Multi-objects Generation with Amortized Structural Regularization

Kun Xu, Chongxuan Li, Jun Zhu, Bo Zhang

Department of Computer Science, Tsinghua University

Introduction

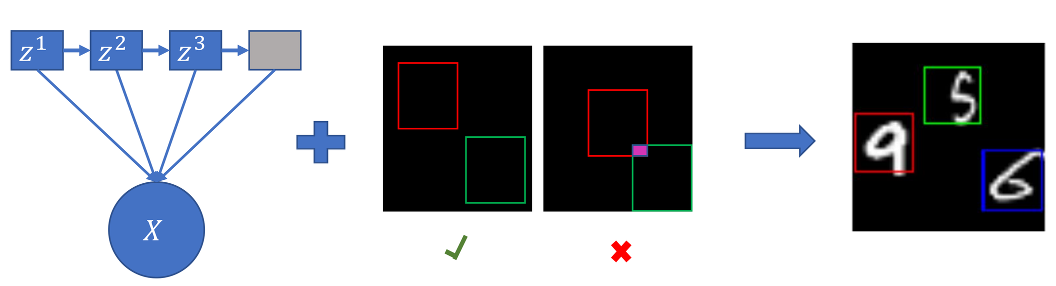


Figure: How to embed human knowledge into structured DGMs?

Embedding human knowledge into DGMS:

- The most straightforward way: Designing proper prior distribution.
  - Sometimes Infeasible, such as the non-overlapping case.
  - More challenging to optimization.
- Posterior Regularization:
  - Flexible and general.
  - Easy to optimize.
  - Hard to generalize to DGMs.

ASR: amortized PR for structural regularization

ASR extend PR to DGMs and can:

- Efficiently amortized variational inference with interpretable latent variables.
- Embed human knowledge into DGMs through structural constraints.

Formally, ASR extend PR into the amortized version as follows:

$$\max_{\theta} \log \int_Z p(X,Z;\theta)dZ - \Omega(p(Z|X;\theta))$$

$$\rightarrow \max_{\theta\in\Theta,\phi\in\Phi} \mathbb{E}_{q(Z|X;\phi)} \log \frac{p(X,Z;\theta)}{q(Z|X;\phi)} - R(q(Z|X;\phi)),$$

where $\Omega = \min_{q\in Q} KL(q(Z)||p(Z|X;\theta))$ and Q is defined as $Q = \{q(Z)|R(q(Z)) \leq \mathbf{0}\}$.

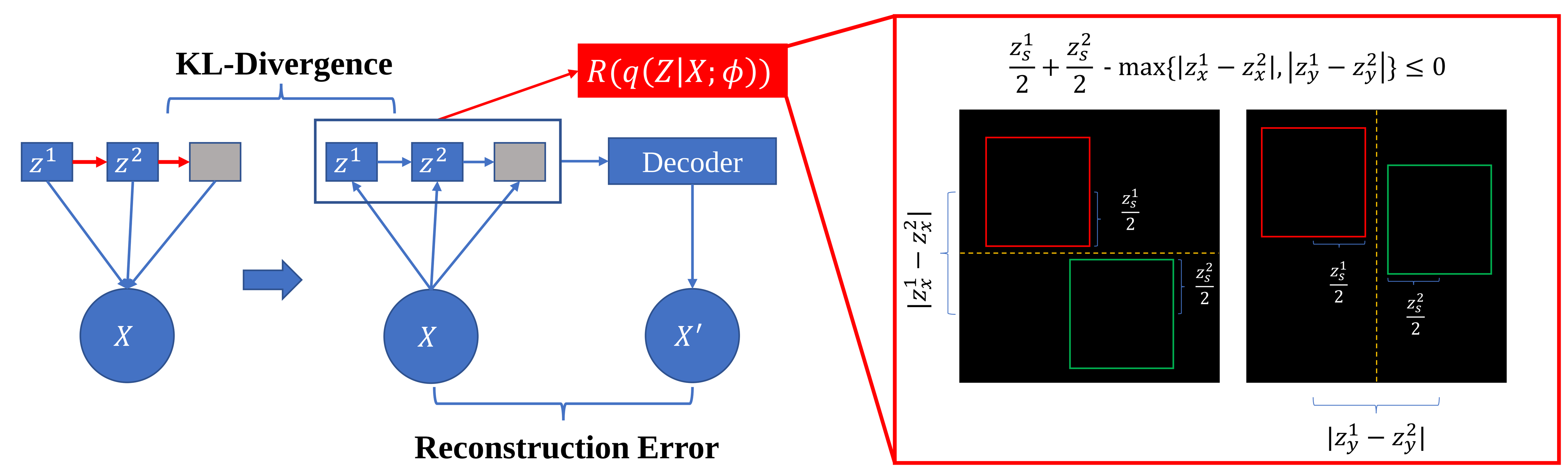


Figure: Example of non-overlapping regularization.

