Georgia Tech
CREATING THE NEXT

MANGO: Learning Disentangled Image Transformation **Man**ifolds with **G**rouped **O**perators

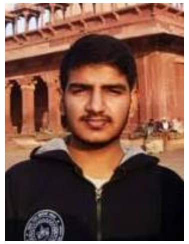

Brighton Ancelin*

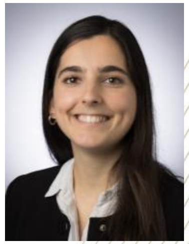

Yenho Chen*


Alex Saad-Falcon*


Peimeng Guan*


Chiraag Kaushik*


Nakul Singh*

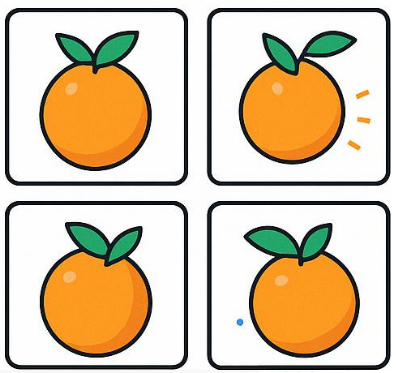

Belen Martin-Urcelay* (speaker)

SampTA - July 29, 2025

*Alphabetical order

1. Background Motivation

Image transformations are everywhere



Data augmentation

Image editing

Virtual Reality

We want a method to <u>learn to generate these transformations</u> from data

- ❑ identity preserving
- ❑ disentangled
- ❑ interpretable
- ❑ with low computational cost

GPT Image 1

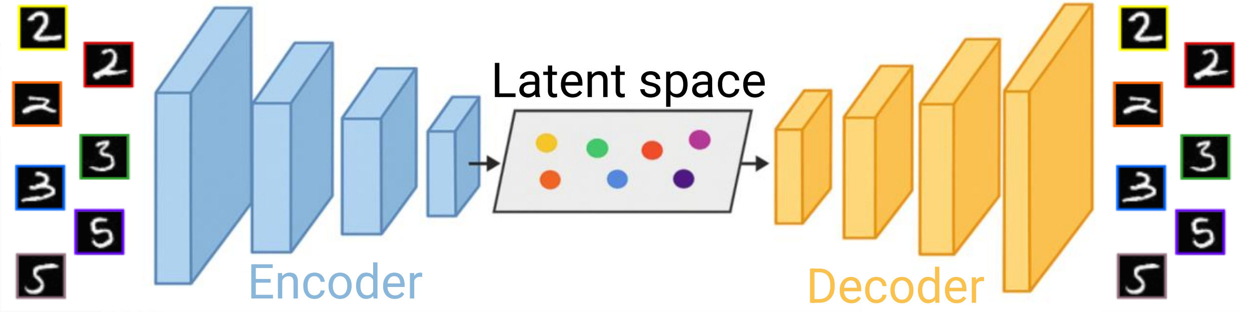
1. Background Motivation

Manifold hypothesis: Within-class object variations lie on or near a low-dimensional, nonlinear manifold and different objects are separated by low density regions. (Cayton, 2005; Narayanan and Mitter, 2010; Bengio et al., 2013)

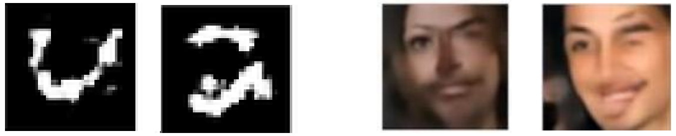
Connor et al. 2021

1. Background Latent Spaces

Autoencoders (AE) transform high dimensional data in a low dimensional latent space



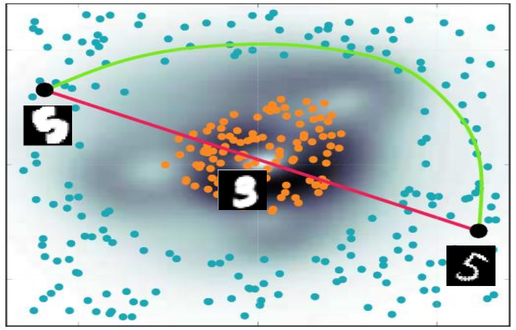
However, **Euclidean transformations** in traditional latent spaces often **lead to unrealistic samples**



(Zhixin et al, 2018)

Regularization procedure which encourages interpolated outputs to appear more realistic by fooling a critic network. Berthelot et al, 2018.

