# VARIATIONAL BAYESIAN FRAMEWORK FOR ADVANCED IMAGE GENERATION WITH DOMAIN-RELATED VARIABLES

Yuxiao Li

IEEE ICASSP 2022: Paper Presentation

Electronic Engineering Department
Tsinghua University

# Outline

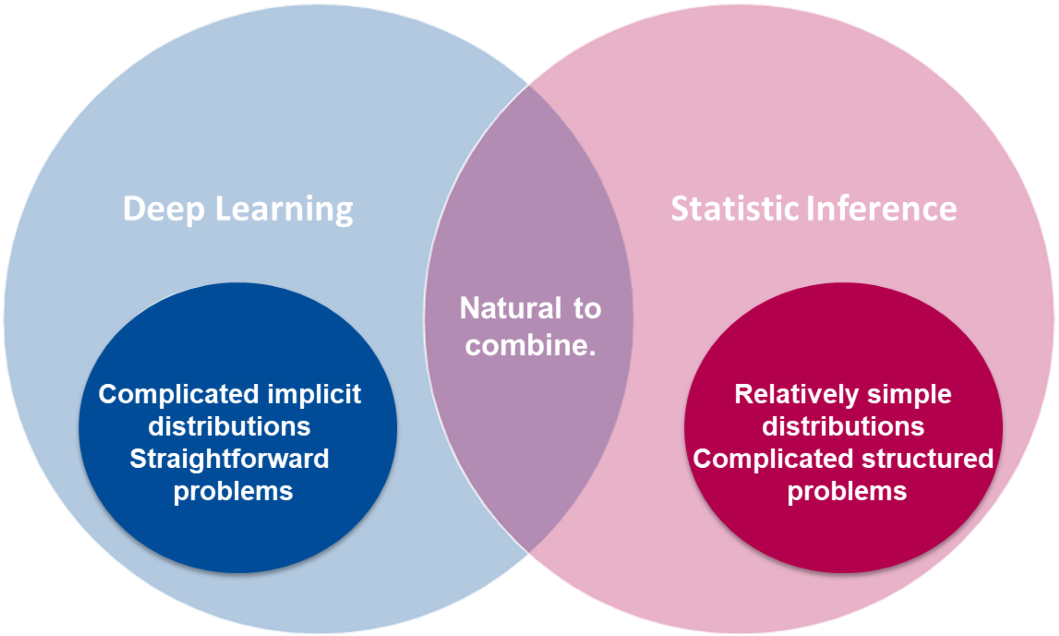
- Introduction
- Variational Bayesian Method
- Realization of Conditional Generation Tasks
- Experimental Results

