UCLA Samueli School of Engineering

Towards Composable Latent Augmentation in Generative Models

Sunay Bhat, Jeffrey Jiang,
Omead Pooladzandi and Greg Pottie

UCLA ECE Department

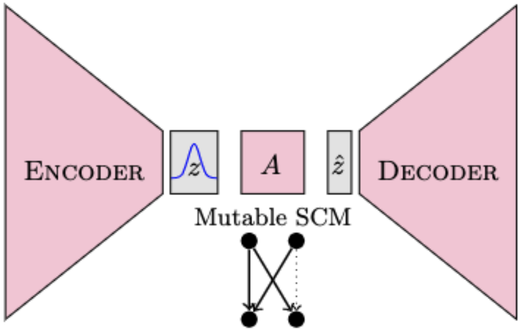
May 2022

Contents

Causal Generative Modeling

Previous work: Causal Counterfactual Generative Models

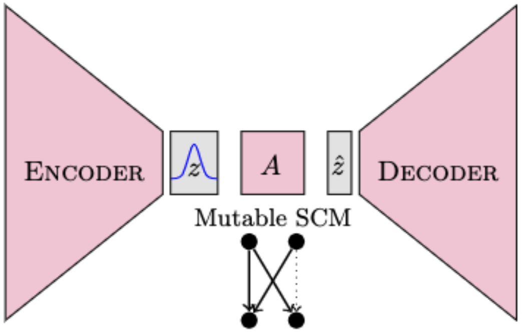
- Used Causal VAE architecture compressing labelled images to a latent space[1,2]
- Learned adjacency matrix and functional relationships with only endogenous/exogenous (parent/child) priors
- Demonstrated ability to modify model and produce Out-of-Distribution counterfactual results on physics based dataset

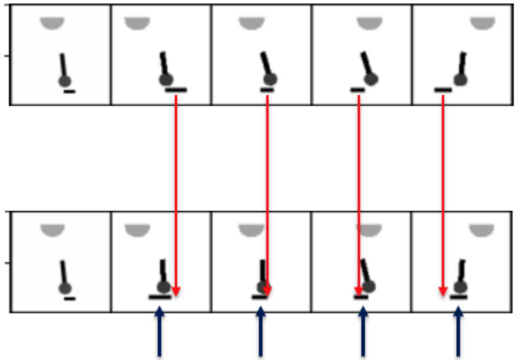


Causal Generative Modeling

Previous work: Causal Counterfactual Generative Models

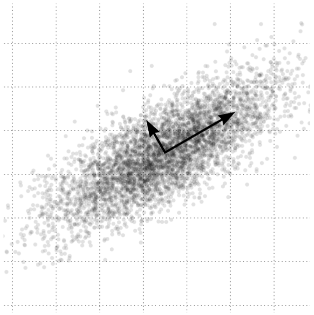
- Used Causal VAE architecture compressing labelled images to a latent space[1,2]
- Learned adjacency matrix and functional relationships with only endogenous/exogenous (parent/child) priors
- Demonstrated ability to modify model and produce Out-of-Distribution counterfactual results on physics based dataset

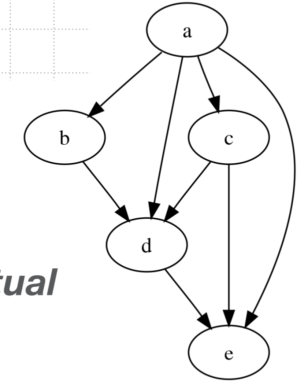




Causal Disentanglement Challenge

- **Causal Concepts are in a low dimensional space while images are relatively high-dimensional**
  - Compression is focused on ***minimum distortion, not interpretability***
    - orthogonal, variance preservation akin to PCA
- **Full causal disentanglement was not possible:**
  - Latent variables were not perfectly-aligned
  - Decoder learned some functional relationships
- **Fundamental Tension of Compression and Causality**
  - Subset of causal variables are usually highly correlated





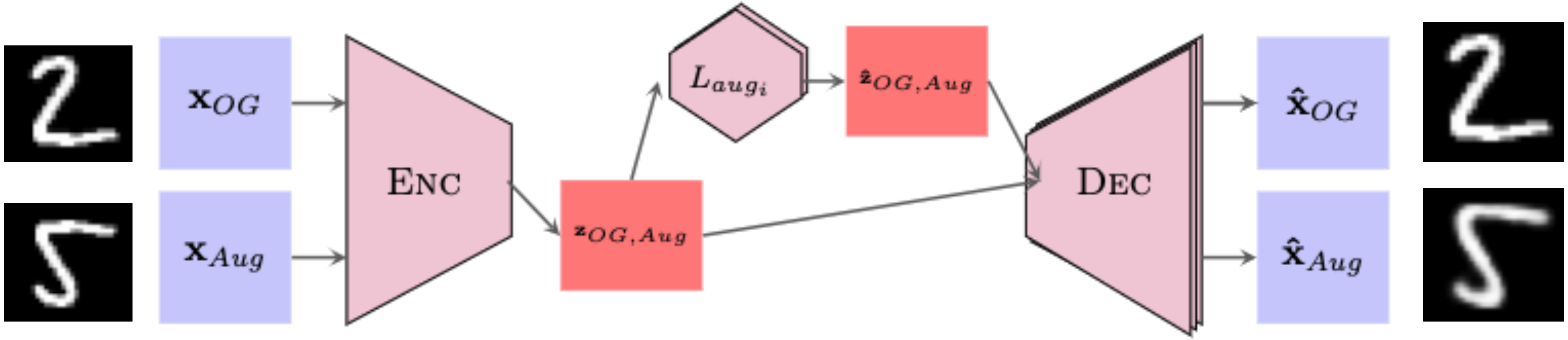
A pure information bottleneck is not the right solution for conceptual variables that are not independent (most of our causal priors)

Priors in Pre-processing

- **Data augmentation is a powerful technique to expand the train/test domain**
- **We don't augment naively, we augment based on priors**
  - We modify images along dimensions invariant to the classification interpretation
- **How might this perspective impact latent space research**
  - Can we better interpret and control the latent space?
  - Can we learn mappings between areas of the latent space?
  - Can we transfer certain properties of latent spaces between augmentations?

Latent Augmentation VAE (LAVAE)

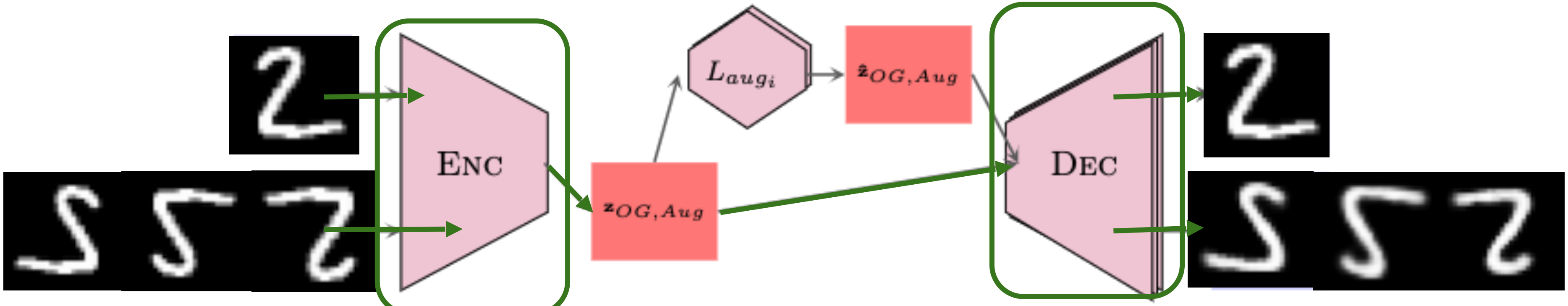
- **Similar architecture to VAE, but**
  - Includes learnable linear transformations between augmented latent spaces
  - Multiple decoder network heads for transferring latent spaces to new augmentations