❌ Identity not preserved



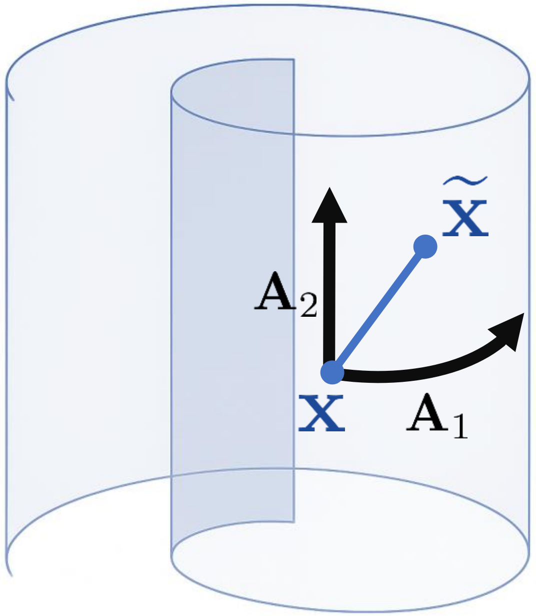
(Arvanitidis et al, 2018)

1. Background Transport Operators

For every point pair $(\mathbf{x}, \widetilde{\mathbf{x}})$ nearby on manifold, we define displacement as **sparse decomposition of Lie group operators:**

$$\mathbf{A} = \sum_{m=1}^{M} \underbrace{\alpha_m}_{\text{latent coefficients}} \underbrace{\mathbf{A}_m}_{\text{transport operators}}$$

$$\widetilde{\mathbf{x}} = \text{expm}(\mathbf{A})\mathbf{x} + \underbrace{\boldsymbol{\epsilon}}_{\text{noise}}$$



Training:

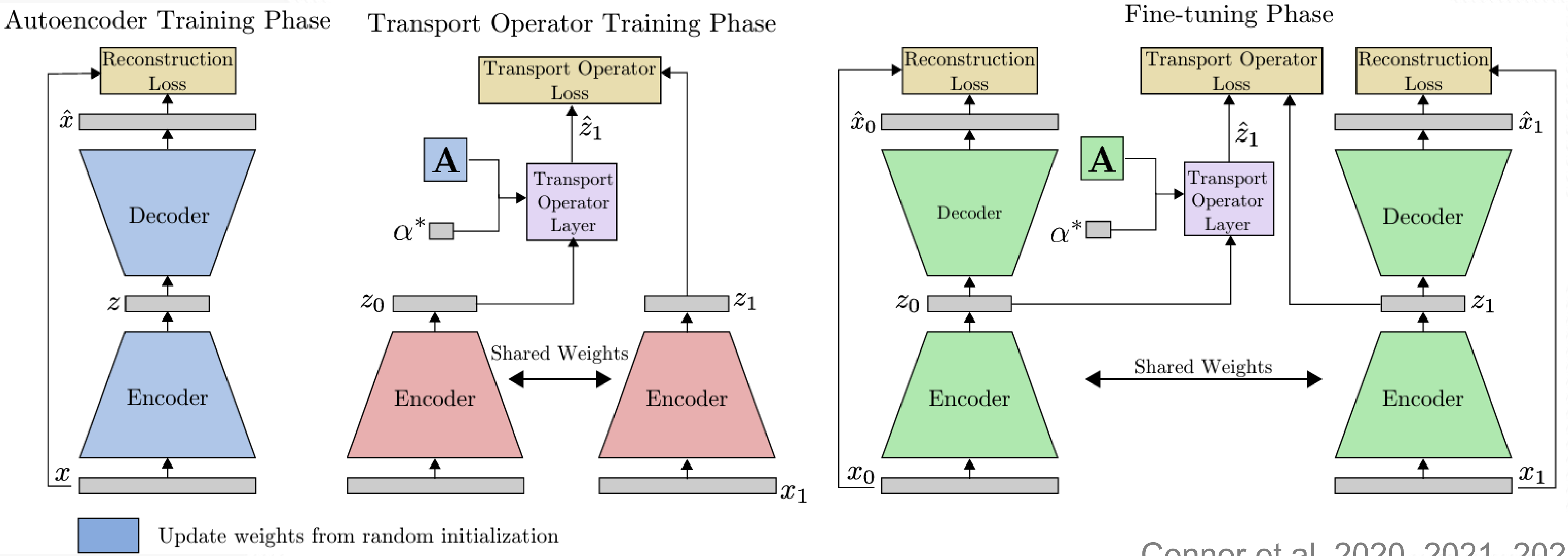
$$L = \frac{1}{2} \left\|\widetilde{\mathbf{x}} - \text{expm}\left(\mathbf{A}\right)\mathbf{x}\right\|_2^2 + \lambda_1 \sum_m \left\|\mathbf{A}_m\right\|_F^2 + \lambda_2 \|\boldsymbol{\alpha}\|_1$$