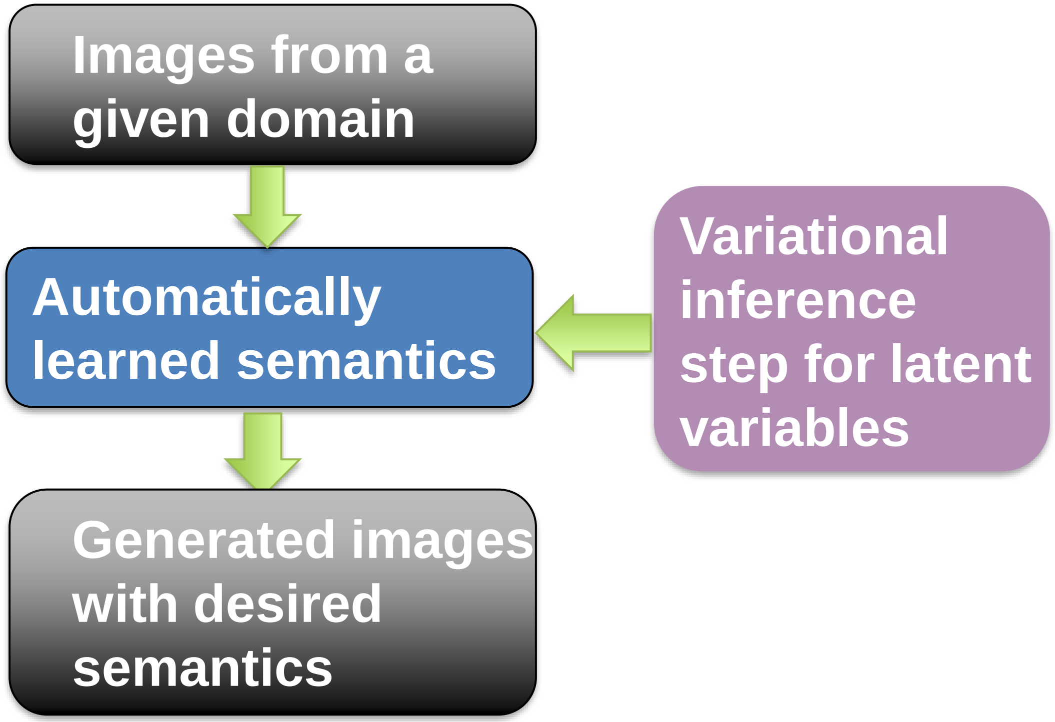
# Introduction

- **Conditional Image generation**
  - Deep generative models (DGMs): Learn complicated data distributions, can scale to the generation of conditional image generation for different problems.
  - Applications: Single Image Super Resolution (SISR), image colorization, image inpainting, style transferring, semantic attribute synthesis.
  - Existing Works for conditional image generation
    - **With explicit label annotations**: InfoGAN, Conditional GAN
    - **With domain annotations**: Pixel GAN, Bicycle GAN, etc.
    - **Without annotations (i.e., *Unsupervised image-to-image translation*): Cycle GAN, Dual GAN, MUNIT, etc.**
  - Description: Images in a target domain are obtained correspond to images in the source domain without paired samples during training.
  - Challenges: The information for condition is indirectly given via samples of marginal distributions in each domain, ill-posed without further assumptions.

# Introduction

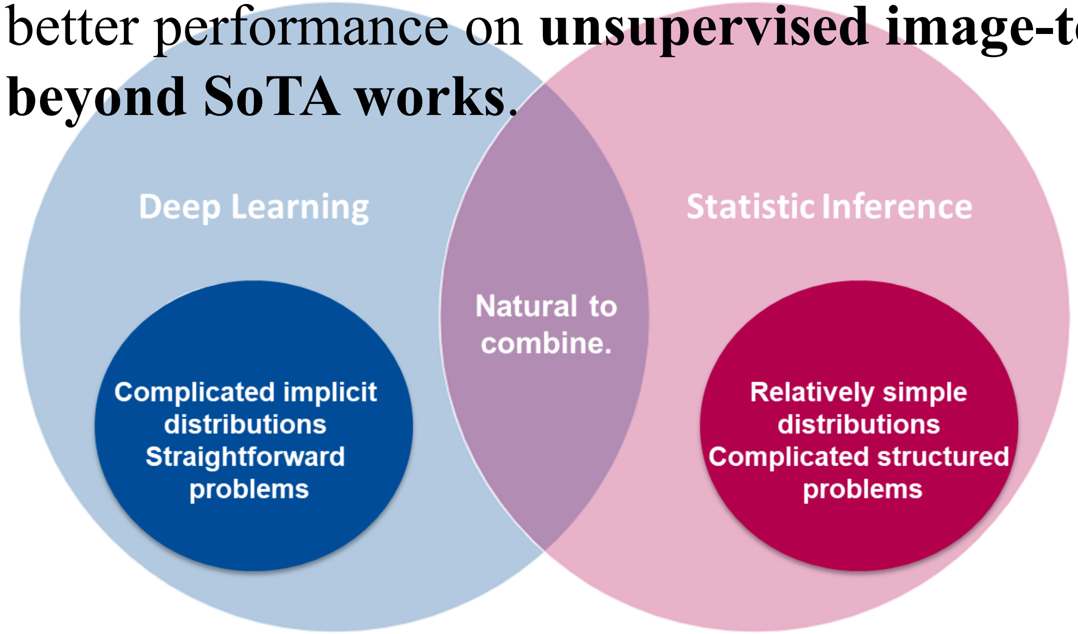
- **Probabilistic Framework**
  - Statistical viewpoint: These problems can be described by a generative model with latent variables, corresponding to different semantics.
  - **Solution**: present **a unified Bayesian framework** with an inference step on latent variables, enabling efficient unsupervised image-to-image translation and semantic editing**.**



# Introduction

- **Contributions**
  - We **propose a deep generative network** for image translation tasks, where latent variables of semantics are inferred via a variational lower bound in learning.
  - Driven **by a rigorous probabilistic model**, the proposed method has a clear interpretation of the loss terms, and improved generality to encompass multiple variants.
  - Experimental results on several public datasets illustrate that the proposed method achieves better performance on **unsupervised image-to-image translation**, and enables **variants beyond SoTA works**.

