GCVAE: Generalized-Controllable Variational Autoencoder

Ezukwoke K.[1], Hoayek A.[1], Batton-Hubert M.[1] and Boucher X.[1]

[1]Mines Saint-Etienne, Univ. Clermont Auvergne, CNRS UMR 6158 LIMOS, Henri FAYOL institute, F-42023, Saint-Etienne, France

Motivation | Methodology

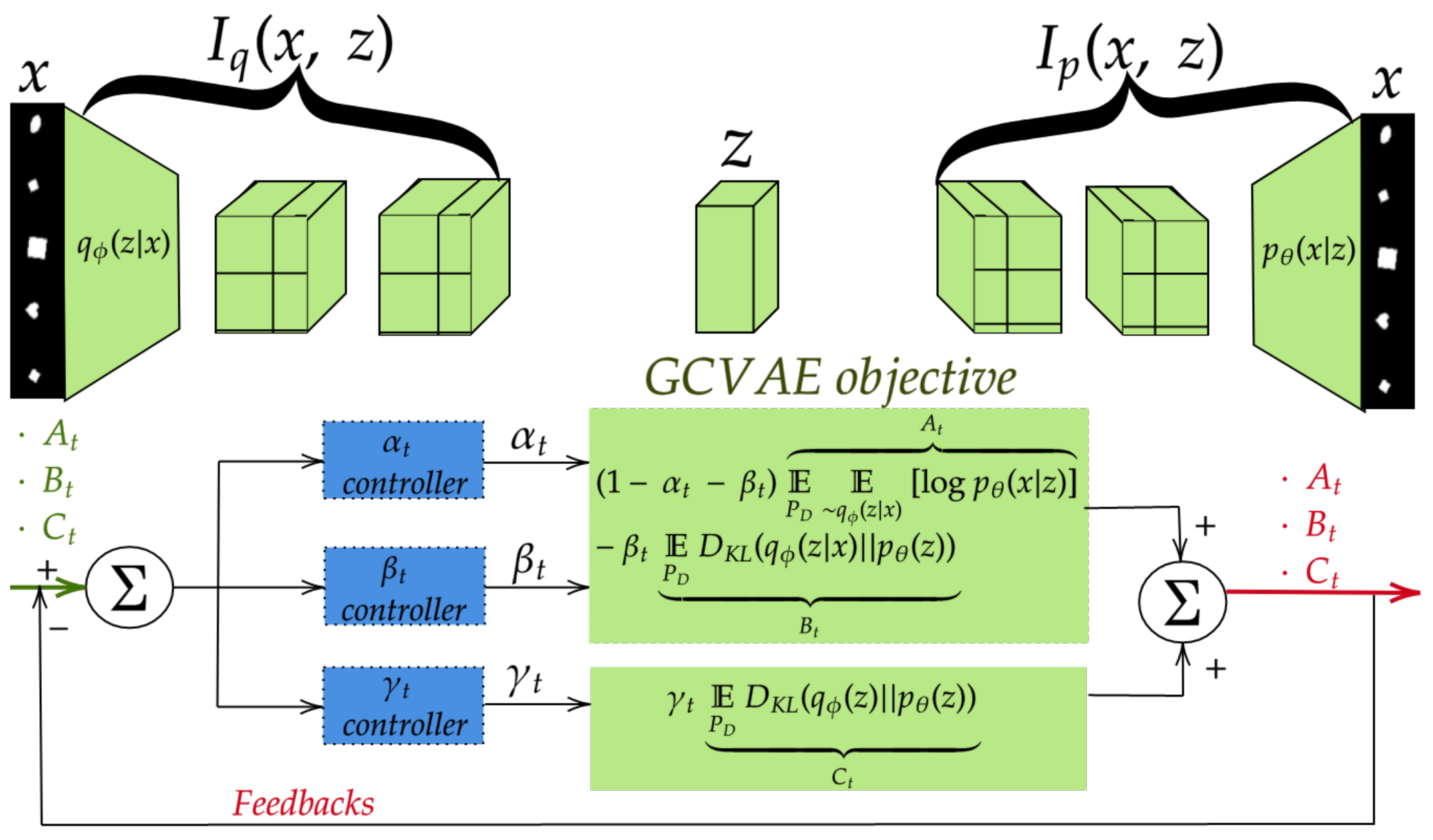
- **Context**: Addressing the challenge of disentanglement learning from an optimization perspective. The optimization framework proposed in **GCVAE** seeks to maximize a new lower bound that is based on the mutual information between reconstructed data x' and latent space z (i.e $I_p(x', z)$), subject to inference constraints and (ii) Automatic control of the hyperparameters of the loss components.
- **Objective**: Simultaneously balancing the tradeoff between reconstruction loss and Kullback–Leibler divergences. The GCVAE loss is defined as follows,