- Model is trained using point pairs
- Alternating between coefficient inference and gradient steps on the transport operators

Culpepper & Olshausen, 2009, Sohl-Dickstein et al, 2017

1. Background Manifold Autoencoder (MAE)

Learn transport operators in low-dimensional latent space of an autoencoder



Connor et al, 2020, 2021, 2023

❌ No guarantee latent operators are disentangled

❌ No guarantee latent operators are interpretable

❌ Expensive training procedure

2. MANGO Transformation Manifolds with Grouped Operators

- ✓ Disentanglement
- Force each transformation to occupy a different block with no overlapping support

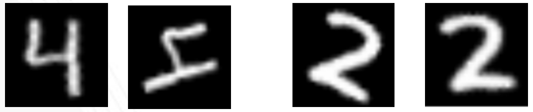
$$\boldsymbol{A}_1 = \begin{bmatrix} \widehat{\boldsymbol{A}}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \vdots \\ \vdots & & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & & \mathbf{0} \end{bmatrix}, \quad \boldsymbol{A}_2 = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \widehat{\boldsymbol{A}}_2 & \ddots & \vdots \\ \vdots & & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & & \mathbf{0} \end{bmatrix}, \ldots, \quad \boldsymbol{A}_M = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \vdots \\ \vdots & & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & & \widehat{\boldsymbol{A}}_M \end{bmatrix}$$

$$\mathbf{A} = \sum_{m=1}^{M} \alpha_m \mathbf{A}_m = \begin{bmatrix} \alpha_1 \widehat{\boldsymbol{A}}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \alpha_2 \widehat{\boldsymbol{A}}_2 & \ddots & \vdots \\ \vdots & & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & & \alpha_M \widehat{\boldsymbol{A}}_M \end{bmatrix}$$

2. MANGO Transformation Manifolds with Grouped Operators

- ✓ Interpretability
- Supervision: Allow practitioners to describe transformations with examples

1) Rotation:



$\alpha = 0$ $\alpha = 0.4$ $\alpha = 0.2$ $\alpha = -0.1$ …

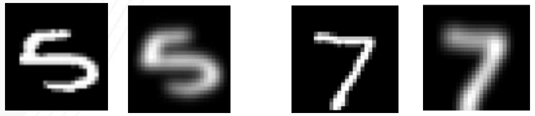
$$\rightarrow T_1 = \|\widetilde{\mathbf{z}} - \text{expm}\left(\textstyle\sum_m \alpha \mathbf{A}_1\right) \mathbf{z}\|_2^2 + \lambda \textstyle\sum_m \|\mathbf{A}_1\|_F^2$$

