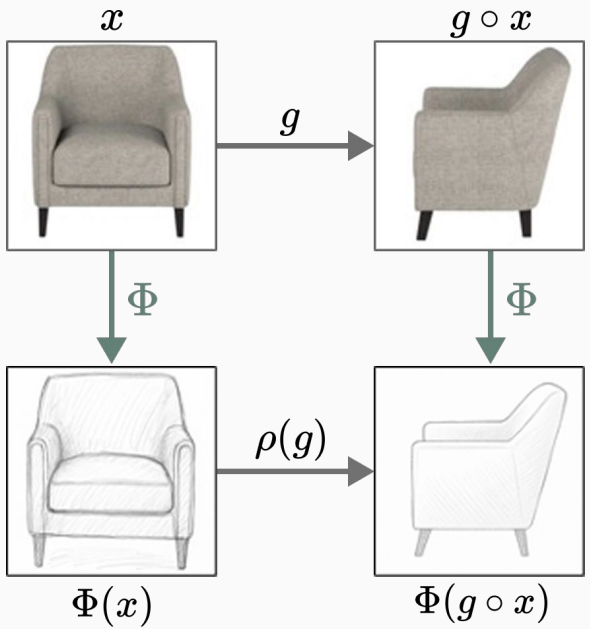
ICML
International Conference
On Machine Learning

Equivariant Latent Alignment via Flow Matching under Group Symmetries

Sunghyun Kim* | Jaehoon Hahm* | Jeongwoo Shin | Joonseok Lee

Seoul National University
University of Illinois Urbana-Champaign

Background – Equivariant Representation Learning



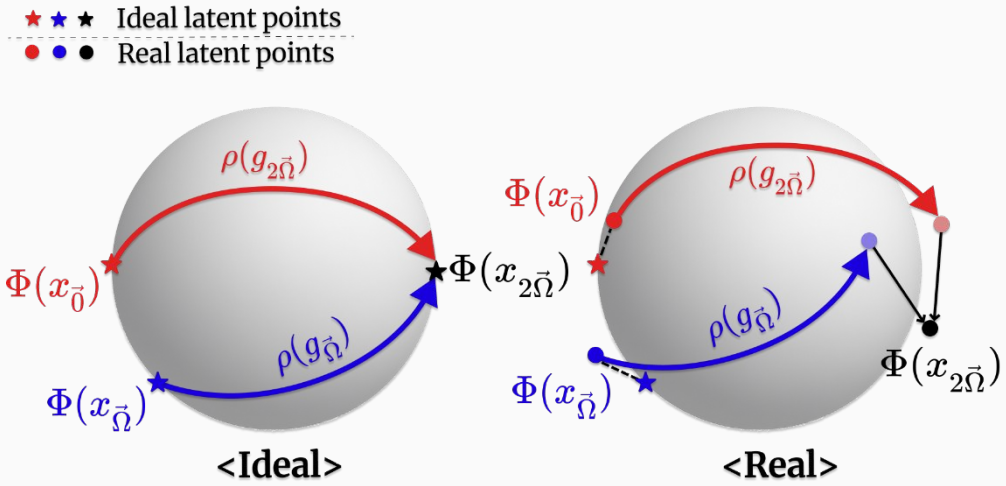
$$\mathcal{L}_{\text{ERL}} = \underbrace{\mathbb{E}[\|\Phi(g \circ x) - \rho(g)\Phi(x)\|_2^2]}_{\text{Equivariance Loss}} + \underbrace{\mathbb{E}[\|g \circ x - \Psi(\rho(g)\Phi(x))\|_2^2]}_{\text{Decoder Loss}}$$

- **Method for Novel View Synthesis (NVS)**
- ERL framework typically consists of two losses:
  - Equivariance loss to ensure that latent features transform consistently with the underlying group action
  - Decoder loss to preserve sufficient information for reconstruction

- Following NFT, we use the Wigner D representation to implement rotations, with our latents residing in the space on which these representations act.