Flip Up/Down

Training LAVAE

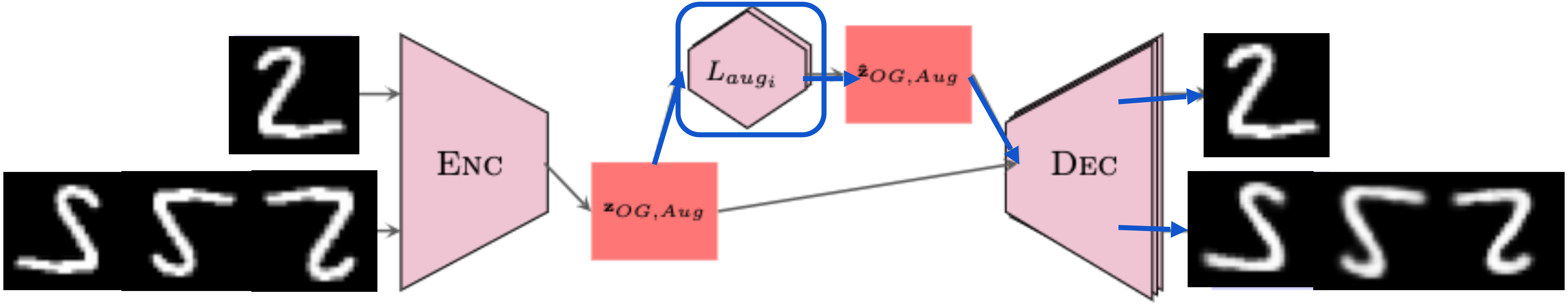
- **Training with Flip Left/Right and Flip Up/Down**
  - **Stage 1:** Populate latent space with original, two types of augmentations, and their composition
  - **Stage 2:** Learn L_{aug} linear transformations between latent spaces (original ->Aug_1, original -> Aug_2)
  - **Stage 3:** Transfer trained latent space by training new decoder on any other pair of augmentations *(X-direction shear and Canny edge-detect augmentations pictured)*



Flip Left/Right, Flip Up/Down

Training LAVAE

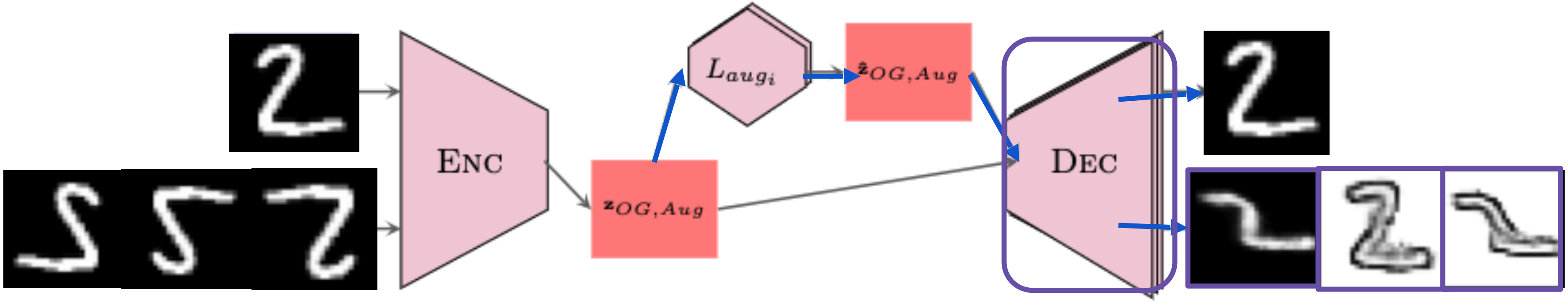
- **Training with Flip Left/Right and Flip Up/Down**
  - **Stage 1:** Populate latent space with original, two types of augmentations, and their composition
  - **Stage 2:** Learn L_{aug} linear transformations between latent spaces (original ->Aug_1, original -> Aug_2)
  - **Stage 3:** Transfer trained latent space by training new decoder on any other pair of augmentations *(X-direction shear and Canny edge-detect augmentations pictured)*



Flip Left/Right, Flip Up/Down

Training LAVAE

- **Training with Flip Left/Right and Flip Up/Down**
  - **Stage 1:** Populate latent space with original, two types of augmentations, and their composition
  - **Stage 2:** Learn L_{aug} linear transformations between latent spaces (original ->Aug_1, original -> Aug_2)
  - **Stage 3:** Transfer trained latent space by training new decoder on any other pair of augmentations *(X-direction shear and Canny edge-detect augmentations pictured)*



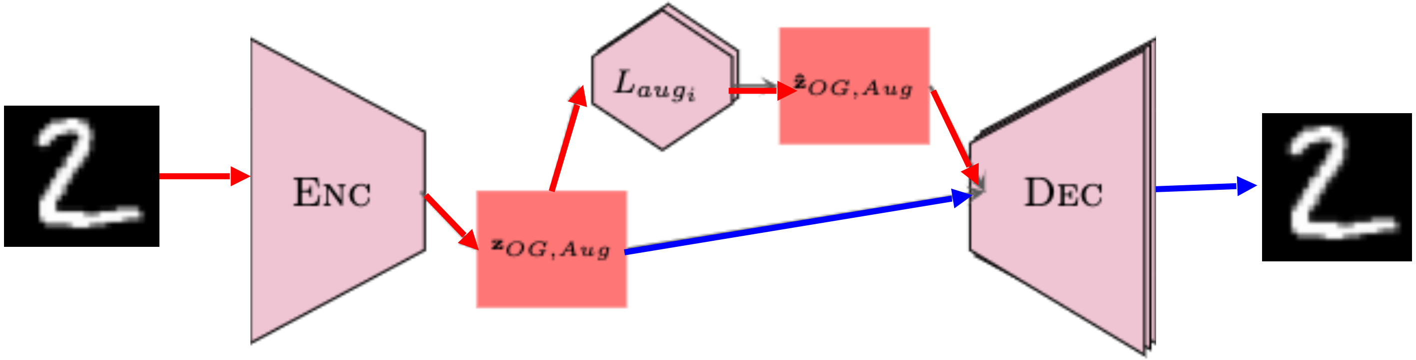
X-dir Shear, Canny Edge-Detect

LAVAE Capabilities

- Once trained, LAVAE offers tremendous flexibility for latent augmentation

Basic Reconstruction:

$$\hat{x}_i = DEC(ENC(x_i)) \mid i \in \{Orig, 1, 2, Compose\}$$

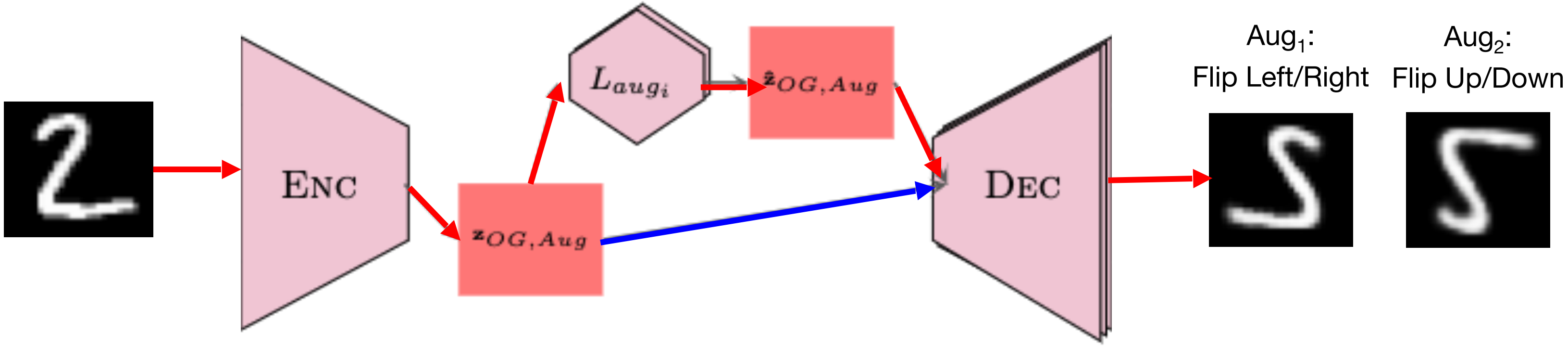


LAVAE Capabilities

- Once trained, LAVAE offers tremendous flexibility for latent augmentation

Latent Augmented Reconstructions:

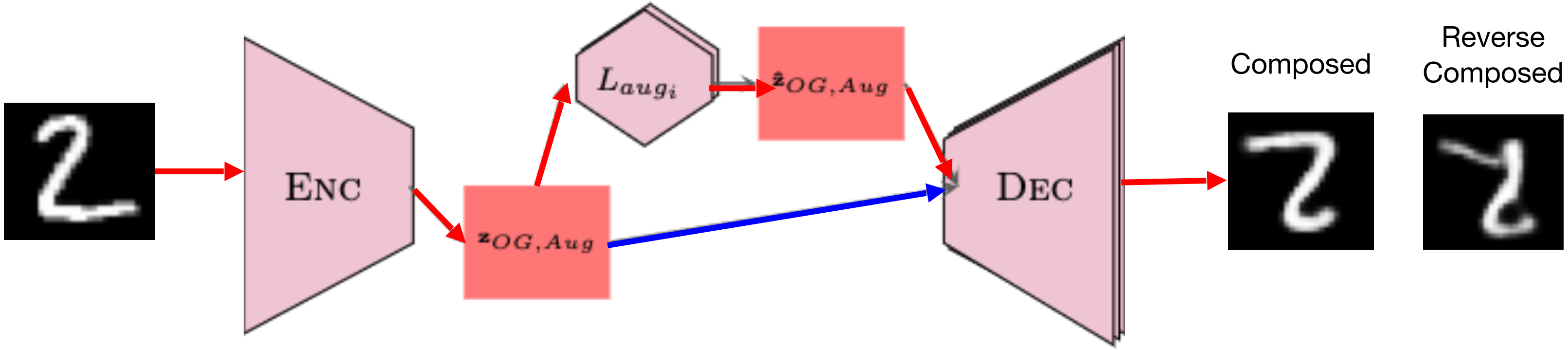
$$\hat{z}_i = ENC(x_{Orig}) \cdot L_{aug_i}$$
$$\hat{x}_i = DEC(\hat{z}_i) | i \in \{1, 2\}$$



LAVAE Capabilities

- **Once trained, LAVAE offers tremendous flexibility for latent augmentation Latent Composable Augmentations (*without explicit training*), and reverse composition with some loss:**

$$\hat{z_{compose}} = ENC(x_i) \cdot L_{aug_1} \cdot L_{aug_2}$$
$$\hat{z_{r\_compose}} = ENC(x_i) \cdot L_{aug_2} \cdot L_{aug_1}$$
$$\hat{x_{compose}} = DEC(\hat{z_{compose}}) \text{ and } \hat{x_{compose}} \approx DEC(\hat{z_{r\_compose}})$$

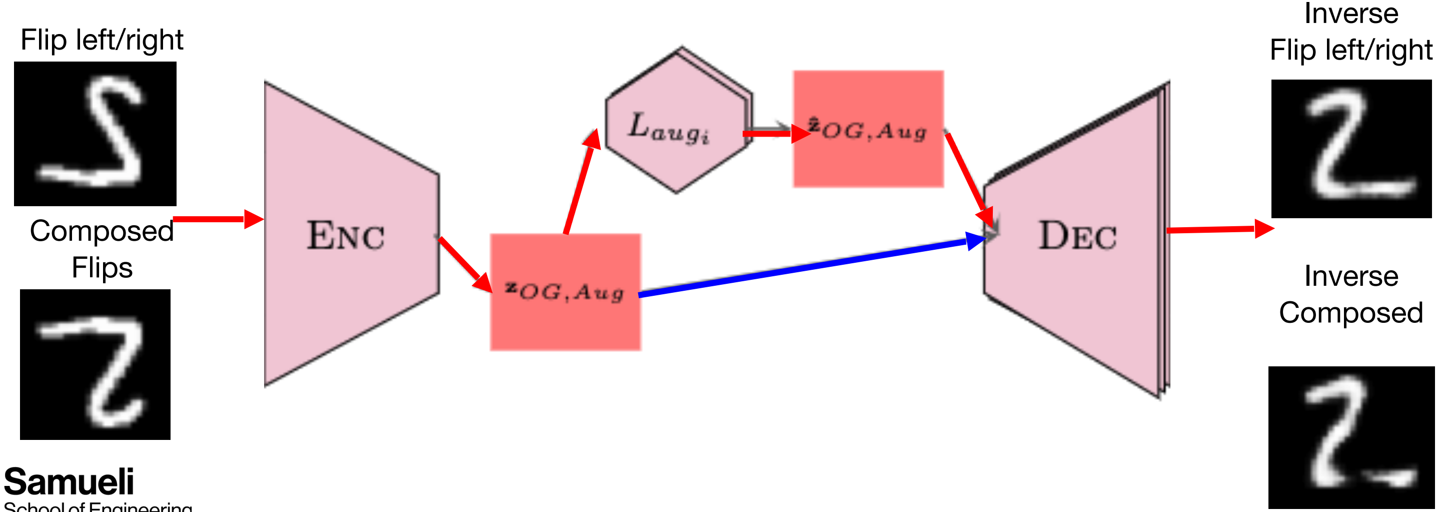


LAVAE Capabilities

- **Once trained, LAVAE offers tremendous flexibility for latent augmentation Inverse Augmentations to go from any input to any output**

