Learning Group Actions on Latent Representations

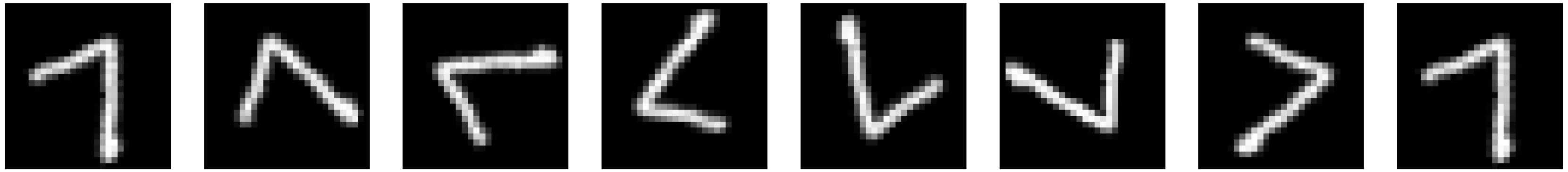
Yinzhu Jin, Aman Shrivastava, P. Thomas Fletcher
Department of Computer Science
Department of Electrical and Computer Engineering
University of Virginia

Learning Group Actions

Group actions represent symmetries and geometric transformations of data.

Recent work shows that explicitly modeling and learning group actions enhances performance across various tasks.

An example: 2D rotations



Group Actions on Latent Factors

Group actions on latent factors but not on the image itself.

We propose to learn group actions on latent representations.

A group element g in G acting on z is denoted as $g.z$

A group action satisfies:

Identity: $\exists e \in G, \quad e.g = g$

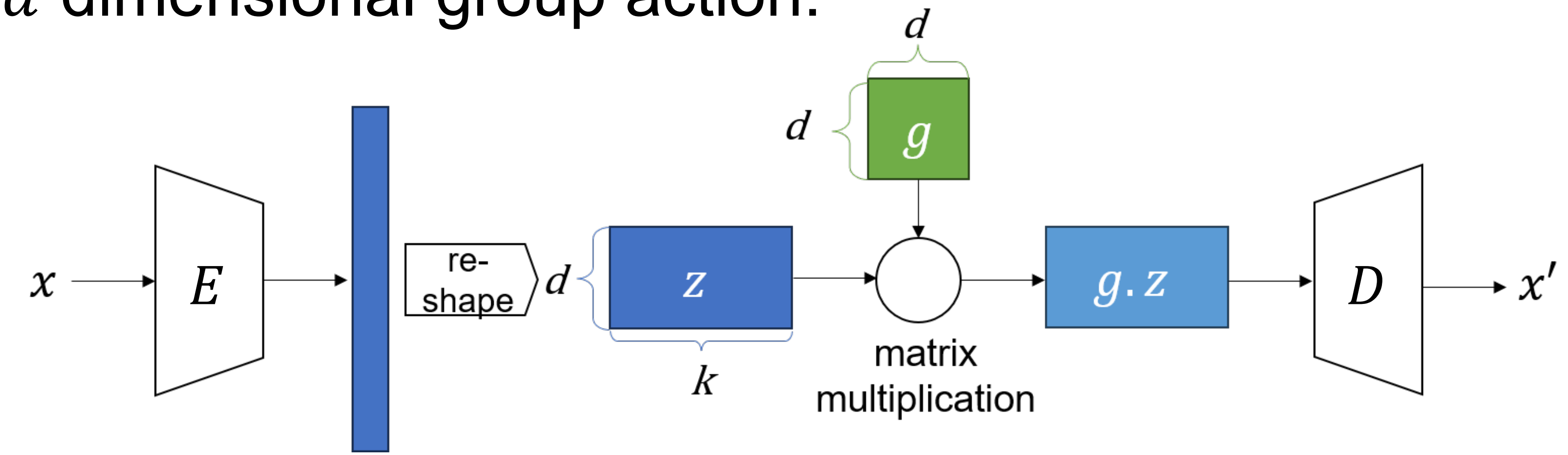
Compatibility: $\forall g_1, g_2 \in G, \quad g_2.(g_1.z) = (g_2 g_1).z$

Related Work

	Winter et al.	Hwang et al.	Dupont et al. (novel view synthesis only)	Sajjadi et al. (novel view synthesis only)	Ours
Models group action	✓	✓	✓	✗	✓
Can handle general groups	✓ (given tailored architecture)	✗ (additive $\mathbb{R}^n$ group only)	✗	✗	✓
Models latent group action	✗	✗	✓	-	✓
No group tailored architecture	✗	✓	✓	-	✓