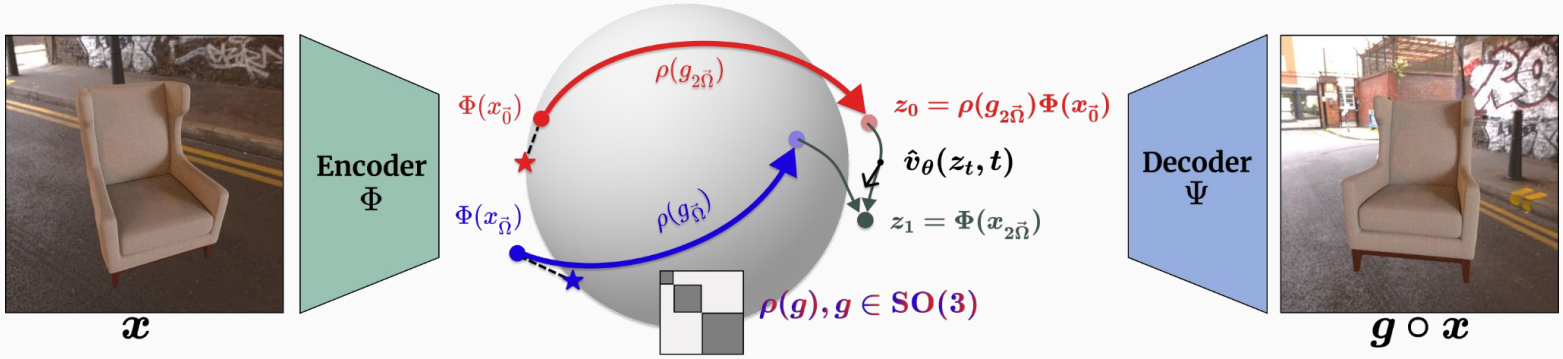
Koyama, Masanori, et al. "Neural fourier transform: A general approach to equivariant representation learning." *International Conference on Learning Representations*. Vol. 2024. 2024.

Motivation – Latent Misalignment



- For a given same object, analytically rotated and true encoding diverges.
- We aim to mitigate this gap by introducing transportation mechanism while preserving the group-theoretic foundation.

Method – Residual Latent Flow



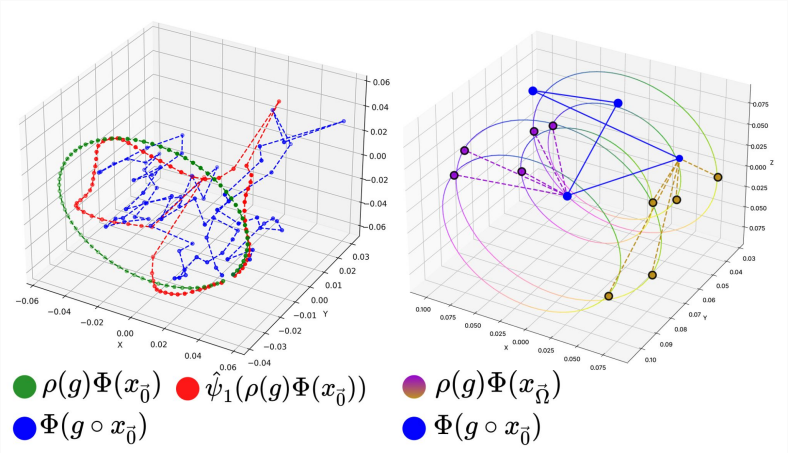
- Let (x,g)~q be a data pair sampled from a dataset, where x is an image and g is a known group transformation.
- The latents z0 and z1 are not precisely aligned in practice due to the imperfect encoder and data variability.

$$z_0 := \rho(g)\Phi(x), \qquad z_1 := \Phi(g \circ x)$$

- To correct this misalignment, we connect z0 to the correct z1 via training an appropriate flow with the RLF loss:

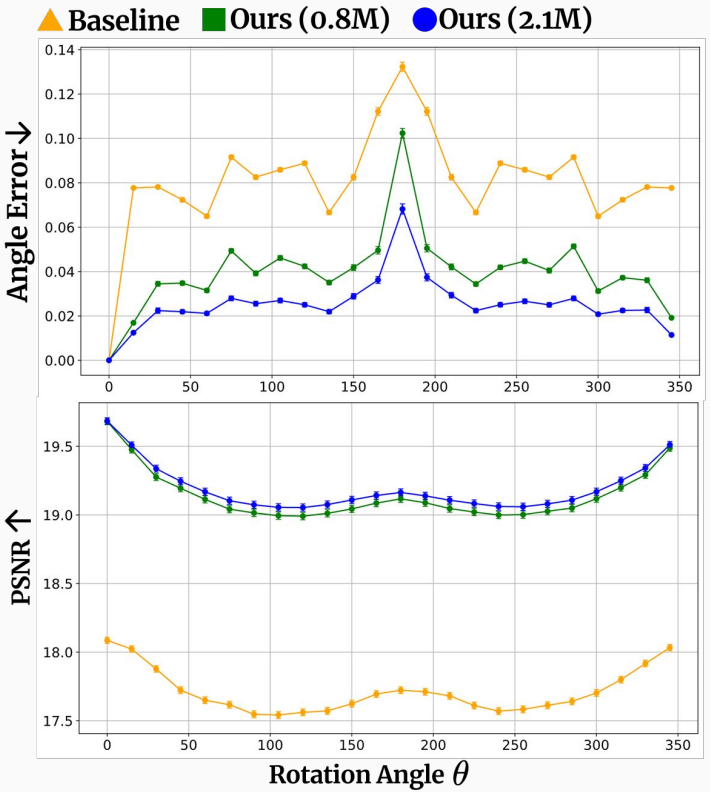
$$\mathcal{L}_{\text{RLF}}(\theta) = \mathbb{E}_{(x,g)\sim q(\cdot),\, t\sim\mathcal{U}[0,1],\, z_t\sim p_t(\cdot|z_0,z_1)}\left[\|v_\theta(z_t,t) - (z_1 - z_0)\|_2^2\right]$$

Why Flow Matching?



- **Latent correction is a distribution transport problem**
  Multiple transformed latent representations should align to the same equivariant target latent.
- **Encoder imperfections create multimodal targets**
  Different inputs can collapse to the same latent, leading to multiple valid target latents after group transformations.
- **Flow Matching naturally handles multimodal transport**
  Instead of deterministic regression, it learns a flexible distribution-to-distribution mapping and can transport arbitrary latent distributions without relying on the choice of priors.

Experiments - Quantitative Comparisons



- **Evaluation on SO(2) dataset (ComplexBRDFs (OOD)) across angular displacements.**
- This measures reconstruction quality with angle error and PSNR.
- The results are shown for two RLF models with 0.8M and 2.1M parameters.

Experiments - Quantitative Comparisons

Group	Dataset	Method	Pred Err.↓	LPIPS↓	PSNR↑	Latent Err.↓	Angle Err.↓
SO(3)	ABO	Base	0.0667 ± 0.0004	0.4434 ± 0.0002	11.85 ± 0.01	7.3×10^{-4} ± 1.3×10^{-6}	0.0079 ± 0.0001
		Ours	**0.0565** ± 0.0004	**0.4267** ± 0.0002	**12.57** ± 0.01	**1.9×10^{-4}** ± 5.1×10^{-7}	**0.0010** ± 0.0001
	ABO (OOD)	Base	0.0641 ± 0.0008	0.4405 ± 0.0003	12.14 ± 0.01	8.1×10^{-4} ± 1.7×10^{-6}	0.0088 ± 0.0002
		Ours	**0.0564** ± 0.0008	**0.4251** ± 0.0003	**12.73** ± 0.01	**2.1×10^{-4}** ± 7.6×10^{-7}	**0.0012** ± 0.0001
	ModelNet10-SO(3) (OOD)	Base	0.1079 ± 0.0025	0.1176 ± 0.0005	10.09 ± 0.03	7.2×10^{-4} ± 7.1×10^{-6}	0.1746 ± 0.0064
		Ours	**0.1018** ± 0.0026	**0.1084** ± 0.0005	**10.43** ± 0.03	**4.1×10^{-4}** ± 7.4×10^{-6}	**0.0430** ± 0.0018
	SmallNORB (OOD)	Base	0.0052 ± 0.0001	0.2729 ± 0.0002	23.13 ± 0.02	7.9×10^{-5} ± 2.7×10^{-7}	0.2429 ± 0.0023
		Ours	**0.0050** ± 0.0001	**0.2473** ± 0.0002	**23.28** ± 0.02	**4.8×10^{-5}** ± 2.6×10^{-7}	**0.0728** ± 0.0012
SO(2)	ABO Day-to-Night	Base	0.0056 ± 0.0001	0.2151 ± 0.0012	22.73 ± 0.05	1.9×10^{-4} ± 3.0×10^{-6}	0.0456 ± 0.0006
		Ours	**0.0039** ± 0.0001	**0.1973** ± 0.0011	**24.32** ± 0.05	**1.5×10^{-4}** ± 2.7×10^{-6}	**0.0163** ± 0.0003
	ABO Day-to-Night (OOD)	Base	0.0079 ± 0.0005	0.2179 ± 0.0016	21.80 ± 0.08	3.4×10^{-4} ± 6.6×10^{-6}	0.0776 ± 0.0014
		Ours	**0.0065** ± 0.0005	**0.1998** ± 0.0015	**22.84** ± 0.08	**2.6×10^{-4}** ± 6.0×10^{-6}	**0.0269** ± 0.0009
	ComplexBRDFs	Base	0.0382 ± 0.0009	0.3506 ± 0.0002	16.04 ± 0.02	8.9×10^{-4} ± 1.3×10^{-5}	0.0556 ± 0.0003
		Ours	**0.0296** ± 0.0008	**0.3266** ± 0.0002	**17.38** ± 0.02	**6.7×10^{-4}** ± 1.3×10^{-5}	**0.0172** ± 0.0004
	ComplexBRDFs (OOD)	Base	0.0480 ± 0.0026	0.3522 ± 0.0002	17.71 ± 0.02	1.4×10^{-3} ± 3.8×10^{-5}	0.0859 ± 0.0009
		Ours	**0.0404** ± 0.0023	**0.3285** ± 0.0002	**19.19** ± 0.03	**1.1×10^{-3}** ± 3.8×10^{-5}	**0.0271** ± 0.0010
	RotatedMNIST	Base	0.0016 ± 0.0000	0.0035 ± 0.0000	28.21 ± 0.006	3.9×10^{-5} ± 1.4×10^{-7}	0.0032 ± 0.0000
		Ours	**0.0013** ± 0.0000	**0.0030** ± 0.0000	**29.02** ± 0.065	**3.7×10^{-5}** ± 1.4×10^{-7}	**0.0030** ± 0.0000
	RotatedMNIST (OOD)	Base	0.0016 ± 0.0000	0.0040 ± 0.0000	28.17 ± 0.008	3.9×10^{-5} ± 1.8×10^{-7}	0.0032 ± 0.0000
		Ours	**0.0013** ± 0.0000	**0.0033** ± 0.0000	**28.99** ± 0.009	**3.7×10^{-5}** ± 1.8×10^{-7}	**0.0030** ± 0.0000

