POSTECH
POHANG UNIVERSITY OF SCIENCE AND TECHNOLOGY
ComputerVision Lab.

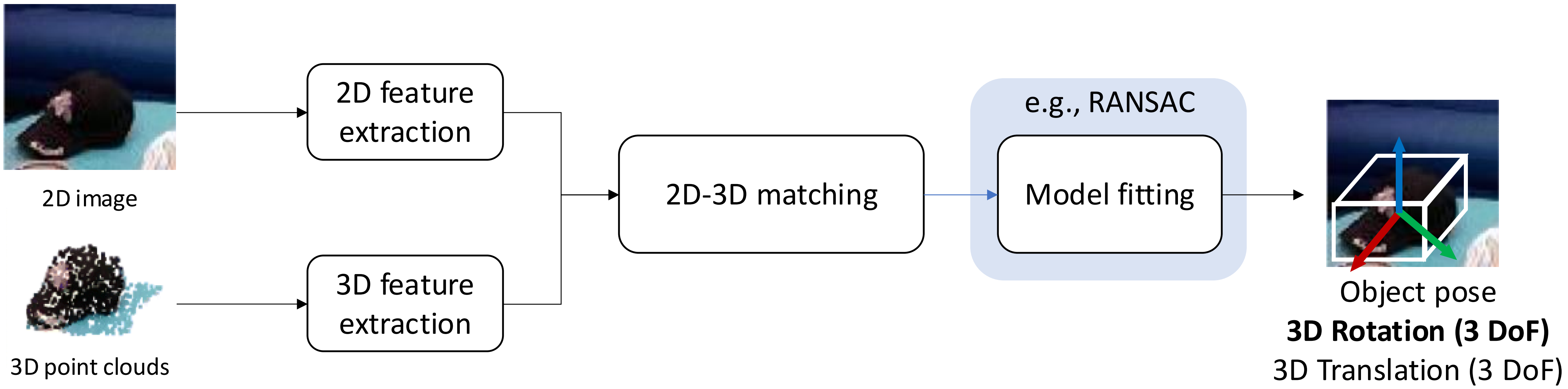
NEURAL INFORMATION PROCESSING SYSTEMS

3D Equivariant Pose Regression via Direct Wigner-D Harmonics Prediction

Jongmin Lee*

Minsu Cho

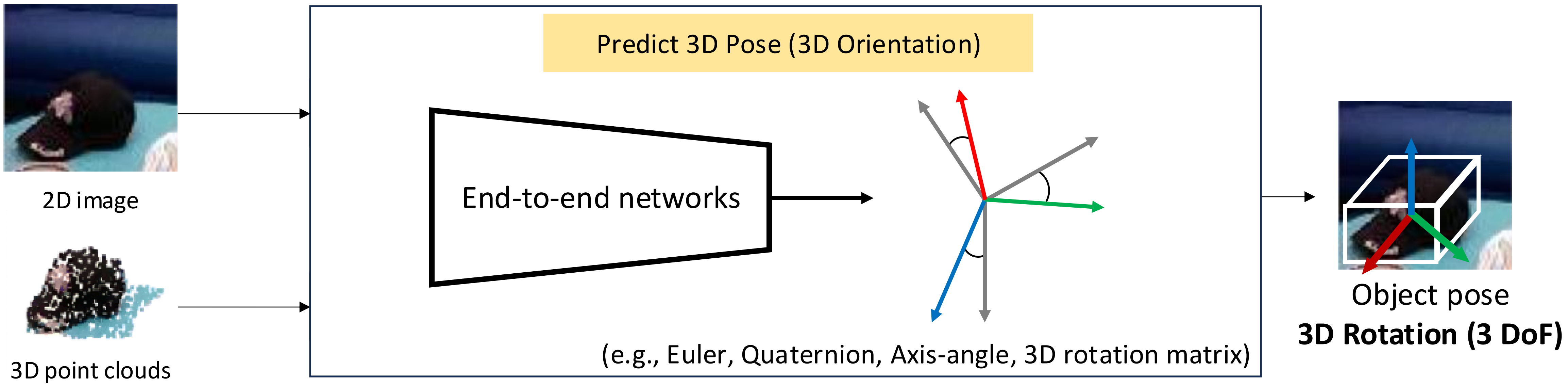
Introduction: Single-View Pose Estimation