$$\hat{x}_i = DEC(ENC(x_i) \cdot L_{aug_i}^{-1}) | i \in \{1, 2\}$$

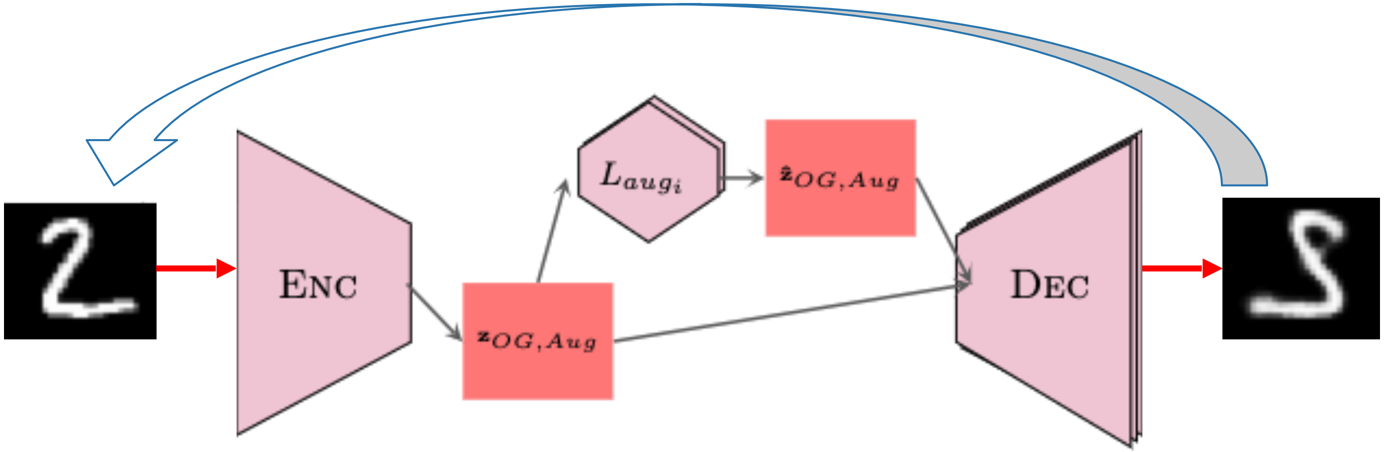
$$\hat{x_{Orig}} = DEC(ENC(x_{compose}) \cdot (L_{aug_1} \cdot L_{aug_2})^{-1})$$



LAVAE Capabilities cont.

- **We can also utilize LAVAE recursively**
  - Sometimes losses accumulate and repeated augmentations diverge and other times it is stable

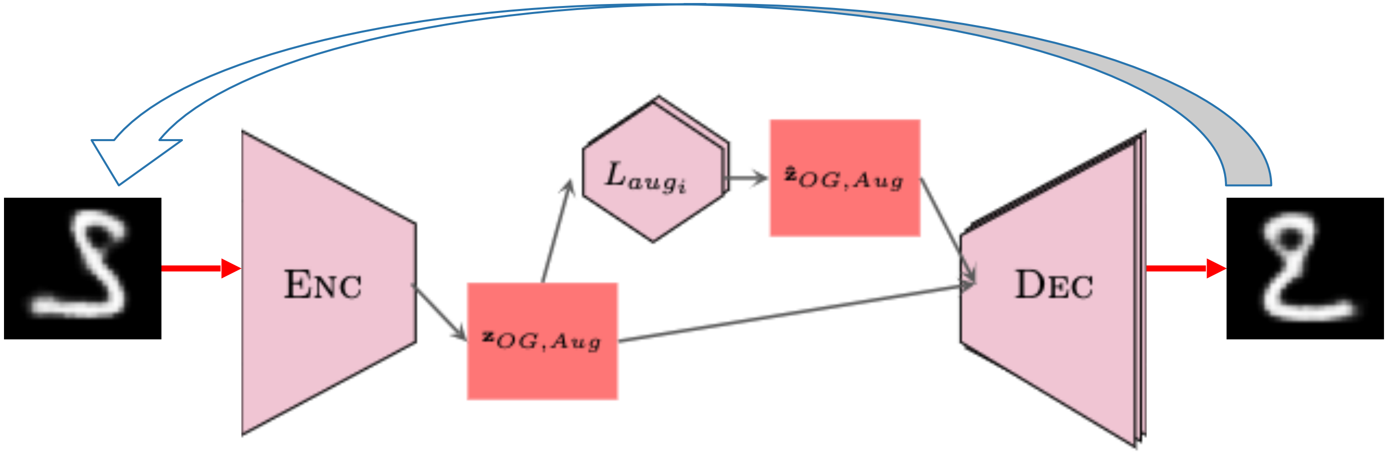
$$\hat{x}_i = DEC(ENC(x_{Orig}) \cdot L_{aug_i})$$
$$for\ j\ in\ range:$$
$$\hat{\hat{x}}_i = DEC(ENC(\hat{x}_i) \cdot L_{aug_i})$$



LAVAE Capabilities cont.

- **We can also utilize LAVAE recursively**
  - Sometimes losses accumulate and repeated augmentations diverge and other times it is stable

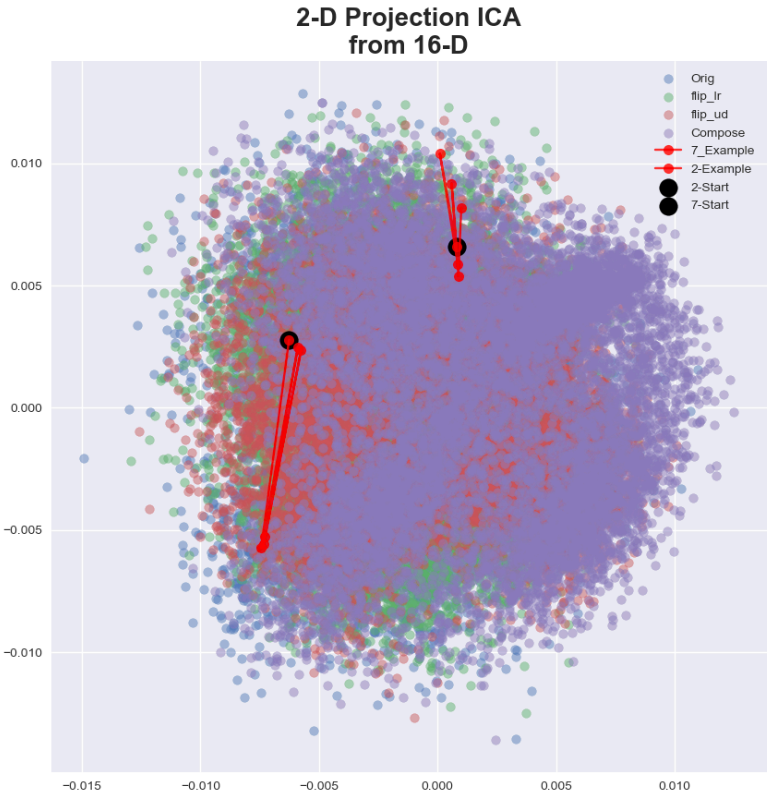
$$\hat{x}_i = DEC(ENC(x_{Orig}) \cdot L_{aug_i})$$
$$for\ j\ in\ range:$$
$$\hat{\hat{x}}_i = DEC(ENC(\hat{x}_i) \cdot L_{aug_i})$$

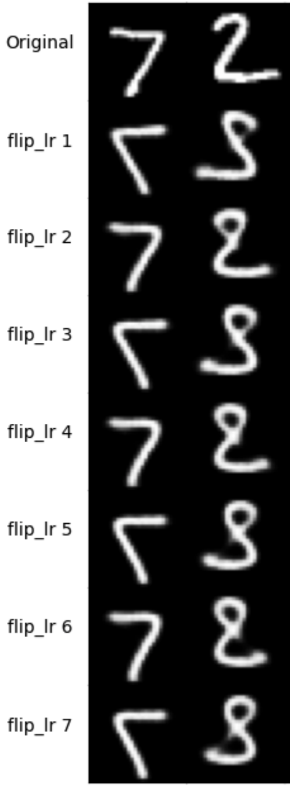


LAVAE Capabilities cont.

