Variational Autoencoder with Arbitrary Conditioning (VAEAC)

Dhruv Panchal (202411042)
Preet Shah (202411053)

Guide: Prof. M.V. Joshi
Subject: Computer Vision

October 1, 2025

Outline

Traditional VAE (Image Generation)

- When we wanted a model to generate entire new images, Variational Autoencoders (VAE) were introduced.
- Learns the distribution $p(x)$ of the dataset.
- Uses a latent variable $z \sim p(z)$ to capture hidden features.
- **Encoder:** $q_\phi(z|x)$ approximates posterior.
- **Decoder:** $p_\theta(x|z)$ reconstructs or generates a full image from z.
- *Output: Entire image generated from latent code.*

Conditional VAE (Image Generation with Condition)

- When there was a need to generate images based on specific conditions, Conditional VAE (CVAE) was developed.
- Learns the conditional distribution $p(x|y)$ where y is a label or attribute.
- **Prior:** $p_\psi(z|y)$ depends on condition y.
- **Posterior:** $q_\phi(z|x, y)$ ensures latent space respects condition.
- *Output: Images generated that are consistent with the given condition y (e.g., digit class, gender attribute).*

Problem Statement

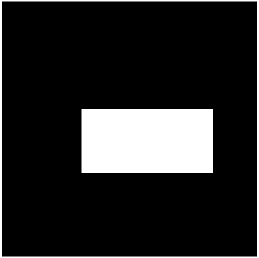
- **Real-world challenge:** Data often has *arbitrary missing features* (e.g., random missing pixels, incomplete records).
- **VAEAC Solution:** Learns $p(x_b|x_{1-b}, b)$, i.e., *generate missing parts of data given any observed parts and a mask* b.
- Paper: "Variational Autoencoder with Arbitrary Conditioning"
- Conference: International Conference on Learning Representations (ICLR), 2019. Author: O. Ivanov, M. Figurnov, and D. Vetrov. [1]

Dataset Information

- **MNIST [2]**
  - 60,000 train, 10,000 test grayscale digit images.
  - Image size: 28×28
- **CelebA [3]**
  - 162,770 train, 19,867 validation, 19,962 test color face images.
  - Image size: 178×218

Data Representation in VAEAC [1]

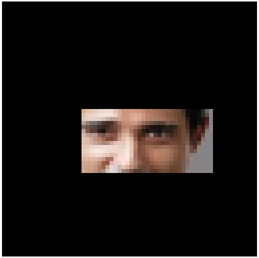
x (Input Image)



b (Mask)

Data Representation in VAEAC

x_{1-b} (Observed Part)



x_b (Missing Part)

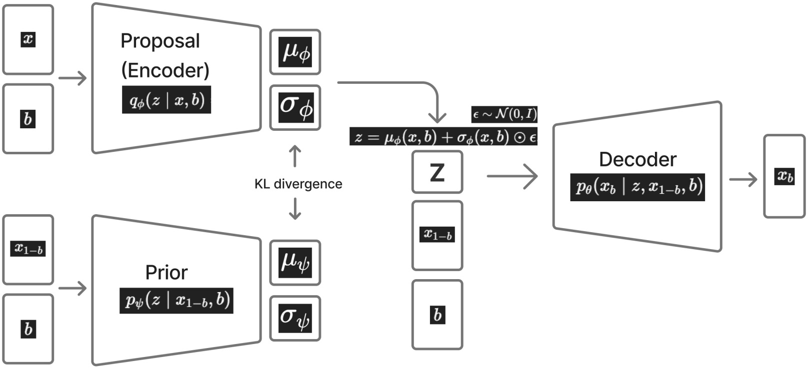
What Is Our Goal?

- Goal: Learn $p(x_b|x_{1-b}, b)$ for arbitrary mask b
- Handles missing features and arbitrary conditioning

Model Architecture

- Proposal network: $q_\phi(z|x, b)$
- Generative network: $p_\theta(x_b|z, x_{1-b}, b)$
- Prior network: $p_\psi(z|x_{1-b}, b)$

Model - Training Pipeline

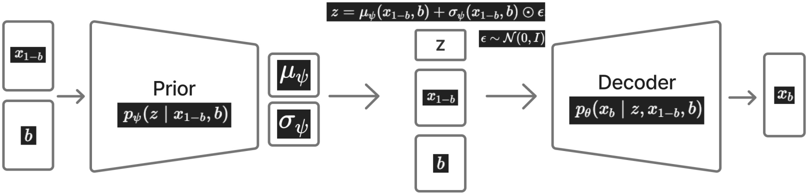


Training Objective Function

$$\mathcal{L}_{VAEAC}(x, b; \theta, \psi, \phi) =$$

$$\mathsf{E}_{q_\phi(z|x,b)} \log p_\theta(x_b|z, x_{1-b}, b) - D_{KL}(q_\phi(z|x, b)||p_\psi(z|x_{1-b}, b))$$

Model At Inference Time



Our Next Steps

- Model Reproduction & Validation
- Experimenting with Sequential Conditioning

References

[1] O. Ivanov, M. Figurnov, and D. Vetrov, "Variational autoencoder with arbitrary conditioning," in *International Conference on Learning Representations (ICLR)*, 2019.

[2] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[3] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 3730–3738, 2015.

Thank You!