Pipeline of object pose estimation

Introduction: Single-View Pose Estimation

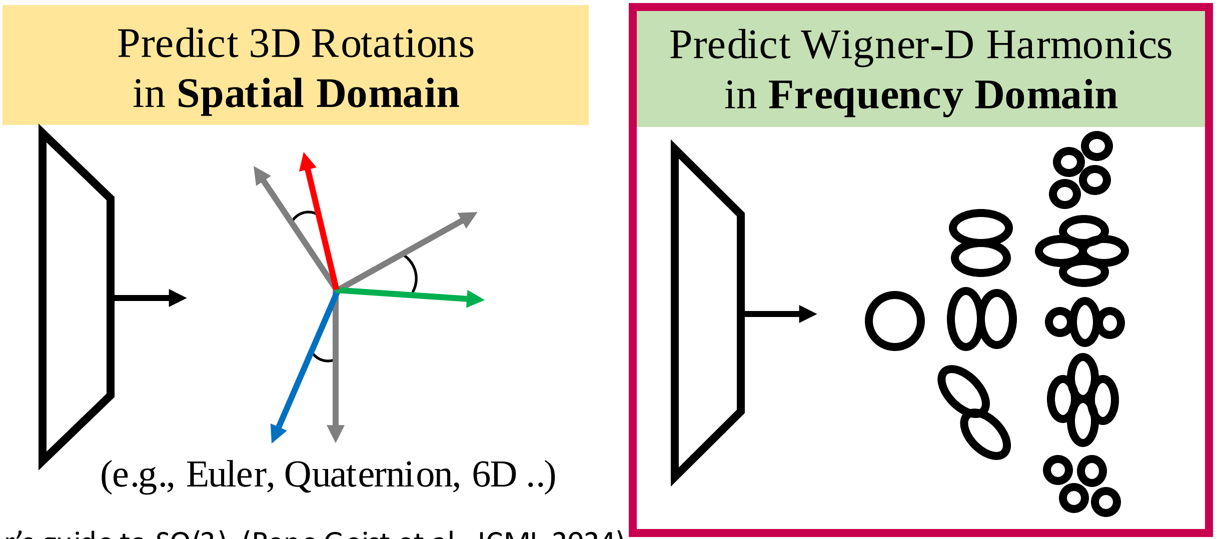
End-to-end pose estimation networks



Pipeline of object pose estimation

Motivation

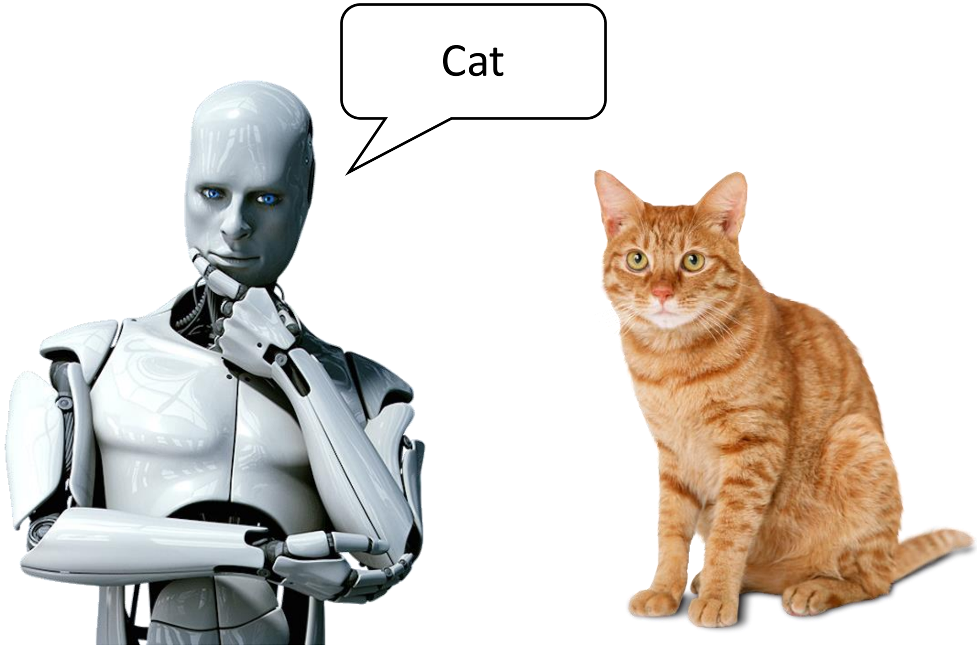
- Problems of existing 3D rotation representations **in spatial domain**[1,2,3]
  - **Discontinuity** (Euler, Quaternions) & **Singularity** (3D rotmats)
- Our solution
  - **Parametrizes** the 3D rotation in **frequency domain**
    - Predicting **Wigner-D matrices** of spherical harmonics
  - Leverages **SO(3) equivariance** for generalization to unseen rotations
    - Using **spherical CNNs** operating in frequency domain