Repeated Flip left/right

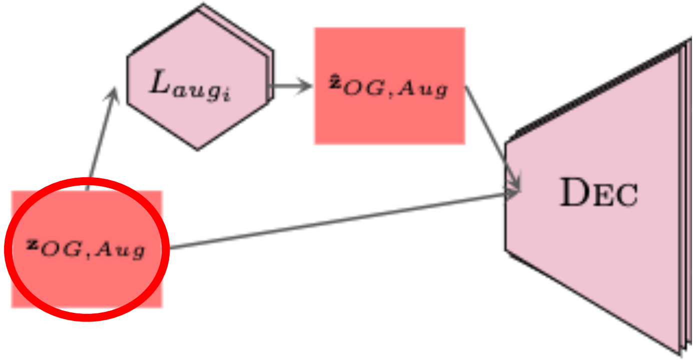
- 2 diverges (possibly to 8's)
- 7 is relatively stable showing minimal loss with repeated flip left/right 's

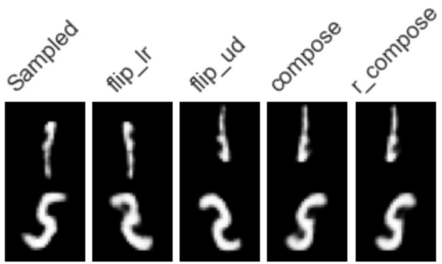




LAVAE Interpolation and Sampling

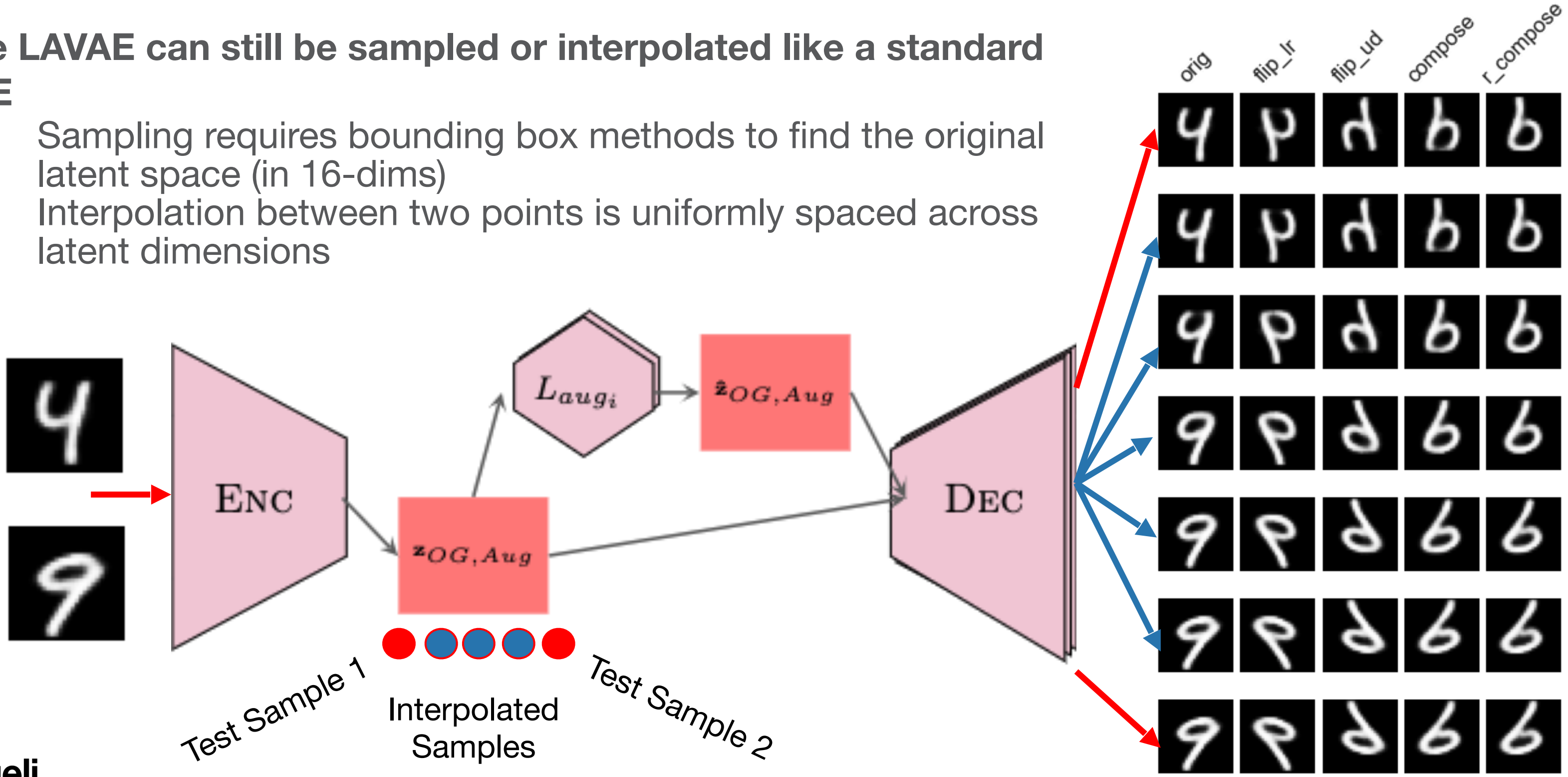
- **The LAVAE can still be sampled or interpolated like a standard VAE**
  - Sampling requires bounding box methods to find the original latent space (in 16-dims)
  - Interpolation between two points is uniformly spaced across latent dimensions





LAVAE Interpolation and Sampling

- **The LAVAE can still be sampled or interpolated like a standard VAE**
  - Sampling requires bounding box methods to find the original latent space (in 16-dims)
  - Interpolation between two points is uniformly spaced across latent dimensions