Structured generative models and its inference

- Generative model is defined as:

$$p(x,z,n;\theta) = p(z_{pres}^{n+1} = 0|z^{\leq n};\theta)\left(\prod_{t=1}^{n} p(z_{pres}^{t} = 1|z^{<t};\theta)p(z^t|z^{<t};\theta)\right)p(x|z,z_{pres};\theta).$$

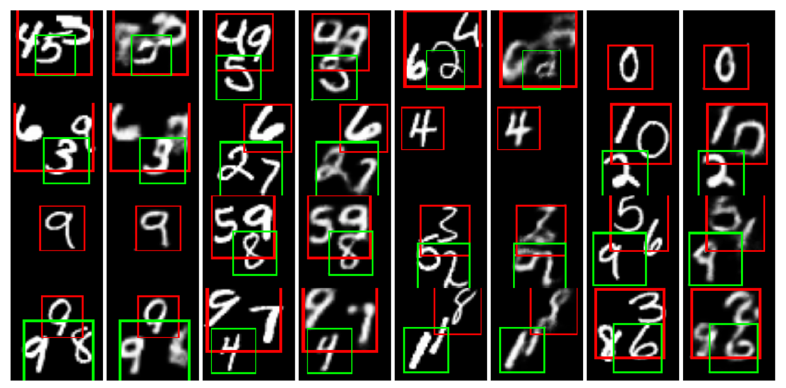
- Inference model is defined as follows:

$$q(z,n|x;\phi) = q(z_{pres}^{n+1} = 0|z^{\leq n},x;\phi)\prod_{t=1}^{n} q(z_{pres}^{i} = 1|z^{<t},x;\phi)q(z^t|z^{<t},x;\phi).$$

ASR on the number of objects

Each image contains a certain number of objects. The valid set is defined as:

$$Q^{num} = \{q(Z|X)|\min_{q_i\in V} KL(q_i||q(Z|X)) \leq 0, KL(q_{uni}(Z)||\mathbb{E}_{p(X)}q(Z|X)). \leq 0\}$$



(a) The reconstruction of AIR-13.

(b) The reconstruction of AIR-pPrior-13.

(c) The reconstruction of AIR-ASR-13.

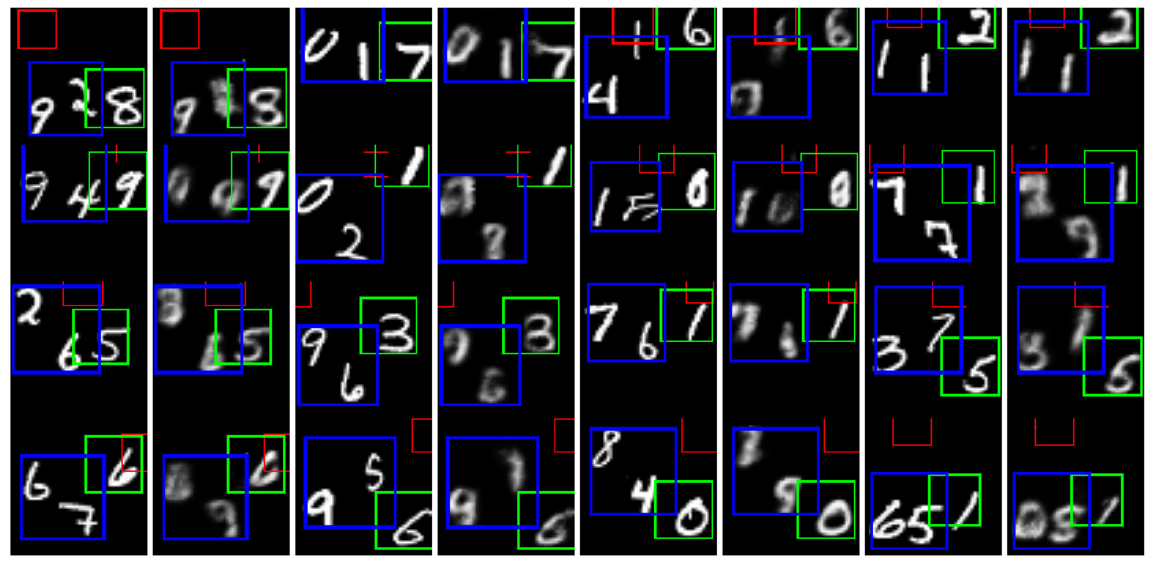
Figure: The reconstruction results of Multi-MNIST on 1 or 3 objects.