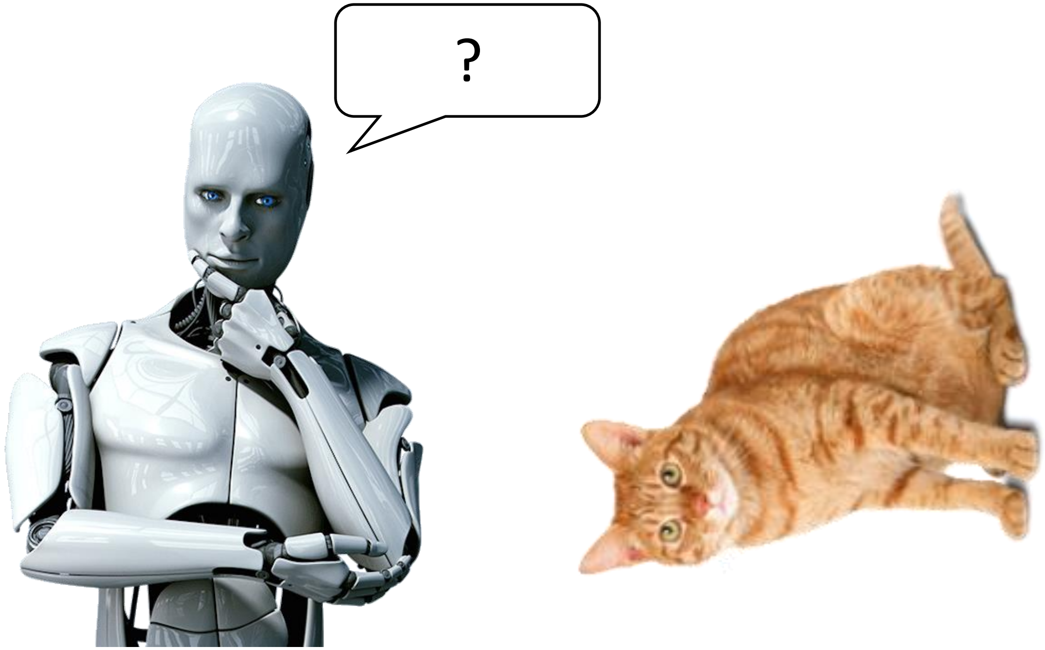


1. Learning with 3d rotations, a hitchhiker's guide to SO(3). (Rene Geist et al., ICML 2024)
2. On the continuity of rotation representations in neural networks. (Zhou et al., CVPR 2019)
3. Learning rotations. (Pepe et al., Applied mathematics 2024)

*Wigner-D matrix: rotation of spherical harmonics

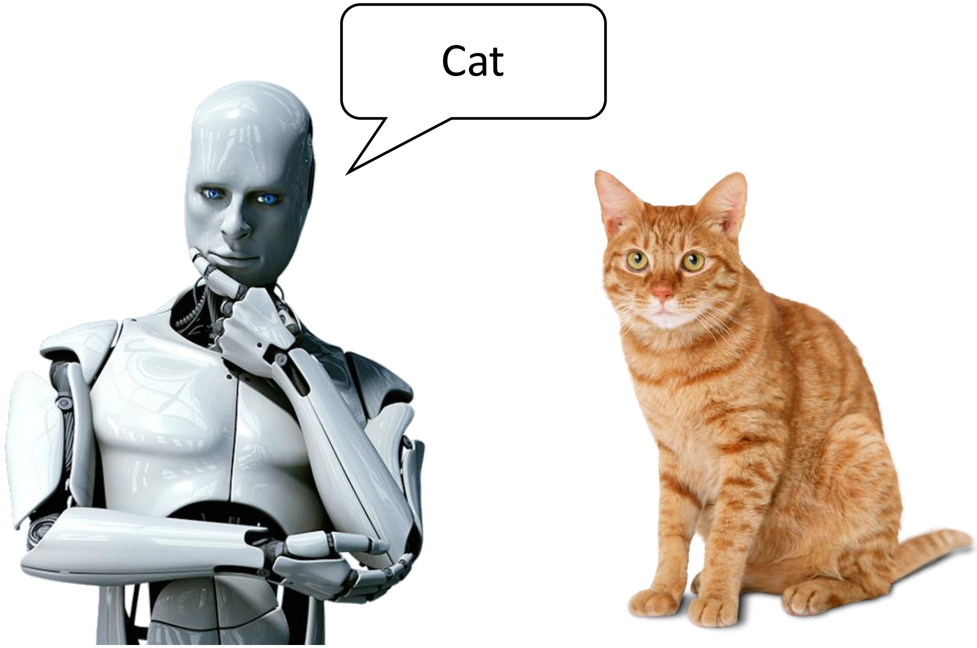
Preliminary: Why Equivariance?