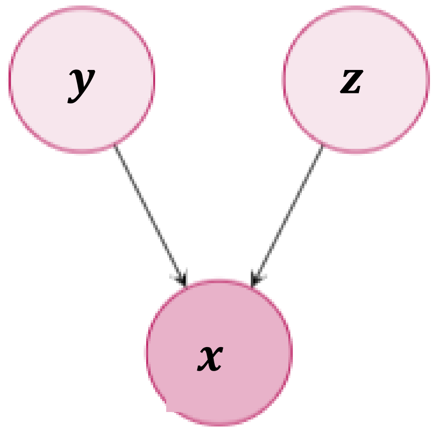


# Outline

- Introduction
- Variational Bayesian Method
- Realization of Conditional Generation Tasks
- Experimental Results

# Variational Bayesian Method

- **Generative Model**
  - The generative process of **an image sample** $x$ in certain domain involves **two latent variables**:
    - A **domain-related variable** $y$ for features specific to the domain, referred to as the 'style' variable.
    - An independent **domain-unrelated variable** $z$ that describes general features, referred to as the 'content' variable (following the classical nomenclature in NST)..
  - The generation from the latent space to the image space can be obtained via the **likelihood distribution** $p(x|y, z)$.
- **How to disentangle different semantics?**
  - Conduct an inference step on latent variables.
  - Construct a deep network to realize the deep variational learning.




$x$: domain image

$y$: **domain-related variable for style**

$z$: **domain-unrelated variable for content**

# Variational Bayesian Method

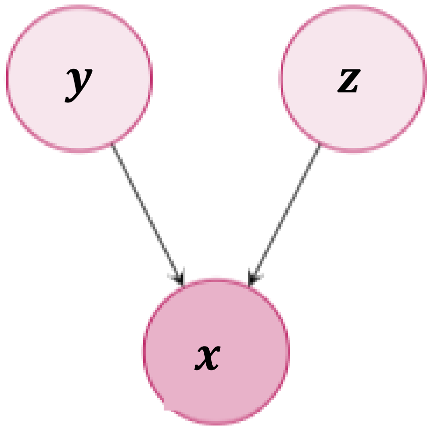
- **Variational Lower Bound for Data Likelihood**
  - The evidence lower bound (ELBO) of the marginal likelihood of data can be written as

$$\log p(\mathbf{x}) \geq \mathcal{L}(q;\mathbf{x})$$
$$= \mathbb{E}_{q(\mathbf{y},\mathbf{z}|\mathbf{x})}\left[\log p(\mathbf{x}|\mathbf{y},\mathbf{z})\right]$$
$$- \mathrm{D_{KL}}\left(\boxed{q(\mathbf{y}|\mathbf{x})}\|p(\mathbf{y})\right) - \mathrm{D_{KL}}\left(\boxed{q(\mathbf{z}|\mathbf{x})}\|p(\mathbf{z})\right)$$

    where $q(\mathbf{y}|\mathbf{x})$ and $q(\mathbf{z}|\mathbf{x})$ are variational distributions introduced to approximated the true posterior distributions.
  - Such bound can then be used to find suitable approximated distributions for $q(\mathbf{y}|\mathbf{x})$ and $q(\mathbf{z}|\mathbf{x})$, i.e.,

$$q^* = \arg\max_{q \in Q} \mathcal{L}(q;\mathbf{x})$$

  - Construct deep neural networks to learn the variational distributions.