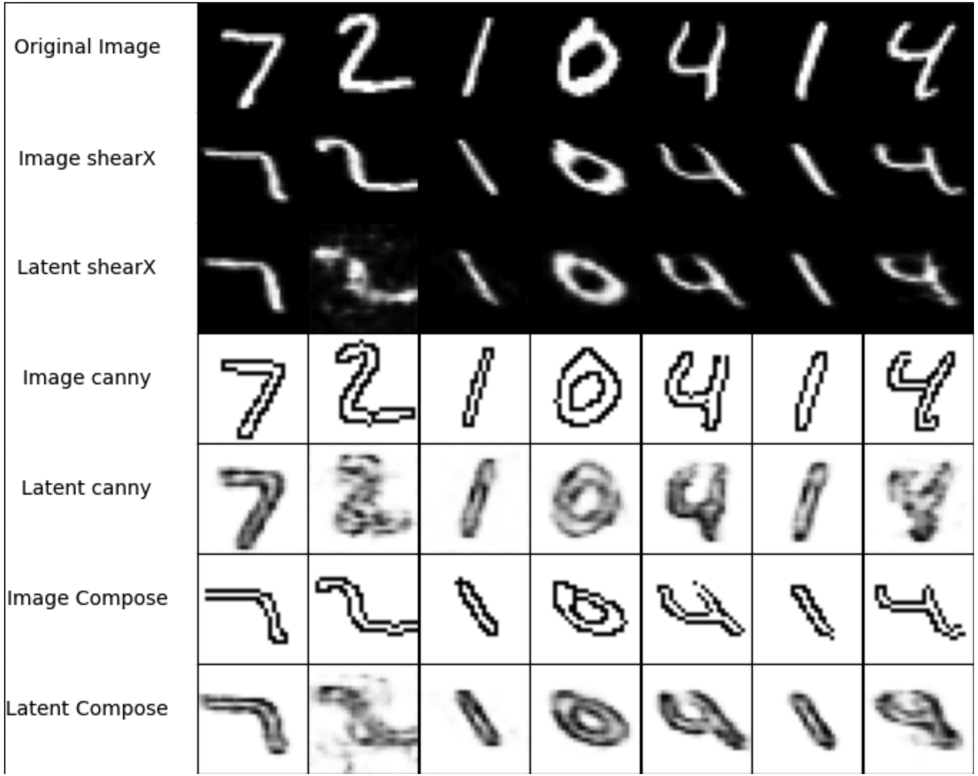


LAVAE Augmentation Transfer

- Finally, with LAVAE, we can transfer our latent space to any pair of augmentations:

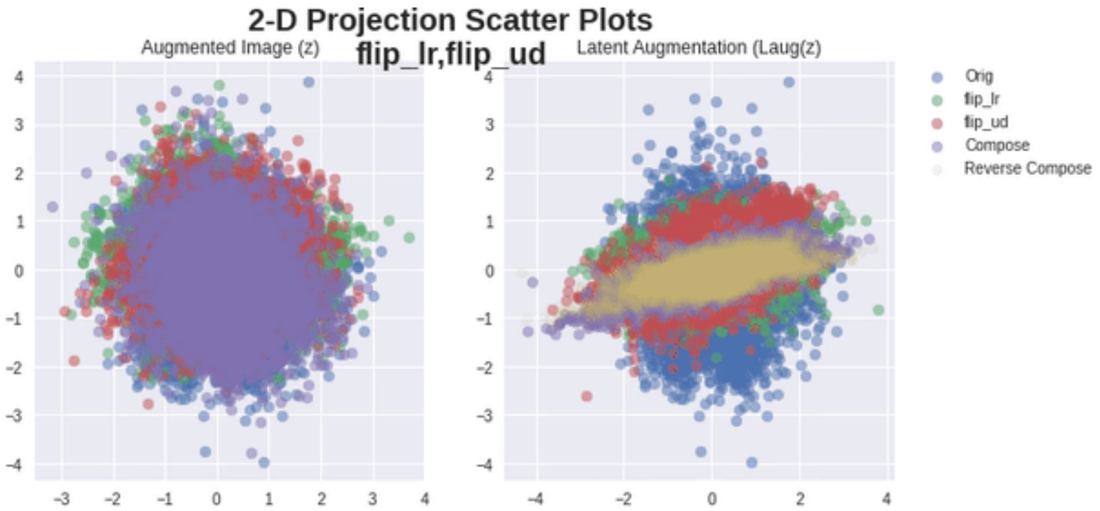
Flip left/right, flip up/down

X-direction Shear, Canny edge-detection



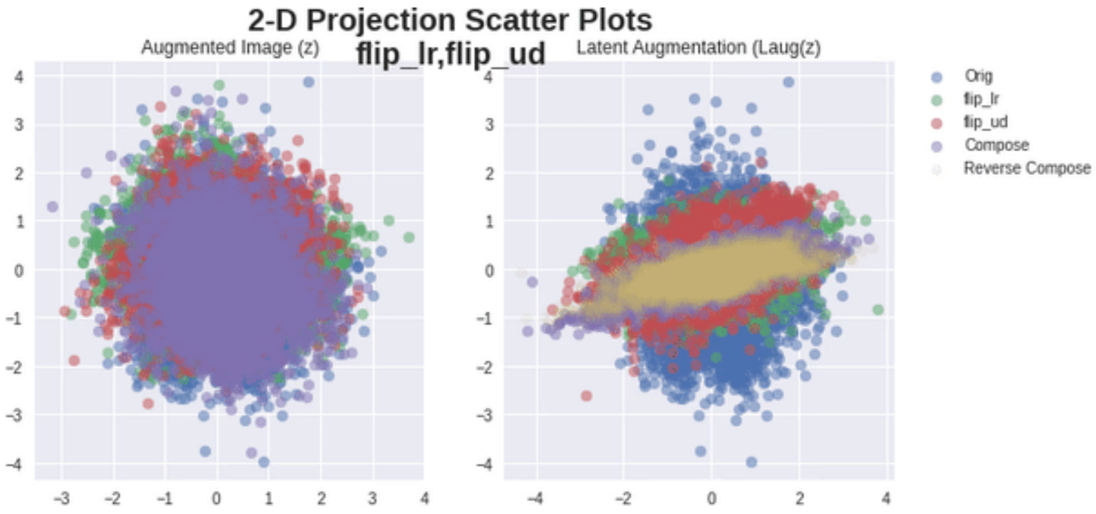
Exploring Latent Geometries

- The **choice of initial augmentation** impacts the latent space properties and how well they transfer to new augmentations
- We explore many augmentations pairs, using 2-D latent dimension or compressions (PCA, tSNE, ICA) to visualize the latent space and understand the geometries



Exploring Latent Geometries

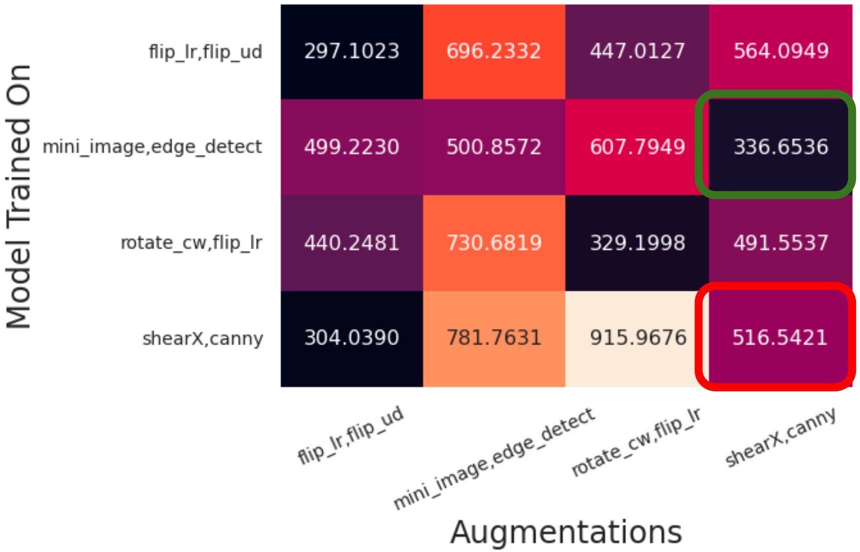
- The **choice of initial augmentation** impacts the latent space properties and how well they transfer to new augmentations
- We explore many augmentations pairs, using 2-D latent dimension or compressions (PCA, tSNE, ICA) to visualize the latent space and understand the geometries