2) Thickness:

$\alpha = 0.2$ $\alpha = 0.7$ $\alpha = 0.2$ $\alpha = -0.2$ …

$$\rightarrow T_2 = \|\widetilde{\mathbf{z}} - \text{expm}\left(\textstyle\sum_m \alpha \mathbf{A}_2\right) \mathbf{z}\|_2^2 + \lambda \textstyle\sum_m \|\mathbf{A}_2\|_F^2$$

3) Blur:



$\alpha = 0$ $\alpha = 0.4$ $\alpha = 0.1$ $\alpha = 0.4$ …

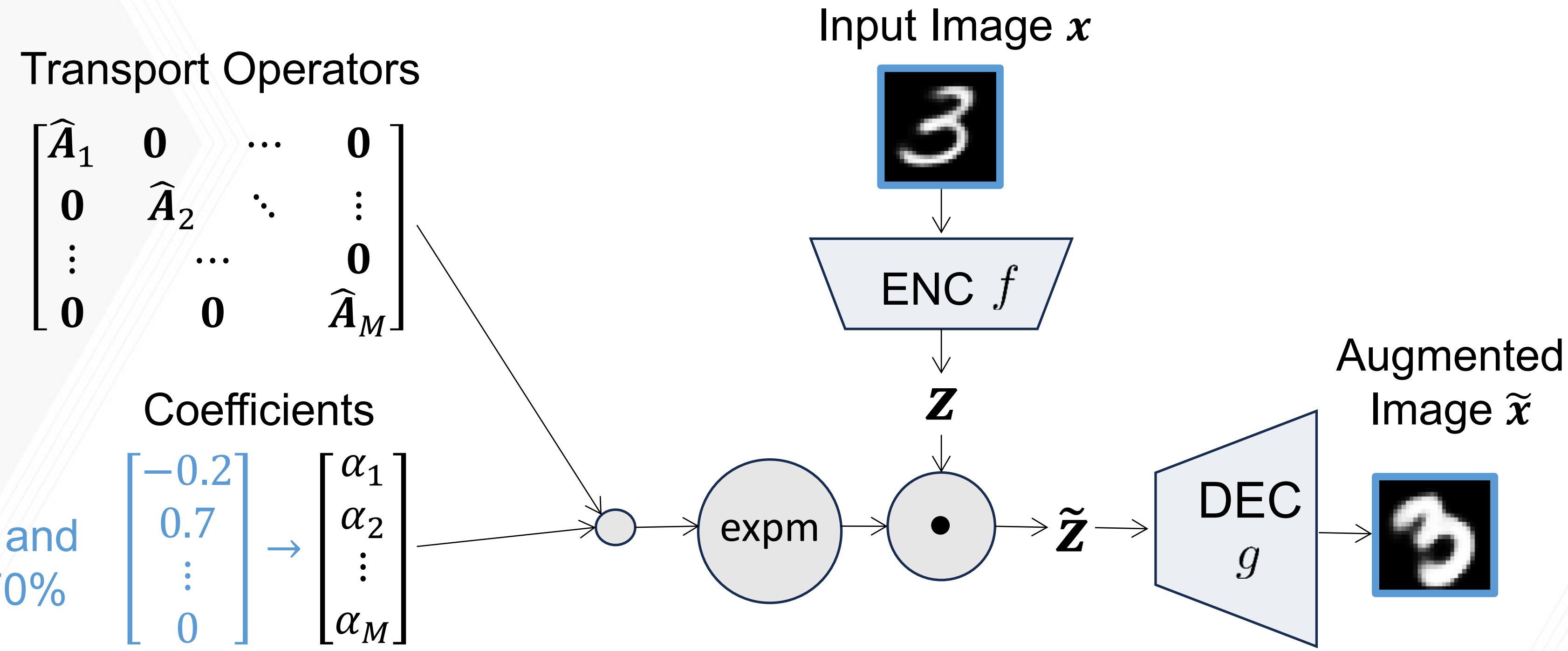
$$\rightarrow T_3 = \|\widetilde{\mathbf{z}} - \text{expm}\left(\textstyle\sum_m \alpha \mathbf{A}_3\right) \mathbf{z}\|_2^2 + \lambda \textstyle\sum_m \|\mathbf{A}_3\|_F^2$$