Methods	nELBO	ACC	SE	mIoU
AIR-13	404.41 ± 4.58	0.81 ± 0.23	31.94 ± 4.68	$\mathbf{0.61} \pm 0.13$
AIR-pPrior-13	405.21 ± 1.17	0.48 ± 0.00	49.42 ± 0.24	0.43 ± 0.01
AIR-ASR-13	$\mathbf{360.20} \pm 19.67$	$\mathbf{0.96} \pm 0.00$	$\mathbf{28.84} \pm 1.11$	$\mathbf{0.61} \pm 0.00$

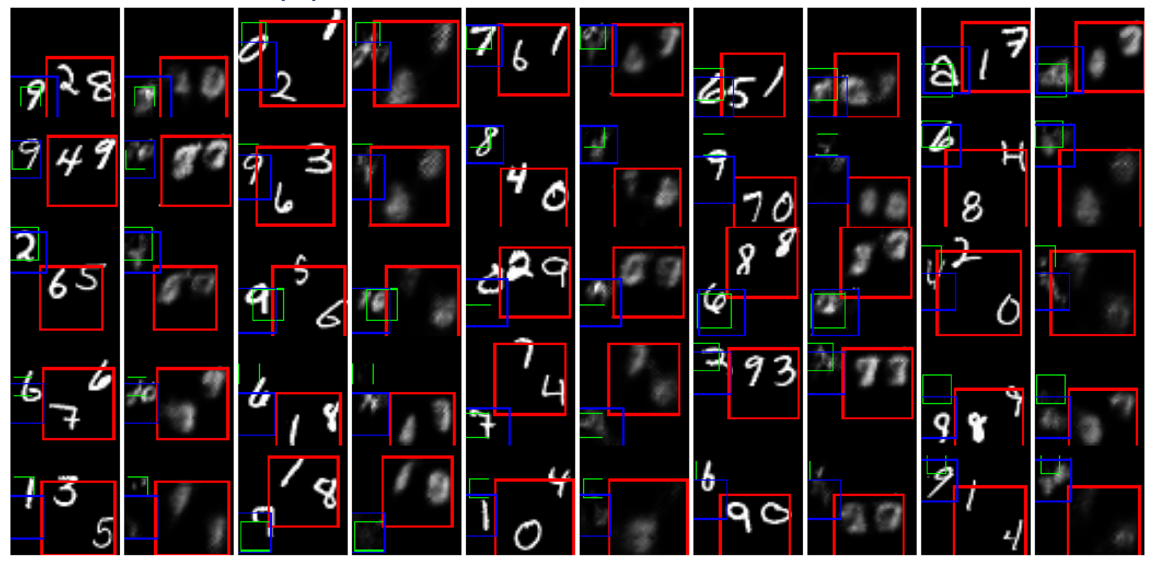
Table: Results on regularization on the number of objects.

ASR on non-overlapping

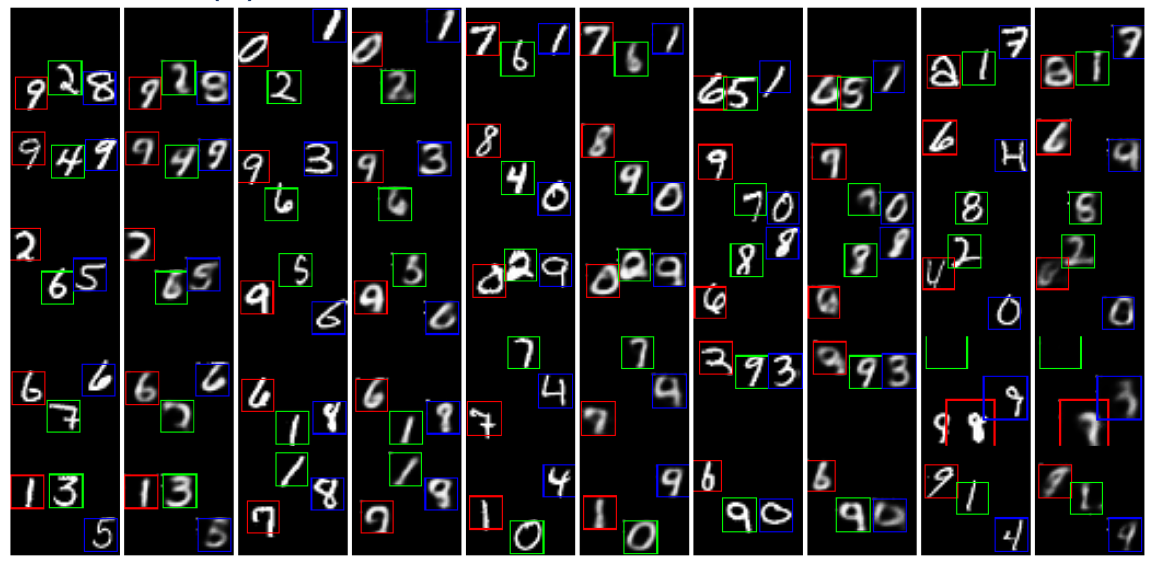
The objects are non-overlapped within an image.



(a) The reconstruction of AIR-3.



(b) The reconstruction of AIR-pPrior-3.



(c) The reconstruction of AIR-ASR-3.

Figure: The reconstruction results of Multi-MNIST on 3 objects. There is no overlapping among objects in the training data. ASR can successfully infer the underlying structures, and improve the reconstruction results.

Contact Information

- Source Code: https://github.com/taufikxu/MOG-ASR
- Email: kunxu.thu@gmail.com