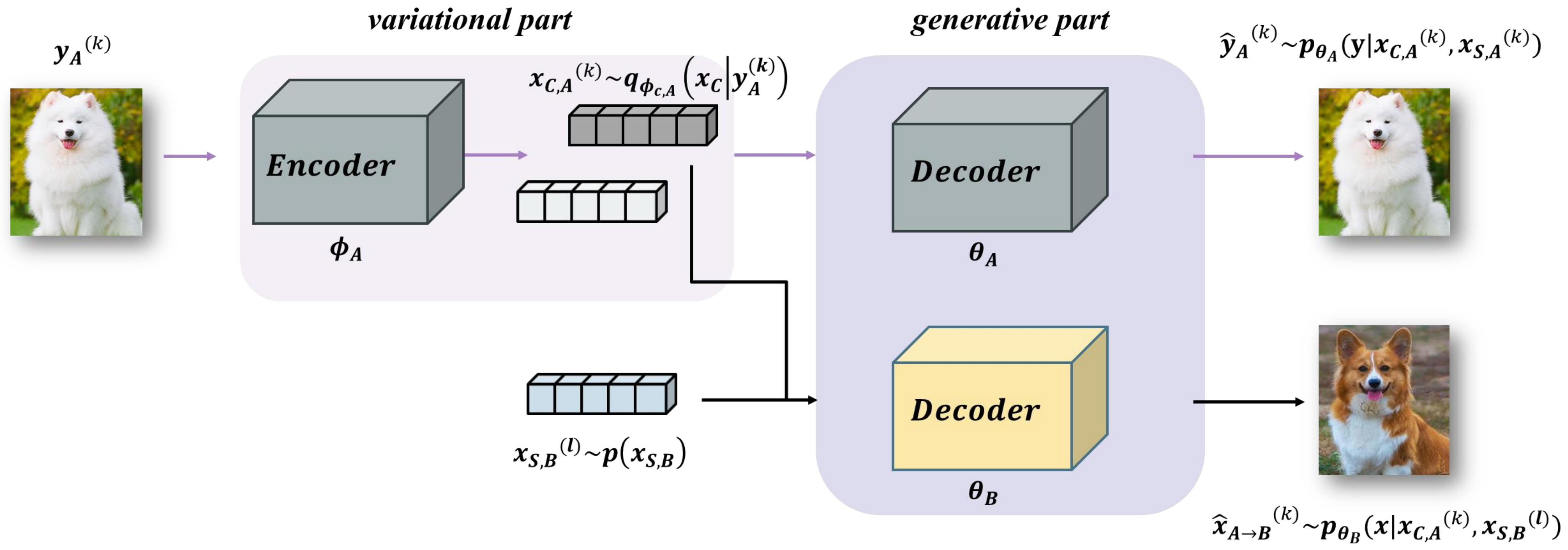

$x$: domain image

$y$: **domain-related variable for style**

$z$: **domain-unrelated variable for content**

# Variational Bayesian Method

- **Variational Bayesian Image Translation Network**
  - **Target: to disentangle different semantics**
  - Conduct an inference step on latent variables. → Variational Inference
  - Construct a deep network to realize the deep variational learning. → Deep Learning



# Outline

- Introduction
- Variational Bayesian Method
- Realization of Conditional Generation Tasks
- Experimental Results

# Variational Bayesian Method

- **Unsupervised Image-to-Image Generation**
  - Unsupervised image-to-image translation from painting to photo obtained by our model is illustrated in Fig.1.
  - The painting of a lake in domain $\boldsymbol{S}$ is translated to a photo of the same lake, with the style variable $y_T$ from photo domain $\boldsymbol{T}$ and the content variable $\boldsymbol{z_S}$ from painting domain $\boldsymbol{S}$.

- **Multiple Variants**
  - Multimodal style editing
  - Multimodal content editing
  - Mixed domain translation