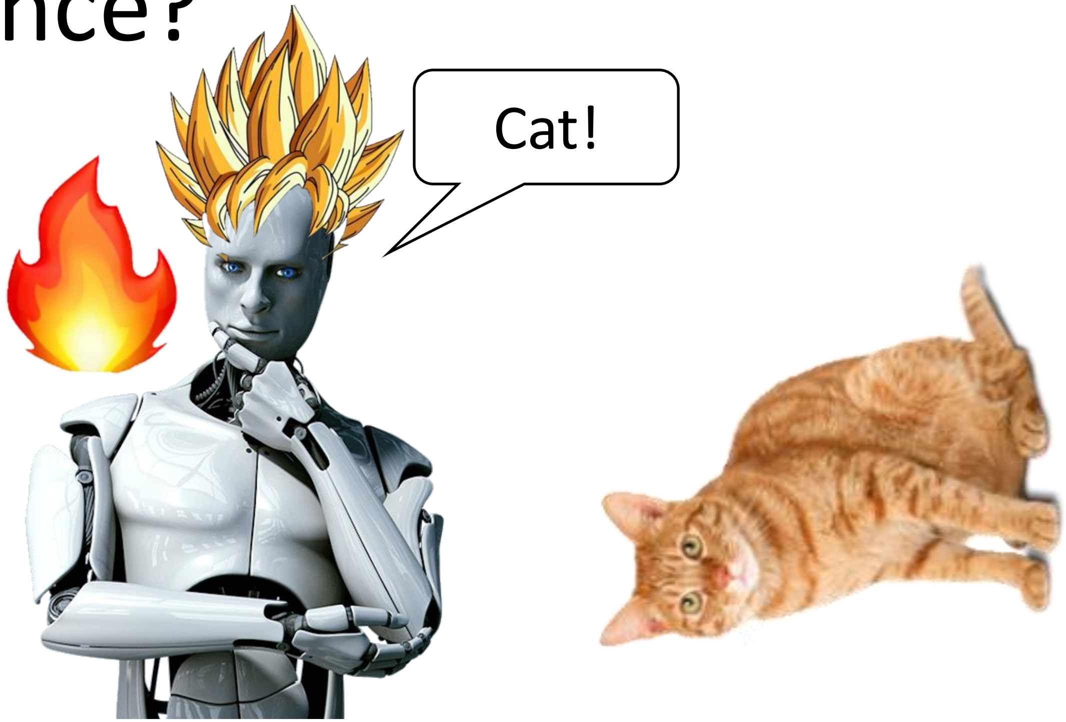




- Existing AI systems are not generalizable to **unseen spatial context**.
    - It depends on large-scale training data, with strong data augmentation.
    - Memorize the samples in training time!

Preliminary: Why Equivariance?



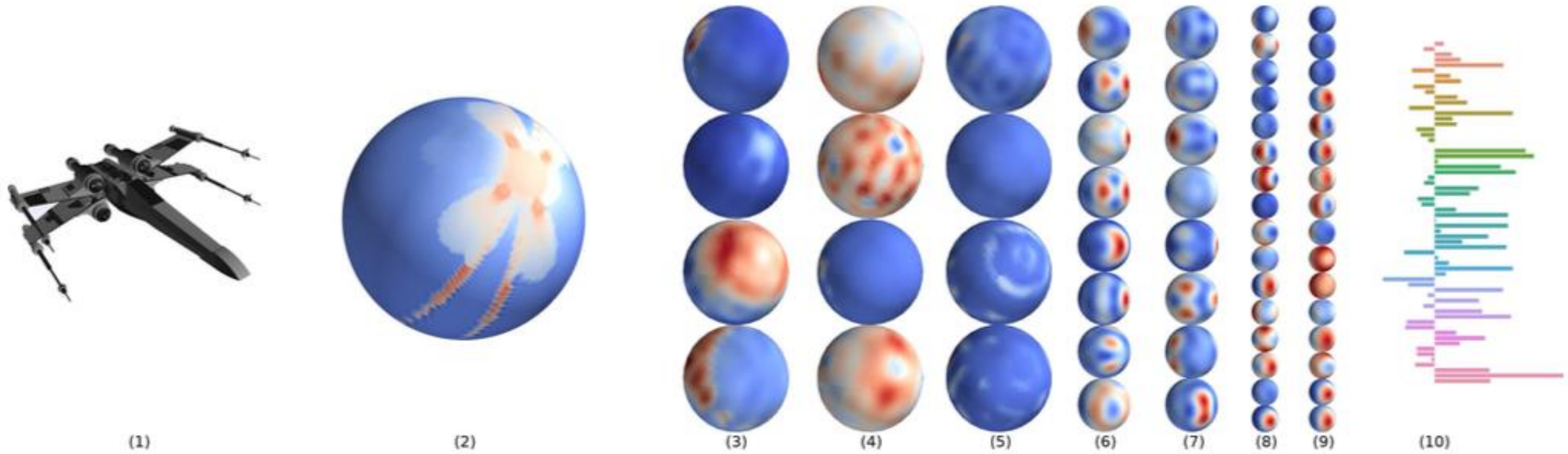


Equivariant Model

- Existing AI systems are not generalizable to **unseen spatial context**.
  - It depends on large-scale training data, with strong data augmentation.
  - Memorize the samples in training time!
- Equivariant networks can **generalize unseen geometric transformations**!
  - It can reduce the # of training data, without data augmentation.