⋮

2. MANGO Transformation Manifolds with Grouped Operators

Training:

$$L = \underbrace{\|\boldsymbol{x} - g(f(\boldsymbol{x}))\|_2^2 + \|\tilde{\boldsymbol{x}} - g(f(\tilde{\boldsymbol{x}}))\|_2^2}_{\text{Reconstruction error}} + \sum_{m=1}^{M} \gamma \left(\underbrace{\|f(\tilde{\boldsymbol{x}}) - \mathrm{expm}\left(\sum_m \alpha \mathbf{A}_m\right) f(\boldsymbol{x})\|_2^2 + \lambda \sum_m \|\mathbf{A}_m\|_F^2}_{m\text{-th transport operator error } (T_m)} \right)$$



Task
Rotate -36º and thickness 70%

✓ Fast training

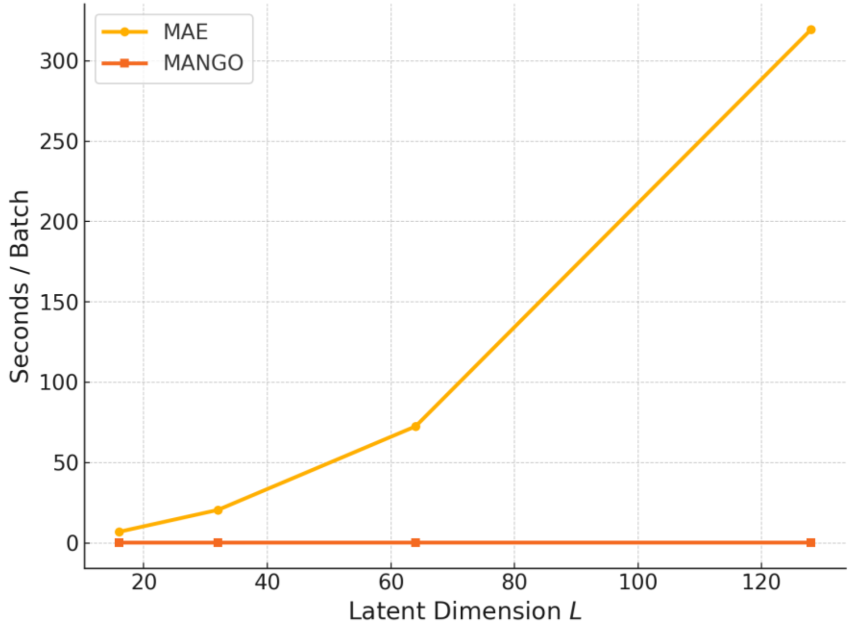
3. Results Improved Training Time

✓ Fast Training

- Supervision leads to simple backpropagation → avoid costly L1 regularization.
- MNIST experiments:
  - MAE (baseline)138 hours vs. MANGO (ours) 12 mins → **0.14% of the time**
  - MAE scales with L^2 while MANGO has nearly constant runtimes.