$$\mathcal{L}(\theta, \phi, \xi^+, \xi^-, \xi_p, \alpha, \beta, \gamma) = (1 - \alpha_t - \beta_t)\underset{p_\mathcal{D}}{\mathbb{E}} \underset{z\sim q_\phi(z|x)}{E}[\ln p_\theta(x|z)] - \beta_t \underset{p_\mathcal{D}}{\mathbb{E}} D_{KL}(q_\phi(z|x)||p_\theta(z)) \\ + \gamma_t \underset{p_\mathcal{D}}{\mathbb{E}} D_{KL}(q_\phi(z)||p_\theta(z)) \quad (1)$$

Where α_t, β_t and γ_t are proportional–integral–derivative (PID) controllable Lagrangian hyperparameters. GCVAE reduces into other lower bounds,

$$\mathcal{L}(\theta, \phi, \xi^+, \xi^-, \xi, \alpha, \beta, \gamma) = \begin{cases} ELBO & \text{if } \alpha_t = \alpha = -1, \beta_t = \beta = 1, \gamma = 0 \\ ControlVAE & \text{if } \alpha_t = \alpha = 0, \beta_t > 0, \gamma = 0 \\ InfoVAE & \text{if } \alpha_t = \alpha = 0, \beta_t = \beta = 0, \gamma_t > 1 \\ FactorVAE & \text{if } \alpha_t = \alpha = -1, \beta_t = \beta = 1, \gamma = -1 \end{cases} \quad (2)$$

Note that the ELBO is used interchangeably to refer to VAE loss function.


x
$I_q(x, z)$
$q_\phi(z|x)$
z
$I_p(x, z)$
x
$p_\theta(x|z)$
GCVAE objective
· A_t
· B_t
· C_t
+
−
Σ
α_t controller
β_t controller
γ_t controller
α_t
β_t
γ_t
A_t
$(1 - \alpha_t - \beta_t) \underset{P_D}{\mathbb{E}} \underset{\sim q_\phi(z|x)}{\mathbb{E}} [\log p_\theta(x|z)]$
$- \beta_t \underset{P_D}{\mathbb{E}} D_{KL}(q_\phi(z|x)||p_\theta(z))$
B_t
$\gamma_t \underset{P_D}{\mathbb{E}} D_{KL}(q_\phi(z)||p_\theta(z))$
C_t
+
+
Σ
· A_t
· B_t
· C_t
Feedbacks

GCVAE framework. α_t, β_t and γ_t respectively provide automatic balancing of the log-likelihood and KL divergences for optimal reconstruction and disentanglement. The feed-in A_t, B_t and C_t are the expectations of the variational loss.

Results

- **Evaluation metric**: **Mutual Information Gap (MIG) score** [1] reports the compactness of the latent code $(I(y_k, z_I) - I(y_k, z_{II}))/(\Sigma_{j=1}^{d} I(y_i, z_j))$. **Modularity score** [3] expresses the number of latent factors z_j with high mutual information that explains the ground-truth factors. **Joint Entropy Minus Mutual Information Gap (JEMMIG) score** [2] expreses the fact that a single latent factor may explain more than one ground-truth factor; $H(y_k, z_I) - I(y_k, z_I) + I(y_k, z_{II})$.