Preliminary: Spherical CNNs for SO(3)-Equivariance

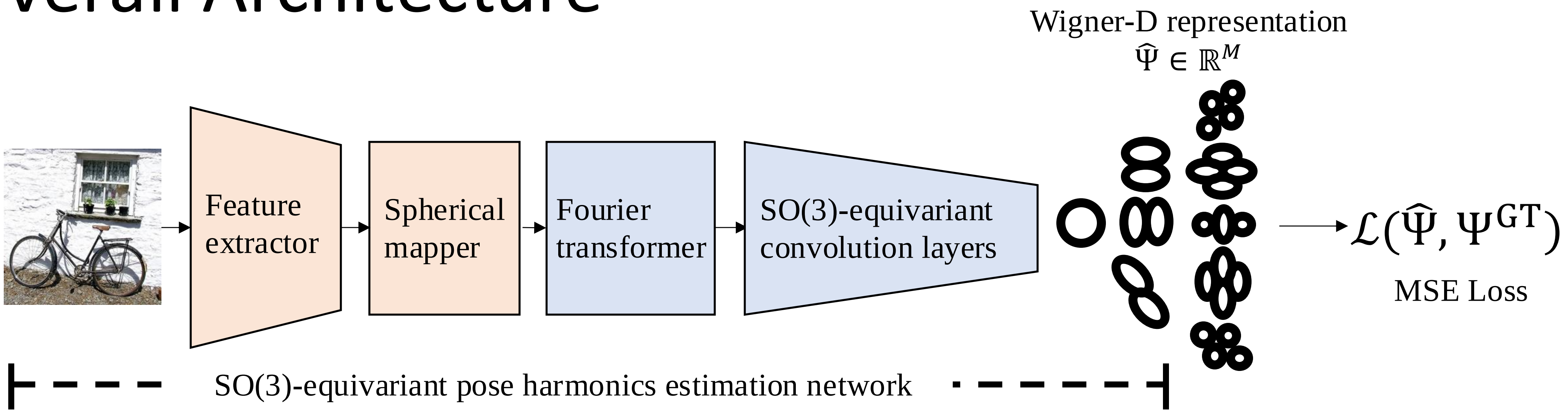
- Spherical convolution neural networks [1, 2, 3, 4, 5, 6]
  - Compute cross-correlation on **frequency domain for efficiency**, by FFT
  - Guarantee SO(3)-equivariance: reducing the # of training data, **improving data efficiency**
  - Obtain reliable spherical representation in pose space



Example of SO(3) equivariant CNNs[2]

1. Spherical CNNs (Cohen et al., ICLR 2018)
2. Learning SO(3) Equivariant Representations with Spherical CNNs (Esteve et al., ECCV 2018)
3. DeepSphere: a graph-based spherical CNN (Defferrard et al., ICLR 2020)
4. Equivariant Networks for Pixelized Spheres (Shakerinava and Ravanbakhsh, ICML 2021)
5. Clebsch-Gordan Nets: A Fully Fourier Space Spherical CNN (Kondor et al., NIPS 2018)
6. Efficient Generalized Spherical CNNs (Cobb et al., ICLR 2021)

Overall Architecture

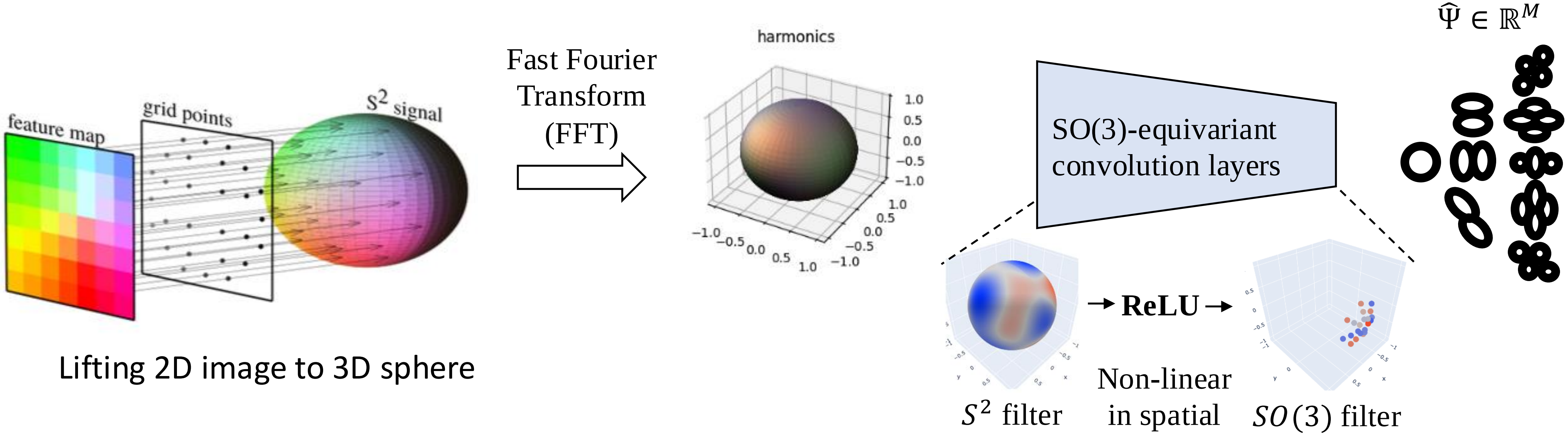


- SO(3)-Equivariant Pose Harmonics Estimation Network[1]
  - Feature extractor: ResNet with ImageNet pretrained weights
  - Spherical mapper: Orthographic projection from 2D image to (hemi-)sphere
  - Fourier transformer: Fast Fourier Transform
  - SO(3)-equivariant convolution layers: Spherical CNNs
- Loss function
  - Frequency-domain specific MSE Loss