Latent Space Group Action Model

For a d dimensional group action:



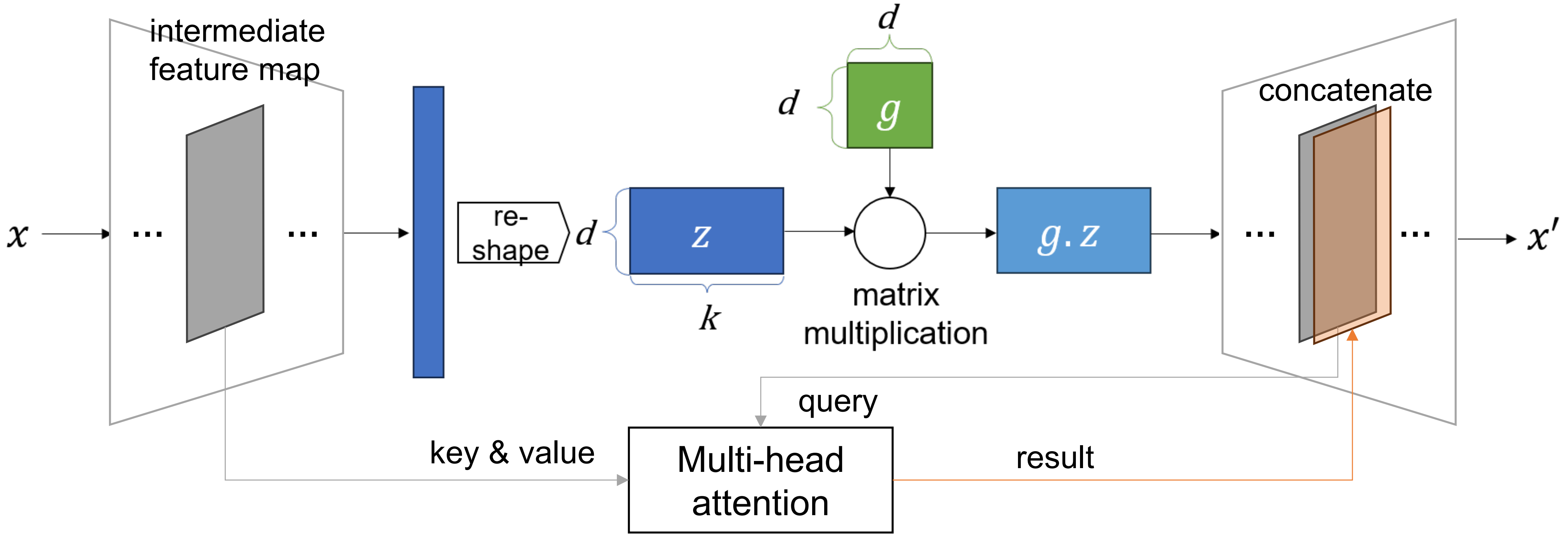
E: encoder, D: decoder

Optionally, decompose z into varying and invariant parts:

$$z = [z_v; z_i], \qquad g.z = [g.z_v; z_i]$$

Latent Space Group Action Model

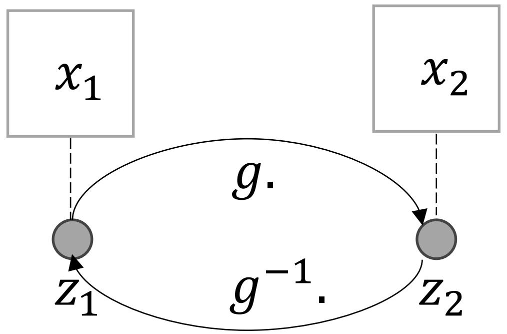
Alternatively, use skip connection with attention to model invariant components and image details better.