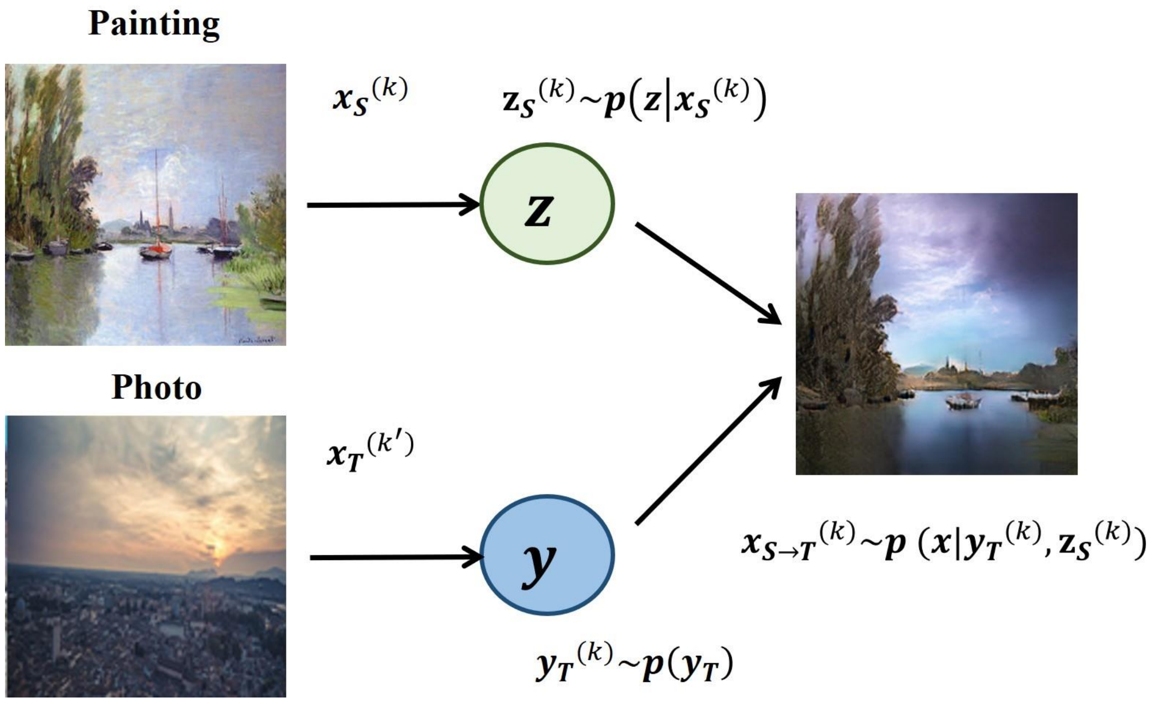
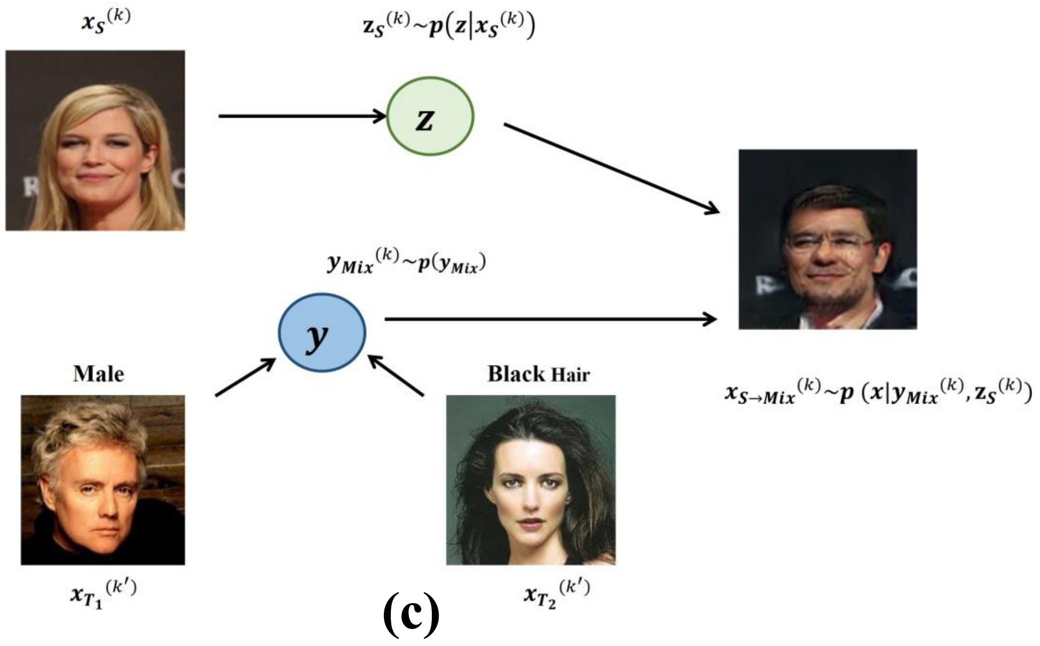



**Figure 1. Unsupervised image-to-image translation.**

# Variational Bayesian Method

- **Multiple Variants**
  - Multimodal style editing → (a)
  - Multimodal content editing → (b)
  - Mixed domain translation → (c)