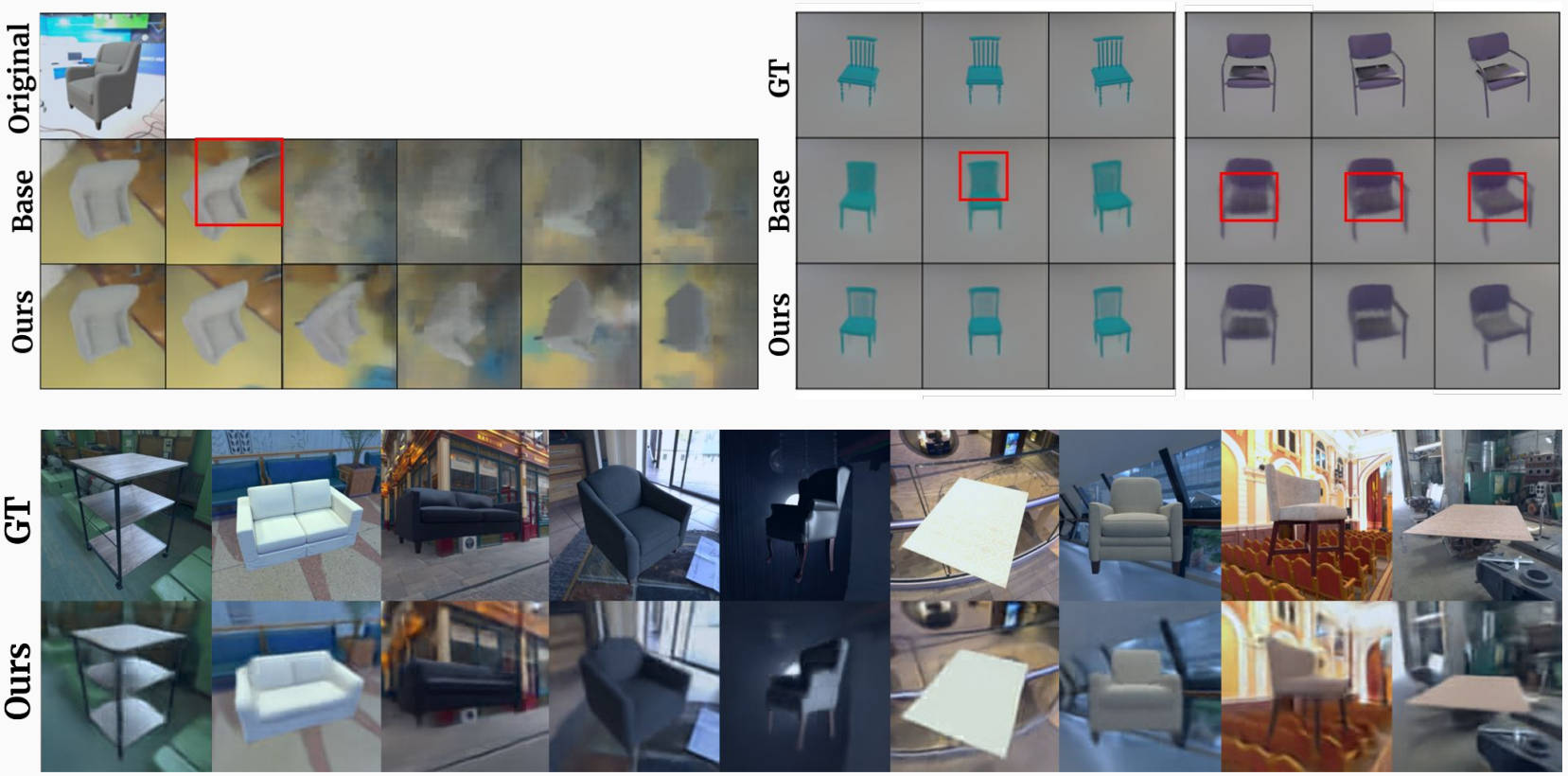
Table 2. **Comparison on in-plane rotation NVS using RotatedMNIST** (SO(2)).