Training Objective

Take a pair of data x_1, x_2, such that their latents lie on the same orbit:

$$z_2 = g.z_1$$



The training loss is the reconstruction loss $\mathcal{L}_{\mathcal{X}}$ with group actions:

$$\mathcal{L} = \mathcal{L}_{\mathcal{X}}\big(x_2, D(g.z_1)\big) + \mathcal{L}_{\mathcal{X}}\big(x_1, D(g^{-1}.z_2)\big),$$
$$z_1 = E(x_1), \qquad z_2 = E(x_2).$$

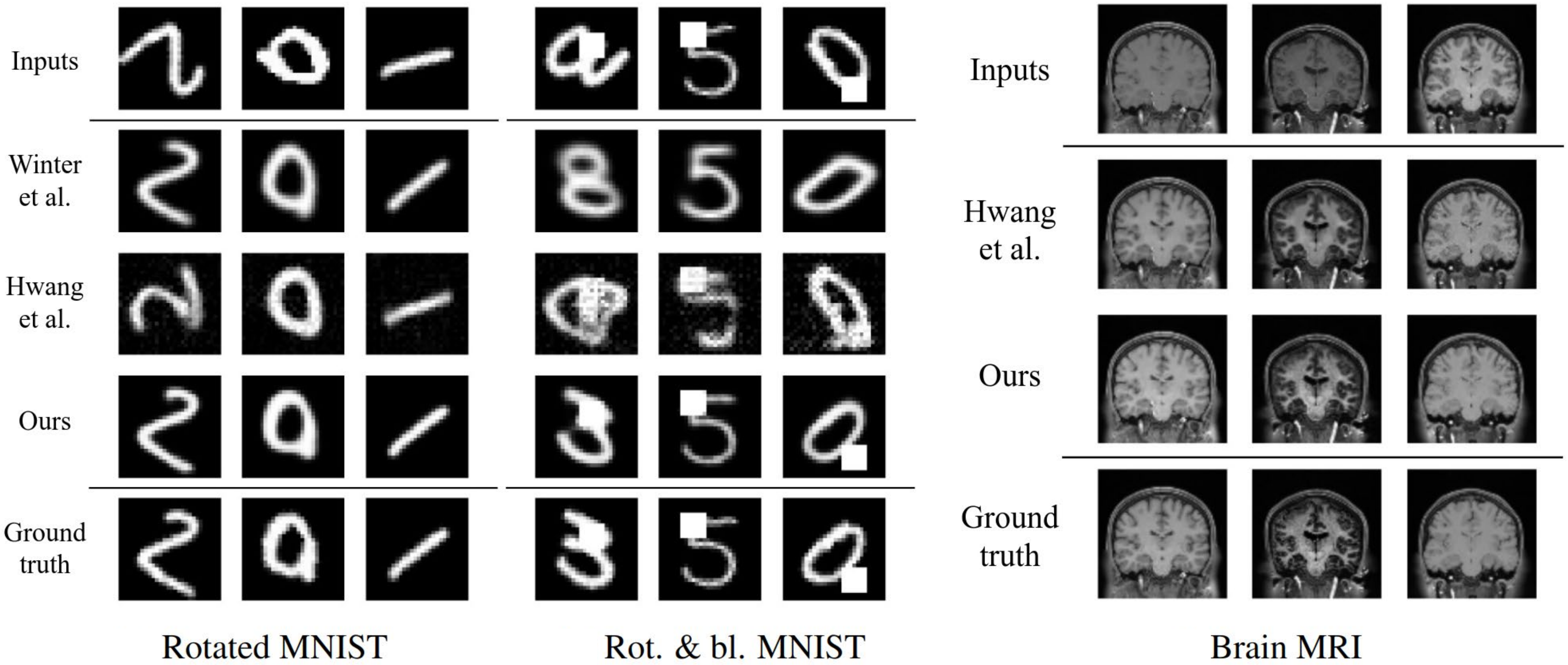
Induced Group Actions on Data Space

We prove that a group action on the latent space $\alpha_g : \mathcal{Z} \rightarrow \mathcal{Z}$, induces a group action $\tilde{\alpha}_g$ on the reconstructed data space $\mathcal{X}'$,