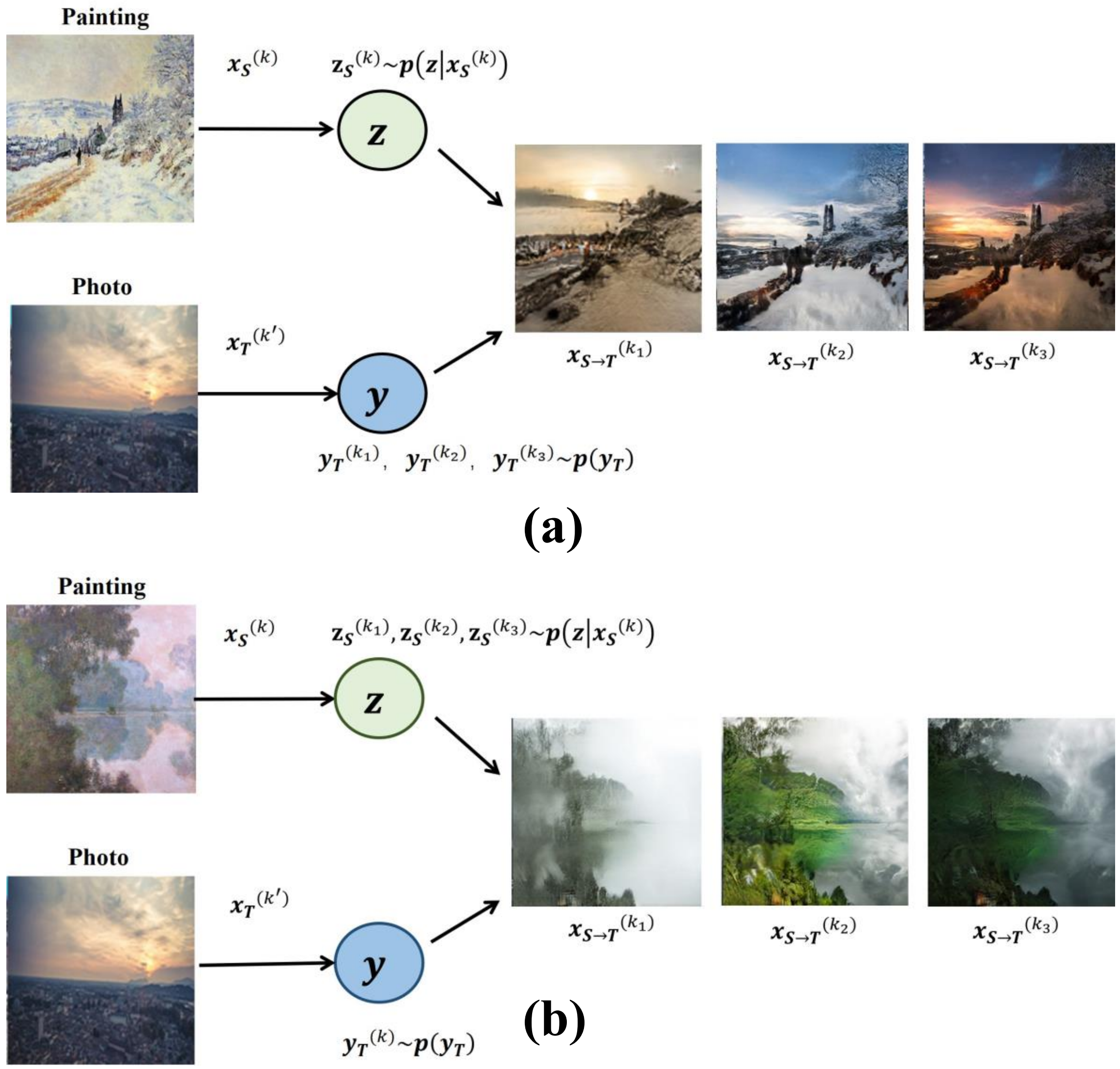




**Figure 2. Three variants for conditional image generation introduced by the proposed Bayesian framework.**

# Outline

- Introduction
- Variational Bayesian Method
- Semi-Supervised Learning Scheme
- Experimental Results