- **Comparison**:

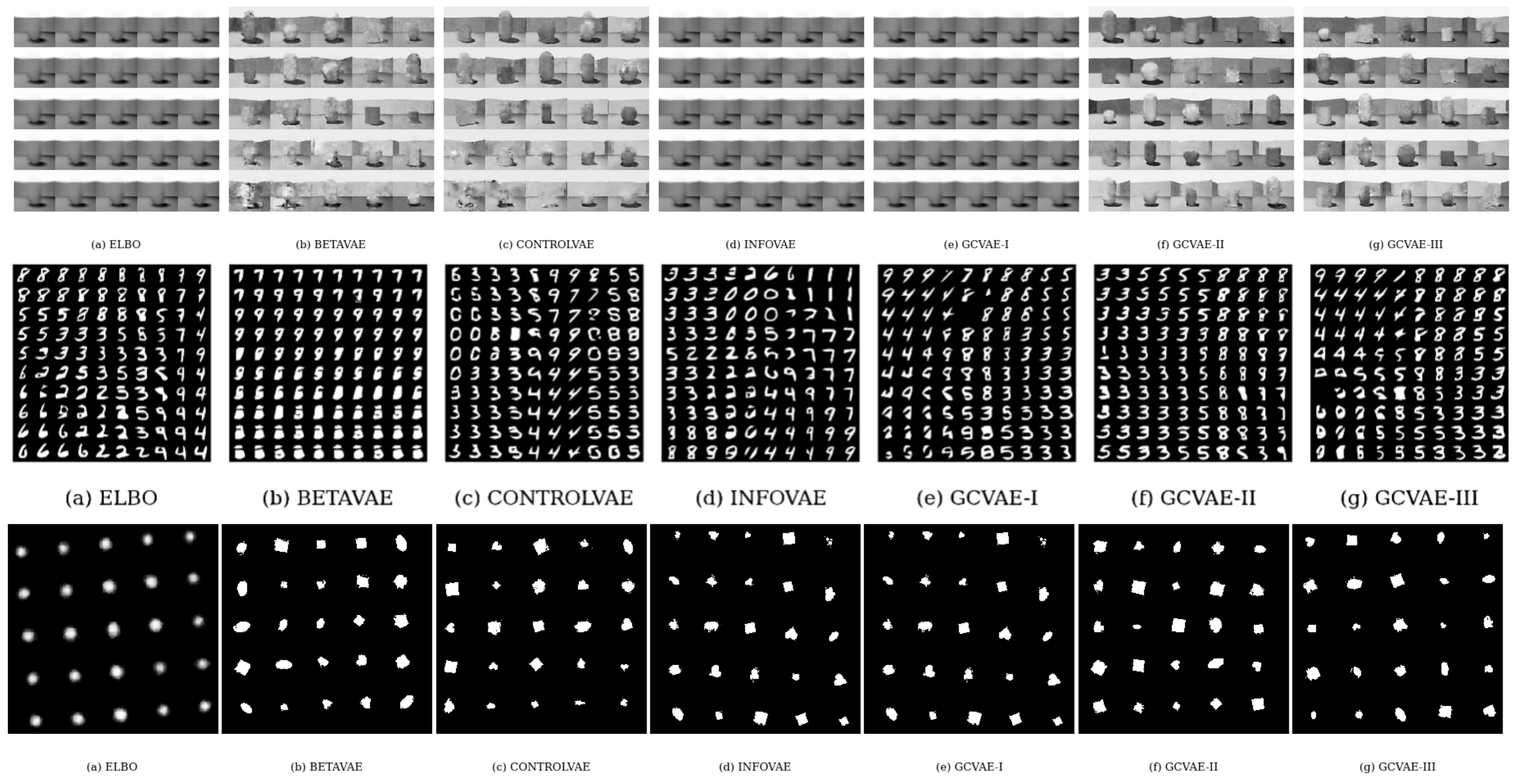
Model	Reconstruction	Disentanglement	Robustness	Interpretability
VAE	✔	✗	✗	✗
β-VAE	✗	✔	✗	✗
ControlVAE	✔	✔	✗	✔
InfoVAE (MMD)	✔	✔	✗	✔
GCVAE	✔	✔	✔	✔

Quality and drawback of the different models with emphasis on disentanglement. With high disentanglement comes poor reconstruction, and β-VAE for large values of β has the poorest quality of reconstruction. Robustness is a measure of how the trade-off between disentanglement and reconstruction are managed.

- To evaluate the strength of disentanglement and the quality of reconstruction, we propose three family of GCVAE according to the metric selected for the $D_{KL}(q_\phi(z)||p_\theta(z))$:
(1) **GCVAE-I**: $D_{KL}(q_\phi(z)||p_\theta(z)) = D^2_{MMD}(q||p)$;
(2) **GCVAE-II**: $D_{KL}(q_\phi(z)||p_\theta(z)) = D^2_{MAH}(q||p)$
(3) **GCVAE-III**: $D_{KL}(q_\phi(z)||p_\theta(z)) = \mathbb{E}\Sigma^{-1}D^2_{MMD}(q||p)$.