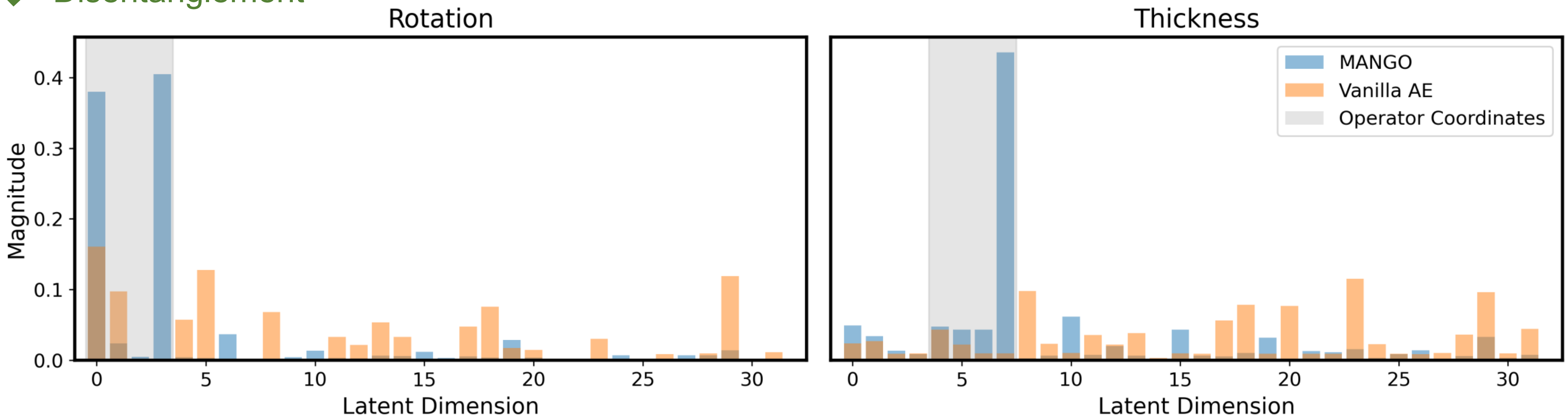
Latent dimension L	16	32	64	128
MAE	6.90	20.52	72.54	319.50
MANGO	0.18	0.18	0.20	0.20

Training runtimes (in seconds) per batch (of size 64)



3. Results Operators are Disentangled