$$\mathcal{L}(\hat{\Psi}, \Psi^{\mathrm{GT}}) = \sum_{l=0}^{L} \sum_{m=-l}^{l} w_l (\hat{\Psi}_{lm} - \Psi_{lm}^{\mathrm{GT}})^2,$$

1. Image to Sphere: Learning Equivariant Features for Efficient Pose Prediction (Klee et al., ICLR 2023)

Spherical Mapper, Spherical Convolution for SO(3)-Equivariance



Spherical mapper by orthographic projection[1]

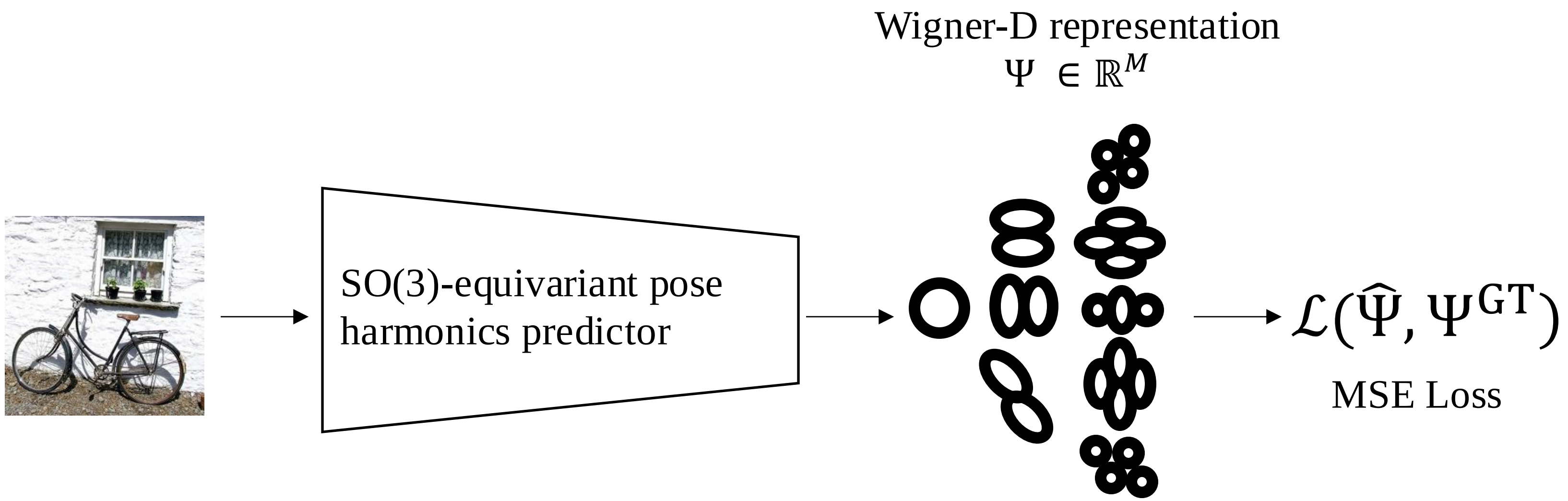
Spherical convolution for SO(3)-equivariance

- Spherical mapper: **orthographic projection** to lift 2D image feature on sphere
- **SO(3)-equivariant** convolution layers: S^2-conv and $SO(3)$-conv in Spherical CNNs[2]
- The output Ψ is **Wigner-D matrix coefficients,** which represents 3D rotation in frequency domain.

1. Image to sphere: Learning equivariant features for efficient pose prediction (Klee et al., ICLR 2023)
2. Spherical CNNs (Cohen et al., ICLR 2018)

*Figure courtesy: Image2Sphere paper, github visualization: https://github.com/dmklee/image2sphere