# Experimental Results

- **Quantitative Results on Supervised Image-to-Image Translation**
  - The first two columns show the input and GT while the next four show the generated.
  - **It can be seen that the proposed method can generated shoes from edges with great diversity as well as realism.**


**Figure 3. Qualitative results on dataset 'Edges to Shoes'.**

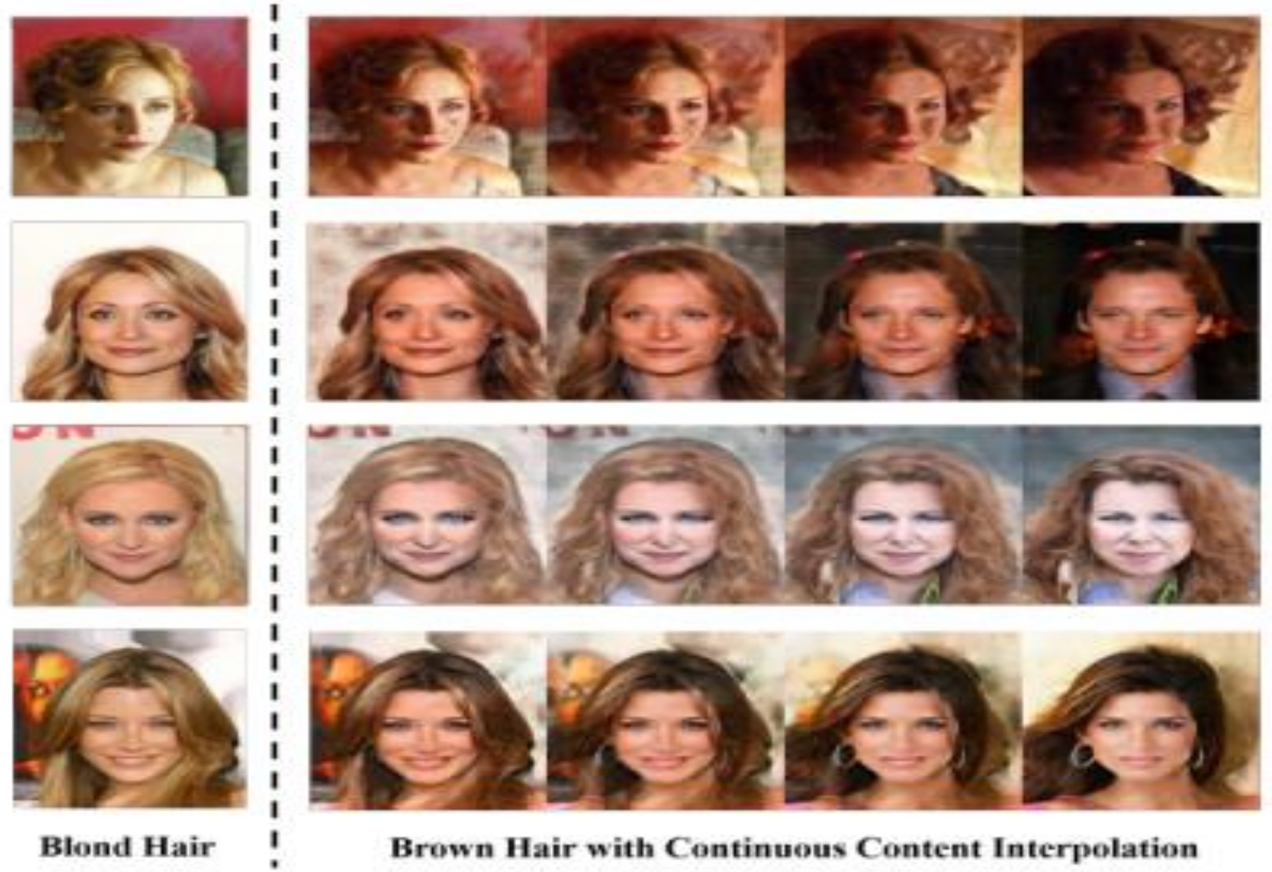
# Experimental Results

- **Quantitative Results on Unsupervised Supervised Image-to-Image Translation**
  - The first columns show the input and the next three show the generated image sampling different values for content and style variables.
  - It can be seen that **paintings are translated to photos with different semantics**.