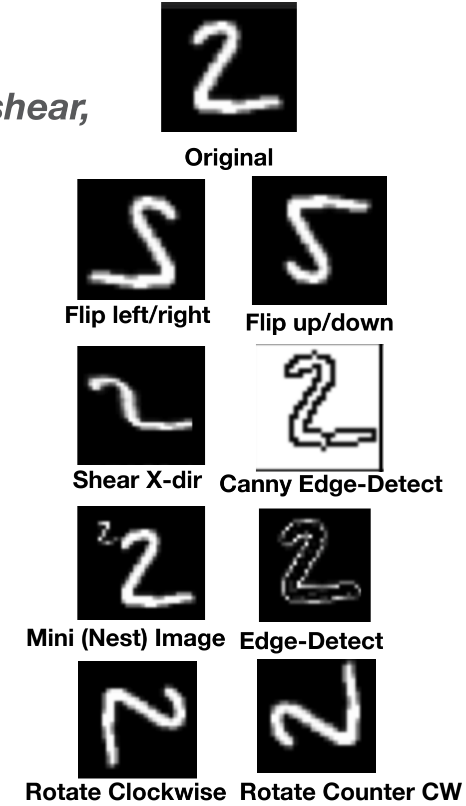


Tradeoff in Augmentation Pairs

The choice of initial augmentation impacts transfer performance:

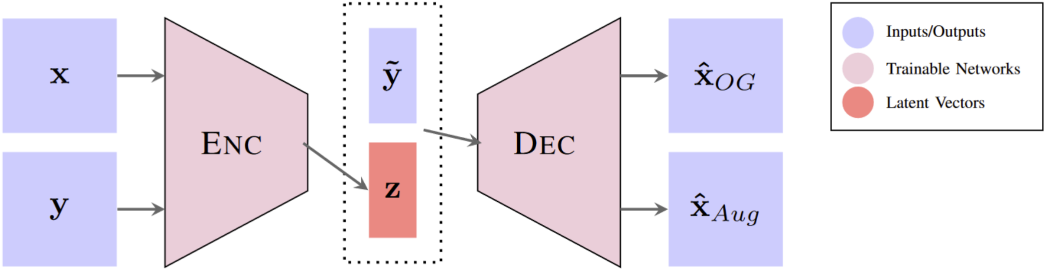
- Mini-Image, edge-detect pair ***performs better transferring to a shear, canny combo*** than training natively on shear,canny





LAVAE vs Conditional VAE (CVAE)

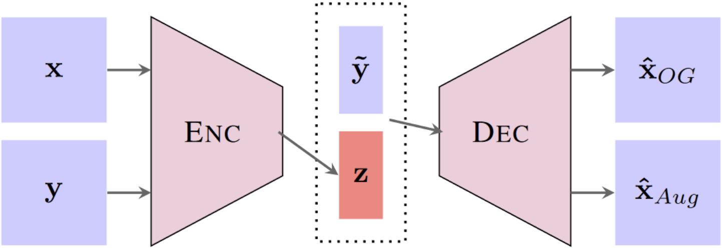
- **CVAE typically conditions on class to generate images within a class**
  - Idea: Enforce latent space partitioning done with conditioning increasing latent dimensions.
  - Can be modified to condition on class, augmentation or both
- **However, CVAE operates on a different philosophy than the LAVAE**
  - Uses same learned distribution, but no learned dependence between classes.
  - No latent space flexibility, meaning no augmentation interpolation.
- **CVAE either**
  - Does not preserve uniqueness (distributions are ind. on conditional)
  - Requires explicit training similar to LAVAE process

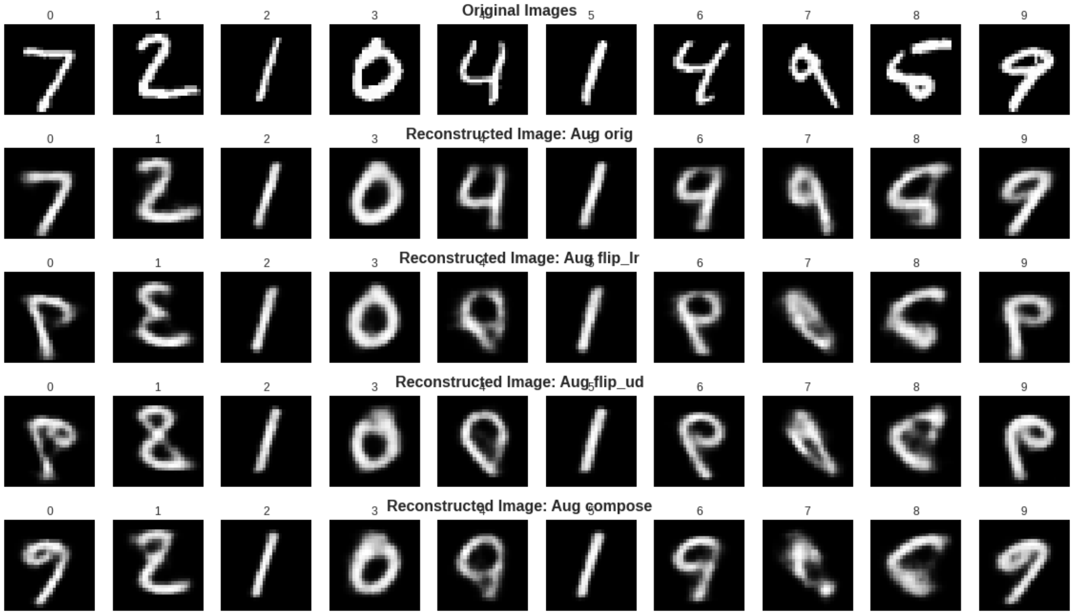


CVAE and Augmentations

CVAE: Class Change

- Suppose $\mathbf{y} \neq \tilde{\mathbf{y}}$
- Conditional does not change output based on augmentations of the original image