✓ Disentanglement



Learned transformations match the coordinates we were aiming for

MIG score (Chen et. al. 2018)

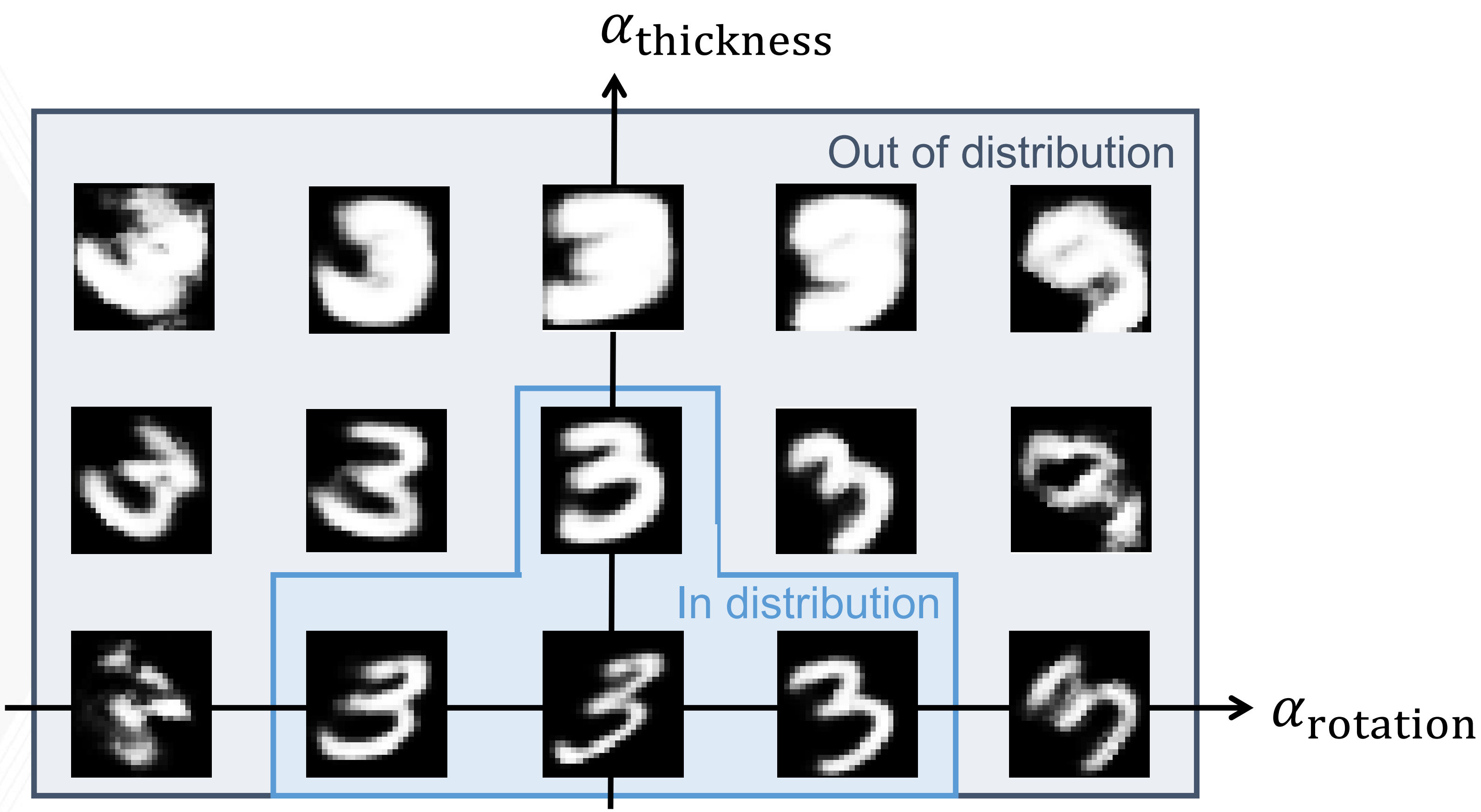
	Vanilla AE	MANGO (ours)
Rotation	0.005	**0.031**
Thickness	0.034	**0.11**

