_{(\mathbf{x}, \mathbf{y}^+)} P(\mathbf{y}^+ | \mathbf{x}) - \log \prod_{(\mathbf{y}, \mathbf{x}^+)} P(\mathbf{x}^+ | \mathbf{y}),$$

where the product is over all training samples, and $\boldsymbol{\theta}$ denotes the parameters (to be learned), which includes all the model parameters in the deep feature extractors. The above cost function is minimized by backpropagation and (mini-batch) stochastic gradient descent.

C. More Results

C.1. Results for the two-step baseline



Figure 1. Generated paired samples trained with WGAN-GP+Pix2pix on paired data.



Figure 2. Generated paired samples trained with WGAN-GP+CycleGAN on unpaired data.

C.2. Additional results for JointGAN on modeling multi-domain images

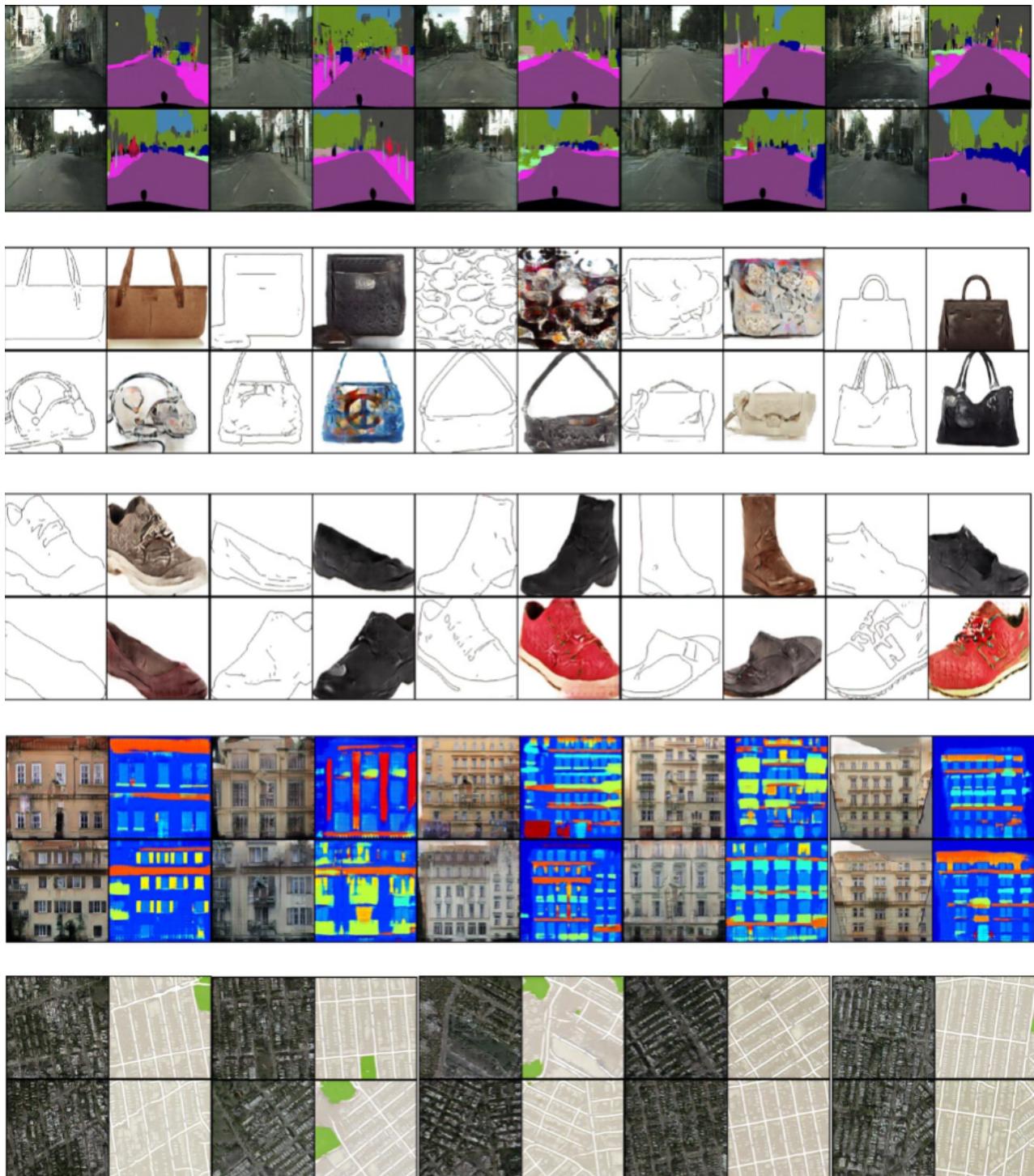


Figure 3. Generated paired samples trained with paired data.



Figure 4. Generated paired samples trained with unpaired data.