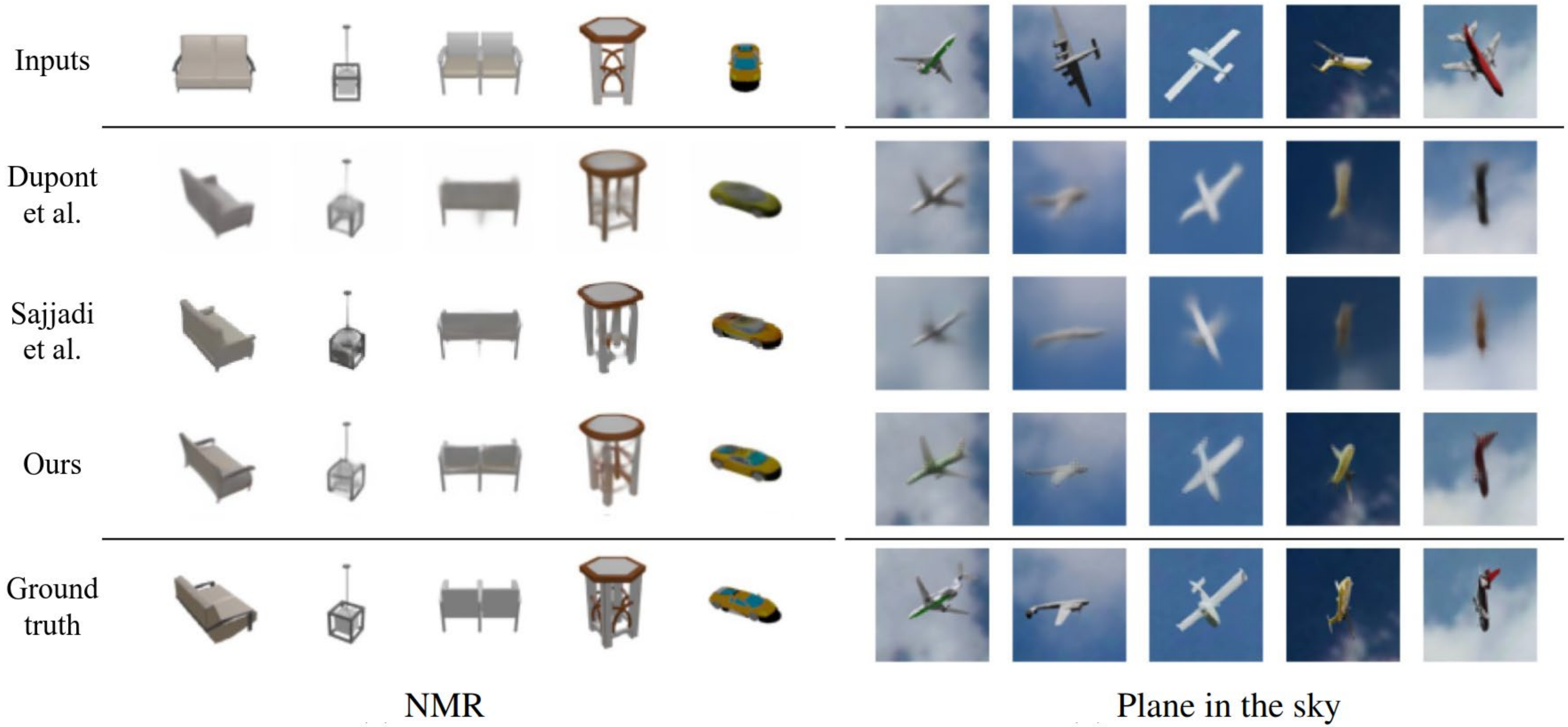
$$\begin{array}{ccc} \mathcal{Z} & \xrightarrow{\alpha_g} & \mathcal{Z} \\ {\scriptstyle D}\downarrow & & \downarrow{\scriptstyle D} \\ \mathcal{X}' & \xrightarrow{\tilde{\alpha}_g} & \mathcal{X}' \end{array}$$

if $\underline{E\big(D(z)\big) = z}$, i.e. the latent representation is reconstructable.

Experiments



Experiments



Experiments

Table 1: Quantitative results on MNIST derived datasets and brain MRI dataset

	Rotated MNIST		Rot. & bl. MNIST		Brain MRI	
	↑PSNR	↑SSIM	↑PSNR	↑SSIM	↑PSNR	↑SSIM
Winter et al. [32]	21.97	0.874	14.05	0.586	NA	NA
Hwang et al. [13]	15.29	0.992	10.19	0.990	27.43	**1.000**
Ours	**26.07**	**1.000**	**23.55**	**1.000**	**35.99**	**1.000**

Table 2: Quantitative results on 3D objects rendered datasets

	NMR			Plane in the sky		
	↑PSNR	↑SSIM	↓LPIPS	↑PSNR	↑SSIM	↓LPIPS
Dupont et al. [9]	26.91	0.899	0.091	24.25	0.773	0.239
Sajjadi et al. [25]	27.87	0.912	0.066	23.53	0.489	0.280
Ours	**28.91**	**0.947**	**0.050**	**25.24**	**0.821**	**0.112**

Swapping Varying and Invariant Parts

Acknowledgement

Thank you!

This work was partially supported by NSF Smart and Connected Health grant 2205417.