(higher is better)

Georgia Tech
CREATING THE NEXT

3. Results Generalization Beyond Training Dataset

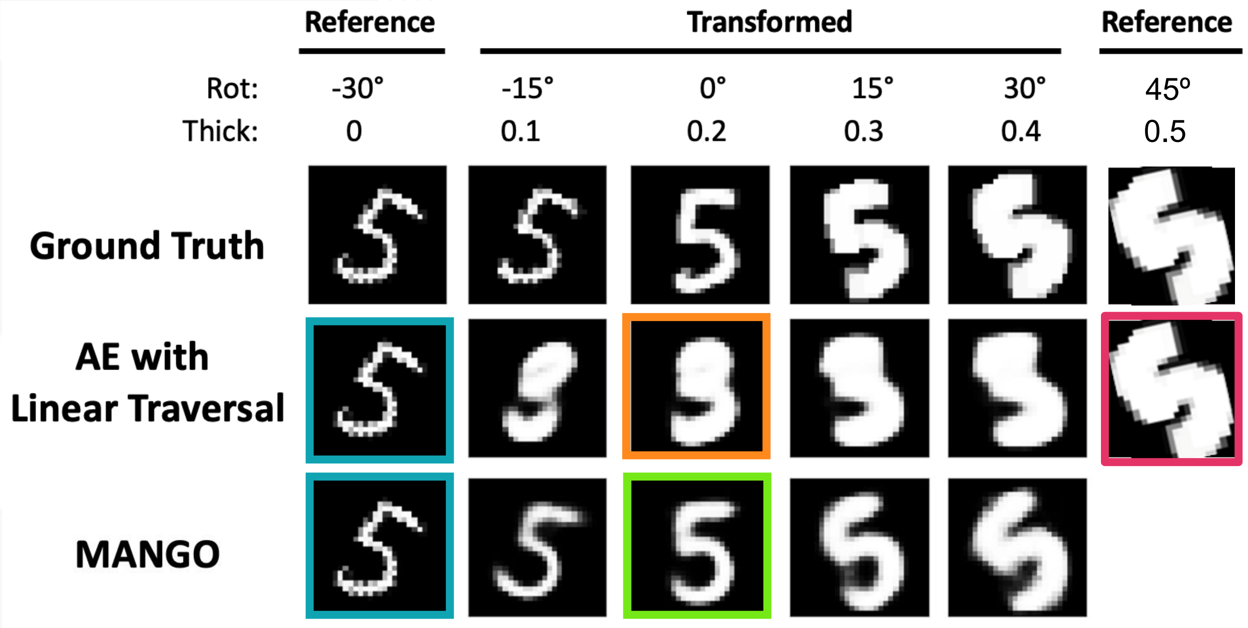
✓ Interpretability

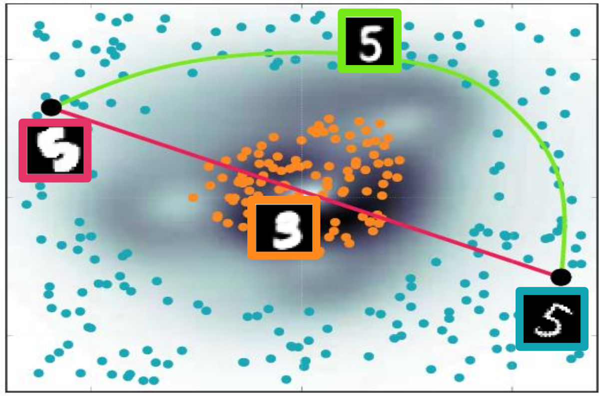


The learned operators are:
1. Capable of transporting images further than the transformations observed during training
2. Composable to achieve complex transformations.

3. Results Image Reconstruction

✓ Identity preserving





60% improvement in transformed reconstructions

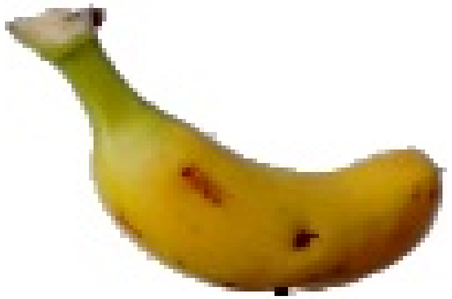
Models	α	Transformed MSE	Transformed LPIPS
AE	rotate	0.072	0.119
	thick	0.093	0.082
	rotate + thick	0.076	0.085
MANGO	rotate	**0.027**	**0.057**
	thick	**0.022**	**0.039**
	rotate + thick	**0.040**	**0.067**

(lower is better)

3. Results Image Reconstruction

4. Summary & Future Work

Training set: Bananas in Fruit 360 dataset with $\alpha \in [-0.25, 0.25]$



Generated images of rotated bananas, for rotation angles **not seen in the training set:**

α =-0.5 | α =-0.62 | α =-0.75 | α =-0.88 | $\alpha = -1$

4. Summary & Future Work

MANGO (transformation **Man**ifolds with **G**rouped **O**perators)
Method to learn to generate image transformations from data

- ✓ Identity preserving
- ✓ Disentangled: Enables composition of image transformations
- ✓ Interpretable: Practitioners define which transformations to model
- ✓ With low computational cost: 100x speed up in training time

Georgia Tech

CREATING THE NEXT

MANGO: Learning Disentangled Image Transformation **Man**ifolds with **G**rouped **O**perators

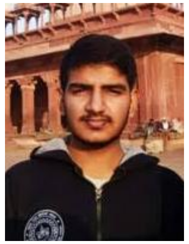

Brighton Ancelin*

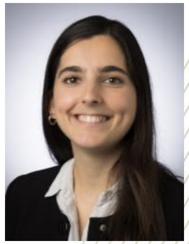

Yenho Chen*


Alex Saad-Falcon*


Peimeng Guan*


Chiraag Kaushik*


Nakul Singh*


Belen Martin-Urcelay* (speaker)

SampTA - July 29, 2025

*Alphabetical order