**Figure 4. Qualitative results on dataset ’Photo to Monet’s Painting’.**

# Experimental Results

- **Quantitative Results on Advanced Variants**

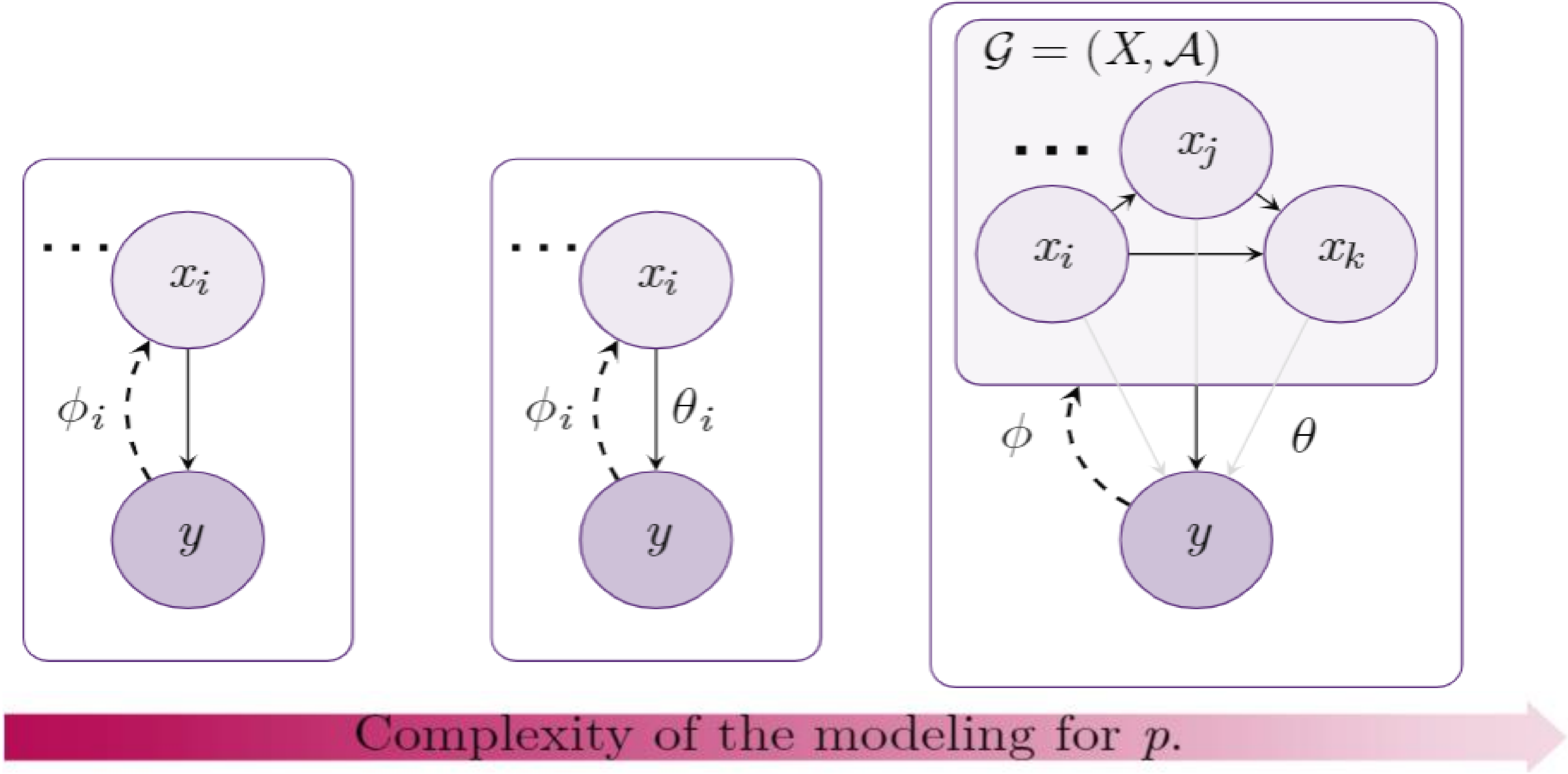



(a)


(b)

**Figure 5. Qualitative results on dataset 'CelebA'. (a) Continuous transformation in content; (b) Multiple styles synthesis.**

# Research Agenda



# Thank you for listening!

Electronic Engineering Department
Tsinghua University