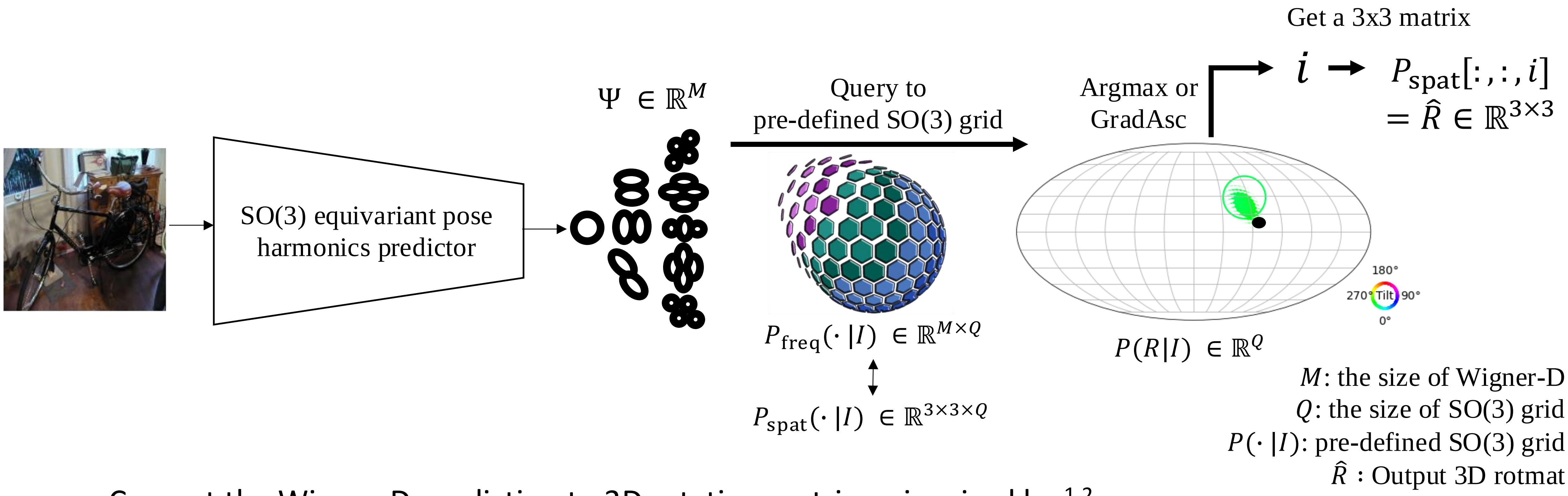
Loss Function



$$\mathcal{L}(\hat{\Psi}, \Psi^{\mathrm{GT}}) = \sum_{l=0}^{L} \sum_{m=-l}^{l} w_l (\hat{\Psi}_{lm} - \Psi_{lm}^{\mathrm{GT}})^2$$

- Frequency-domain regression loss
  - The output Ψ indicates **specific object orientations in an image**.
  - We calculate the MSE by **normalizing each harmonic frequency level l** with w_l.
  - This regression loss allows for **continuous output values**.
  - More **precise predictions of unambiguous object poses** than previous discretization methods.

Inference: Converting Wigner-D coefficients to 3D rotation matrix



- Convert the Wigner-D prediction to 3D rotation matrices, inspired by [1,2].
  - Producing **non-parametric probability distribution** by computing similarity between Ψ and P_{freq} .
  - With pre-defined SO(3) mapping grid ($P_{\text{freq}}(\cdot|I) \rightarrow P_{\text{spat}}(\cdot|I)$).
- Possible to modeling **uncertainty from pose ambiguity and 3D symmetry** with distribution loss[1,2].

1. Implicit-PDF: Non-Parametric Representation of Probability Distributions on the Rotation Manifold (Murphy et al., ICML 2021)
2. Image to sphere: Learning equivariant features for efficient pose prediction (Klee et al., ICLR 2023)

Results: Comparison with Existing Methods

Method	Acc@15	Acc@30	Rot Err. (Median)
Zhou *et al.*, (CVPR 2019)	0.251	0.504	41.1
Breiger (3DV 2021)	0.257	0.515	39.9
Liao *et al.* (CVPR 2019)	0.357	0.583	36.5
Deng *et al.*, (ECCV 2020, IJCV 2022)	0.562	0.694	32.6
Prokudin *et al.* (ECCV 2018)	0.456	0.528	49.3
Mohlin *et al.* (NeurIPS 2020)	0.693	0.757	17.1
Murphy *et al.*, (ICML 2021)	0.719	0.735	21.5
Yin *et al.*, (CVPR 2022)	-	0.751	16.1
Yin *et al.*, (ICLR 2023)	0.742	0.772	12.7
Klee *et al.*, (ICLR 2023)	0.728	0.736	15.7
Liu *et al.*, (Uni) (CVPR 2023)	0.76	0.774	14.6
Liu *et al.*, (Fisher) (CVPR 2023)	0.744	0.768	12.2
Howell *et al.*, (NeurIPS 2023)	-	-	17.8
ours (ResNet-50)	0.759	0.767	15.1
ours (ResNet-101)	**0.773**	**0.780**	**11.9**