C.3. Additional results for JointGAN on modeling caption features and images

	small bird with yellow and black coloring on its body and a small beak		a small brown bird with a grey cheek patch and a long pointy beak
	this bird has a yellow throat and belly with a black eyebrow and a light colored wings		this bird has a yellow belly and breast and a grey crown and a large beak
	this bird has a white belly and cheek patch and a very short beak		a medium sized grey and white bird with a short curved bill and a yellow throat
	this yellow and black bird has a tiny beak and a brown wings		a small bird with a black eye and a grey and white breast and a pointy beak
	this bird has a gray crown and a black beak and orange feet		this is a red and black bird with a red eye and a pointy beak
	this small bird has a white belly and a black head		a small bird with a short pointed bill and a brown and white striped breast and head
	a small brown and yellow bird with a black eye		this small bird has a brown crown and a white and brown belly with a short beak
	a small sized bird that has a red throat and breast and a long pointed black bill		this is a yellow and red bird with long feet
colorful bird with yellow belly yellow breast and yellow and grey wings		this small bird has a red belly and a reddish brown wing and tail and a light red eye ring	
small red and gray colored belly and breast red crown and head and red feet and feet		this bird has a red crown black wings and a brown crown	
small bird with a grey belly and throat and a striped crown with a brown back and wings		a light brown and white bird with spots of brown on its head and a mix of white and brown feathers on its head	
a small yellow and yellow and grey speckled bird with a yellow breast and belly		a small blue bird with a green and white wings and a black crown	
a white and grey bird with a large orange and white crown and a small orange beak		small bird with a brown and white belly and a brown head with a pointy beak	
this bird has a white belly with a brown crown and a small head and a beak		this is a black bird with a gray beak and a pointy red beak	
a small sized bird with black feathers and orange beak and a green tail		small brown and black bird with red eyes and brown beak and head	
a bird with yellow and green feathers and a brown belly and a short pointy beak and a dark brown eye ring		a bright yellow bird with a black head and beak and blue wings with a short pointy bill	

Figure 5. Generated paired samples of caption features and images. Top block: from generated images to caption features. Bottom block: from generated caption features to images.

C.4. Model architecture for JointGAN on modeling caption features and images

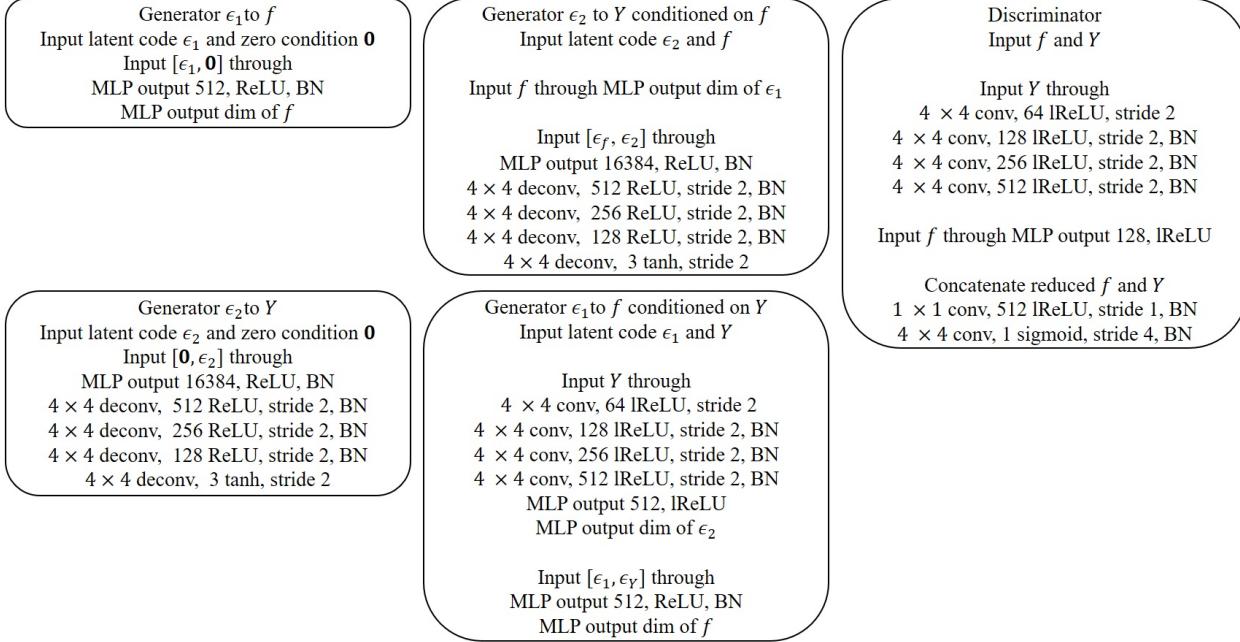


Figure 6. Model architecture for JointGAN on modeling caption features and images.

References

- Huang, P.-S., He, X., Gao, J., Deng, L., Acero, A., and Heck, L. Learning deep structured semantic models for web search using clickthrough data. In *CIKM*, 2013.