	MIG ↑	Modularity ↑	JEMMIG ↑	Reconstruction loss ↓	KL loss ↗
VAE	0.1268	0.798	0.233	3.339	3.0025
β-VAE	0.0778	**0.881**	0.238	**0.012**	35.0295
ControVAE	0.1213	0.782	0.312	0.016	24.3809
InfoVAE	0.1501	0.757	0.188	0.079	10.0621
GCVAE-I	0.1507	0.844	0.236	**0.012**	24.3739
GCVAE-II	**0.2793**	0.858	**0.312**	**0.012**	**24.4316**
GCVAE-III	0.1337	0.825	0.294	0.015	24.2937

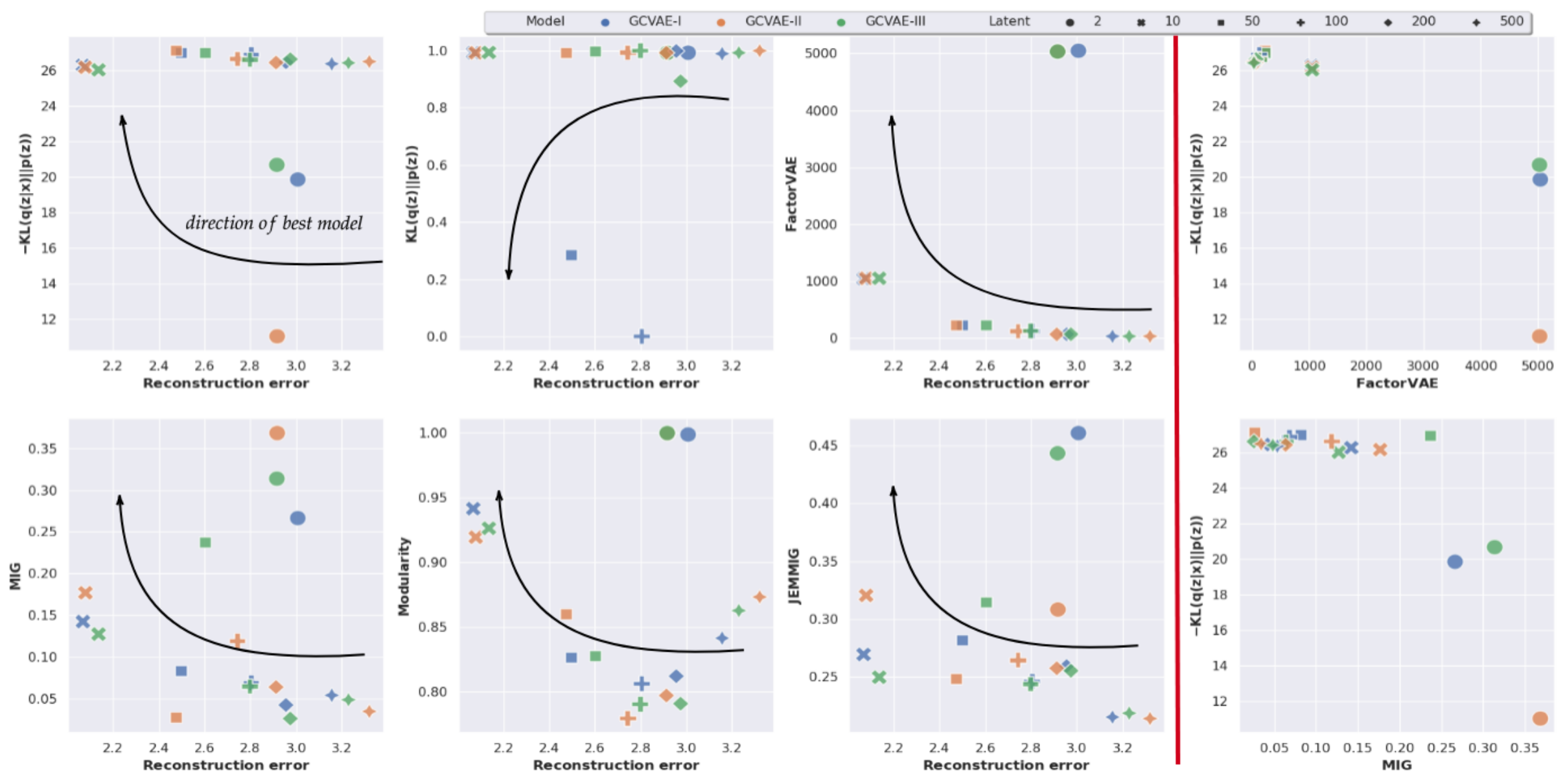
Performance comparison of different models on **DSprites** after training on 737 samples. Comparison metrics MIG [1], Modularity (MOD) [3] and JEMMIG [2] for 10-D Latent representation. The direction of the arrow indicates the best performing model. Higher is better for MIG, Modularity and JEMMIG($1 - JEMMIG$). **GCVAE-II** performs best on MIG disentanglement metric, robustness and interpretability; plus having the lowest reconstruction error. **GCVAE-I, III** and **ControlVAE** also measure up in disentanglement compared to other benchmark models. **GCVAEs** have the least reconstruction error partly due to the normalization introduced by the inverse precision matrix in $D_{KL}(q_\phi(z))||p_\theta(z))$ and the weight on the first term of the GCVAE loss. **β-VAE** (with $\beta = 1e^{-4}$) has a comparable reconstruction loss with GCVAE-II and the best Modularity.


