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

Yuxiao Li†, Santiago Mazuelas‡, and Yuan Shen†

†Department of Electronic Engineering, Tsinghua University, Beijing, China ‡BCAM-Basque Center for Applied Mathematics, Bilbao, Spain

Emails: li-yx18@mails.tsinghua.edu.cn, smazuelas@bcamath.org, shenyuan ee@tsinghua.edu.cn

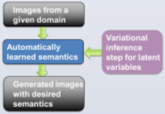
INTRODUCTION

Motivation: Make a general-purpose statistic framework for conditional image generation.

- From a statistical viewpoint, these problems can be described well by a latent variable model (LVM).
- Specifically, semantic features can be viewed as latent variables while the generation can be conducted by inferring the conditional distribution of images given the variables corresponding to desired semantics.
- The idea of disentangling codes for different semantics is partially discussed by [1, 2], while *seldom derived from first principles via statistic modeling*.

In this paper, we present **a novel probabilistic framework for a general class of conditional image generative problems**. Our contributions can be summarized as follows,

- Propose a deep generative network for image translation tasks, with latent variables of semantics inferred via variational inference.
- Driven by probabilistic modeling, the method has clear interpretation and improved generality to multiple variants.
- Experimental results on illustrate that the proposed method achieves better performance on unsupervised image-to-image translation, and enables variants beyond SoTA works.

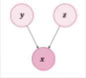


BAYESIAN FRAMEWORK FOR IMAGE GENERATION

The generative process of an image sample $x \in X$ in certain domain involves two latent variables: a domain-related variable y and an independent domain-unrelated variable z, referred to as 'style' and 'content' variable, following the classical nomenclature in NST [3].

The generation from the latent space to the image space can be obtained via the likelihood distribution $p(x|y,z)$.

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



x: domain image

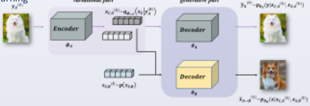
y: domain-related variable for style

z: domain-unrelated variable for content

VARIATIONAL BAYESIAN TRANSLATION NETWORK

Target: to disentangle different semantics for conditional image generation.

- Conduct an inference step on latent variables. → Variational Inference
- Construct a deep network to realize efficient image generation. → Deep Learning



EXPERIMENTAL RESULTS

Fig. 1 Supervised Image-to-Image Translation with Multimodal

Fig. 2 Unsupervised Image-to-Image Translation with Content and Style Editing

Fig.3 Multiple styles synthesis. *Fig.4 Mixed Continuous transformation in content*

Fig. 1

Fig. 2

Fig. 3

Fig. 4

References

[1] X. Huang, M.-Y. Liu, S. J. Belongie, and J. Kautz, "Multimodal unsupervised image-to-image translation," in ECCV, 2018.

[2] T. Park, J.-Y. Zhu, O. Wang, J. Lu, E. Shechtman, A. A. Efros, and R. Zhang, "Swapping autoencoder for deep image manipulation," ArXiv, vol. abs/2007.00653, 2020.

[3] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image style transfer using convolutional neural networks," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2414–2423, 2016.