Method	Pred Err. ↓	LPIPS ↓	PSNR ↑	SSIM ↑
SpatialVAE	0.0373	0.1561	15.59	0.6664
GIAE	0.0146	0.1000	19.88	0.8289
LGA	0.0049	0.0385	24.13	0.9427
NFT	0.0016	0.0035	28.21	0.9953
Ours	**0.0013**	**0.0030**	**29.02**	**0.9961**

Table 3. **Comparison on out-of-plane rotation NVS using SmallNORB** (SO(3)).

Method	Pred Err. ↓	LPIPS ↓	PSNR ↑	SSIM ↑
LGA	0.0174	0.270	17.59	0.785
ENR	0.0156	0.262	18.07	0.800
NFT	0.0052	0.272	23.13	**0.811**
Ours	**0.0050**	**0.247**	**23.28**	0.802

Experiments - Qualitative Comparisons



Thank you!

Do you have any questions?

Email: ksho719@snu.ac.kr, jh141@illinois.edu
Code: https://github.com/jaehoon-hahm/residual-latent-flow

Equation Appendix

$$p_t(z_t) = \iint p_t(z_t|z_0, z_1)\, \pi_{0,1}(z_0, z_1)\, dz_0\, dz_1, \quad z_0 \sim p_0(z_0), \quad z_1 \sim p_1(z_1)$$
$$z_0 := \rho(g)\Phi(x), \qquad z_1 := \Phi(g \circ x)$$

$$\pi_{0,1}(z_0, z_1) \neq p_0(z_0)p_1(z_1)$$

$$\pi_{0,1}(z_0, z_1|x, g) = p_0(z_0|x, g)p_1(z_1|x, g)$$

$$\begin{aligned} p_t(z_t) &= \iiiint p_t(z_t|z_0, z_1)\, p_0(z_0|x, g)\, p_1(z_1|x, g)\, q(x, g)\, dx\, dg\, dz_0\, dz_1 \\ &= \iiiint p_t(z_t|z_0, z_1)\, \delta\big(z_0 - \rho(g)\Phi(x)\big)\, \delta\big(z_1 - \Phi(g \circ x)\big)\, q(x, g)\, dx\, dg\, dz_0\, dz_1 \\ &= \iint p_t\big(z_t|\rho(g)\Phi(x),\ \Phi(g \circ x)\big)\, q(x, g)\, dx\, dg \end{aligned}$$

$$\mathcal{L}_{\mathrm{RLF}}(\theta) = \mathbb{E}_{(x,g)\sim q(\cdot),\, t\sim\mathcal{U}[0,1],\, z_t\sim p_t(\cdot|z_0,z_1)}\Big[\|v_\theta(z_t, t) - (z_1 - z_0)\|_2^2\Big]$$