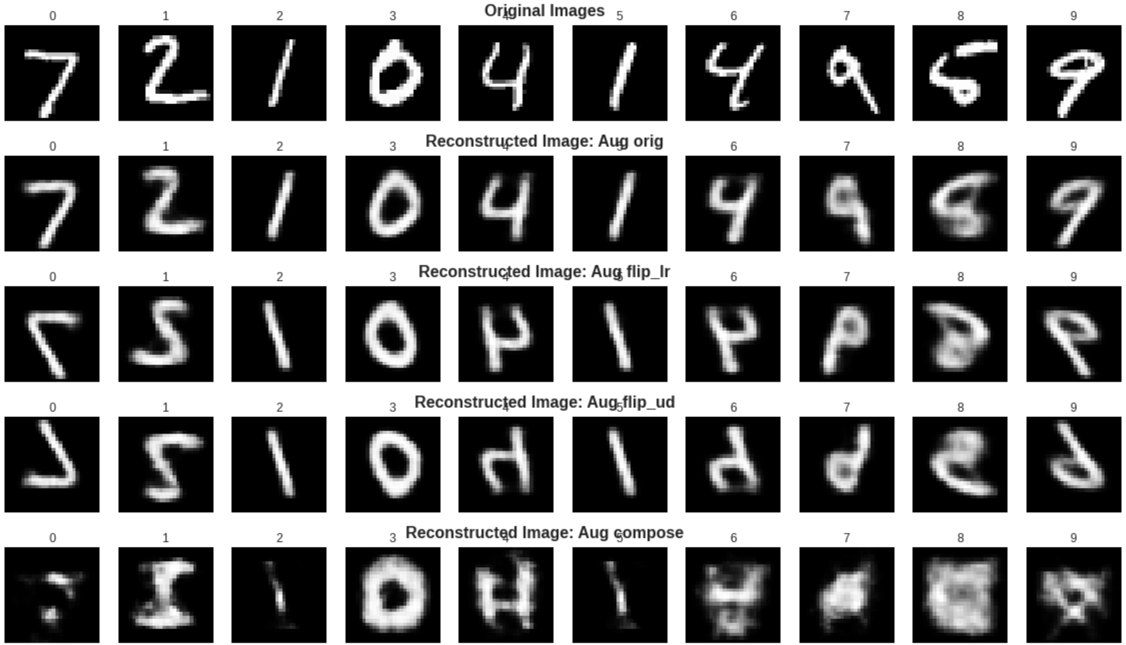




CVAE with LAVAE Training

Training the CVAE like LAVAE:

- Can teach CVAE to learn how to decode uniquely among augmentations.
- Composition does not emerge from training
- Can only input original images (not any-to-any)

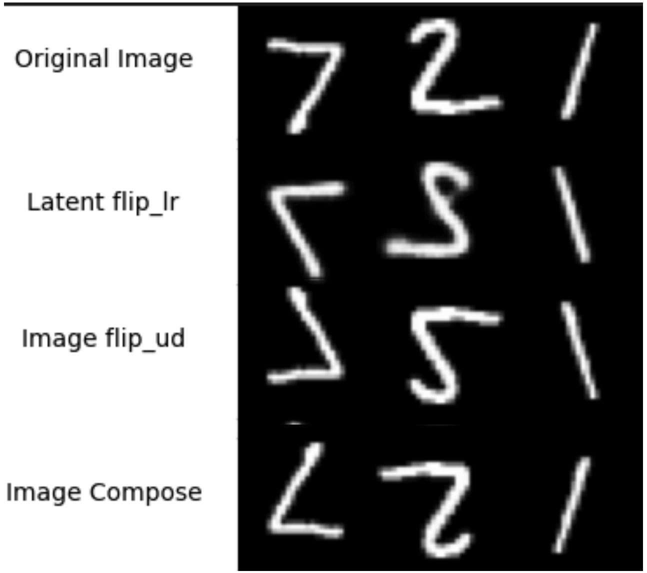


CVAE with LAVAE Training

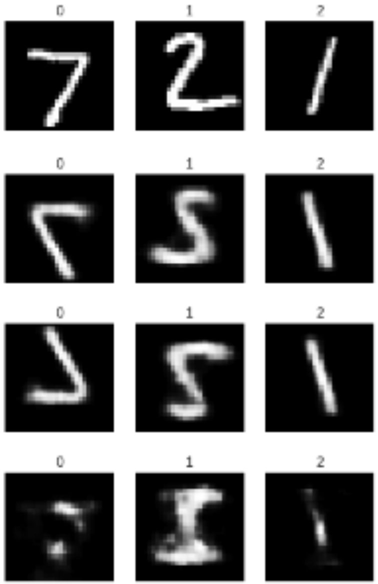
Training the CVAE like LAVAE:

- Can teach CVAE to learn how to decode uniquely among augmentations.
- Composition does not emerge from training
- Can only input original images (not any-to-any)

LAVAE



CVAE



Future Research

- Next phase, potentially **combine CVAE and LAVAE**, and **explore latent flow-based models** to better condition and partition based on dependencies of interest
- Utilize synthetic images with controlled variance instead of augmentation pairs to **construct latent space**
- **Use for causal models**
  - Better causal disentanglement
  - Explicit latent space partitioning and interpretability from priors

Thank You

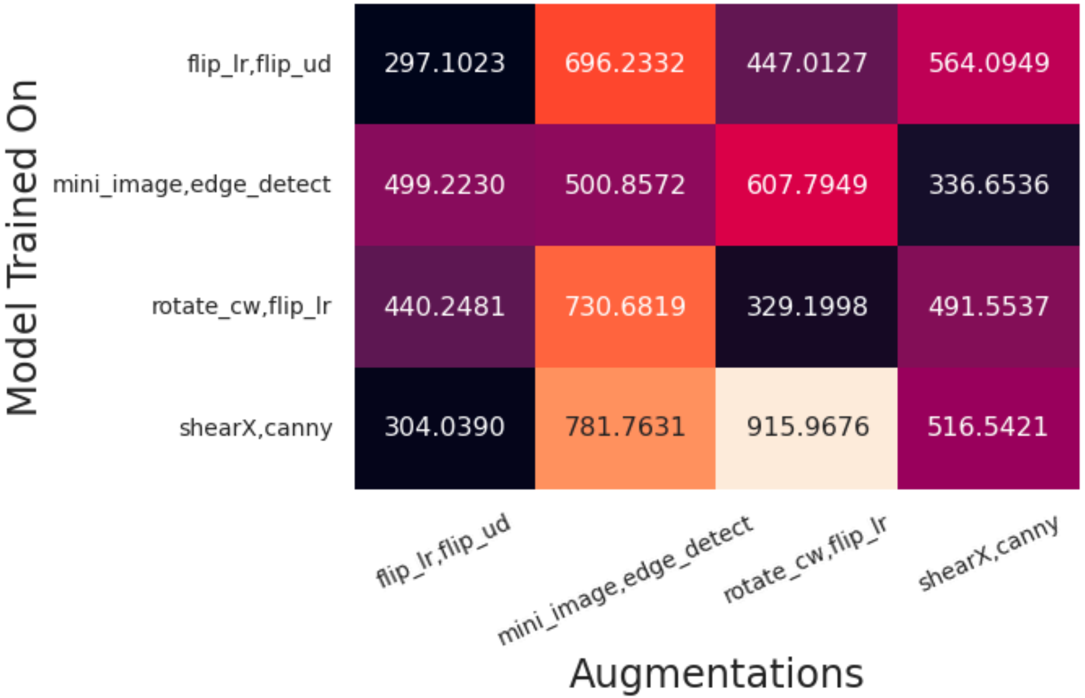
References

[1] S. Bhat, J. Jiang, O. Pooladzandi, and G. Pottie, “De-biasing generative models using counterfactual methods,” in 2022 Information Theory and Applications Workshop, 2022.

[2] M. Yang, F. Liu, Z. Chen, X. Shen, J. Hao, and J. Wang. CausalVAE: Disentangled representation learning via neural structural causal models. arXiv preprint arXiv:2004.08697, 2020.

Backup

Performance on other Augmentation Pairs



Enforcing Involutory Transformations

Training with Naturally Involutory Augmentations

- Flipping an image across any axis through the center of the image, notably horizontally and vertically.
- Rotating an image 180 degrees (composition of two orthogonal flips)

Training with Naturally Involutory Augmentations

- Partitions the latent space into augmentation regions.

Involutory Loss

$$L_{invol} = \| I - A^2 \|^2$$