(a) ELBO (b) BETAVAE (c) CONTROLVAE (d) INFOVAE (e) GCVAE-I (f) GCVAE-II (g) GCVAE-III

(a) ELBO (b) BETAVAE (c) CONTROLVAE (d) INFOVAE (e) GCVAE-I (f) GCVAE-II (g) GCVAE-III

(a) ELBO (b) BETAVAE (c) CONTROLVAE (d) INFOVAE (e) GCVAE-I (f) GCVAE-II (g) GCVAE-III

Generative process $p_\theta(x|z)p(z)$ comparison for the different models after training on **3D shapes** (top), **MNIST** (middle), **DSprites** (bottom) dataset for $250K$, 500 and $250K$ epochs respectively. GCVAE-II and ELBO (VAE) have similar reconstruction quality with better interpretation. GCVAE-II clearly outperformed the benchmark models in generating meaningful, diverse and clear representations of the original data. β-VAE is the least performing at generating interpretable image of the original data.



Model performance comparison on 737 samples of **DSprites** data. Arrows indicate direction of best performing model. **Top**: Comparison of reconstruction error against KL divergence , $D_{KL}(q_\phi(z|x))||p_\theta(z))$ and correlation, $D_{KL}(q_\phi(z))||p_\theta(z))$. High KL-Low reconstruction error observed for Latent-10. Low reconstruction error does and high KL does not imply high disentanglement. High disentanglement using FactorVAE is observed on Latent-2 followed by Latent 10. **Bottom**: Comparing disentanglement metrics with reconstruction loss. Highest disentanglement on MIG metric observed for GCVAE-**II** on Latent-2, however, Best scores is observed for GCVAE-II on Latent-10. JEMMIG is similar in behaviour with MIG. **RHS**: Validating the statement *Latent disentanglement is not correlated with KL maximization*.

Conclusion

We propose a new lower bound we refer to as, Generalized-Controllable Variational Autoencoder (GCVAE). A model built from an constraint optimization perspective to maximize mutual information in the generative phase subject to inference constrains to encourage disentanglement in the latent space. We use the Mahalanobis distance metric as a heuristic to encourage disentanglement in the variables of the latent space and show that the representation obtained GCVAE is both meaningful and interpretable, with low reconstruction loss. GCVAE-II shows extensive strength in disentangling the latent space and reconstructing with minimal mutual information loss compared to other variants. Next objective is using the model together with transformer model for decision (text) generation.

References

[1] Ricky T. Q. Chen, Xuechen Li, Roger Grosse, and David Duvenaud. Isolating sources of disentanglement in variational autoencoders, 2019.

[2] Kien Do and Truyen Tran. Theory and evaluation metrics for learning disentangled representations, 2021.

[3] Karl Ridgeway and Michael C. Mozer. Learning deep disentangled embeddings with the f-statistic loss, 2018.