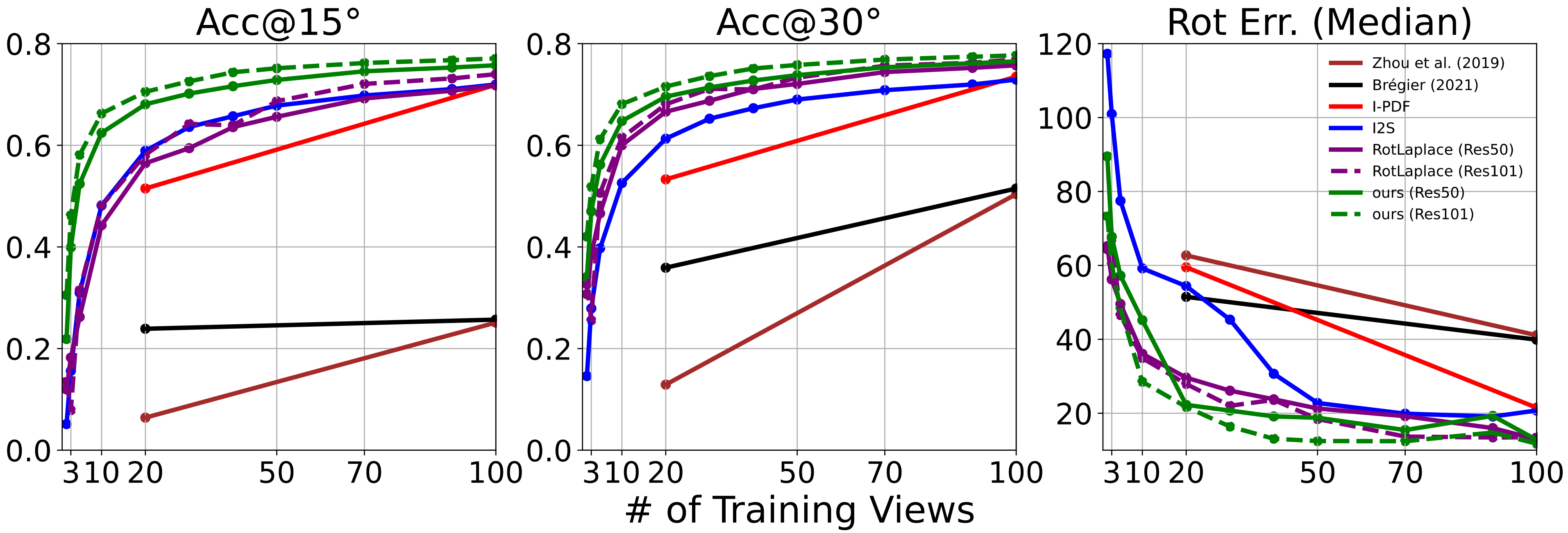
Results on **ModelNet10-SO(3).**

Method	Acc@30	Rot Err. (Median)
Zhou *et al.*, (CVPR 2019)	-	19.2°
Breiger (3DV 2021)	-	20.0°
Liao *et al.* (CVPR 2019)	0.819	13.0°
Prokudin *et al.* (ECCV 2018)	0.838	12.2°
Mohlin *et al.* (NeurIPS 2020)	0.825	11.5°
Tulsiani & Malik (CVPR 2015)	-	13.6°
Mahendran *et al.* (BMVC 2018)	-	10.1°
Murphy *et al.*, (ICML 2021)	0.837	10.3°
Yin *et al.*, (ICLR 2023)	-	9.4°
Klee *et al.*, (ICLR 2023)	0.872	9.8°
Liu *et al.*, (Uni) (CVPR 2023)	0.827	10.2°
Liu *et al.*, (Fisher) (CVPR 2023)	0.863	9.9°
Howell *et al.*, (NeurIPS 2023)	-	9.2°
Ours	**0.897**	**8.9°**

Results on **PASCAL3D+.**

- Our model achieve **state-of-the-art performance** on standard single-view SO(3) pose estimation benchmarks.

Results: Few-shot Training for Sampling efficiency



- Our model consistently obtains best scores **by reducing the # of training data**.
  - Our SO(3)-equivariant network contributes **data sampling efficiency.**

- On the Continuity of Rotation Representations in Neural Networks (Zhou et al., CVPR 2019)
- Deep Regression on Manifolds: A 3D Rotation Case Study (Brégier, 3DV 2021)
- Implicit-PDF: Non-Parametric Representation of Probability Distributions on the Rotation Manifold (Murphy et al., ICML 2021)
- Image to Sphere: Learning Equivariant Features for Efficient Pose Prediction (Klee et al., ICLR 2023)
- RotationLaplace: A Laplace-inspired Distribution on SO(3) for Probabilistic Rotation Estimation (Yin et al., ICLR 2023)

Results: Validation of Design Choices

Method	Acc@15	Acc@30	Rot Err.
Wigner (ours)	**0.6807**	**0.6956**	**22.27°**
Euler	0.0010	0.0072	132.56°
Quaternion	0.0510	0.1629	75.95°
Axis-Angle	0.0124	0.0815	88.66°
Rotmat	0.3909	0.5682	37.54°
w.o equivConv	0.1056	0.1308	149.25°

Comparison of **different rotation parametrizations & w/o SO(3)-equivariant convolution.**

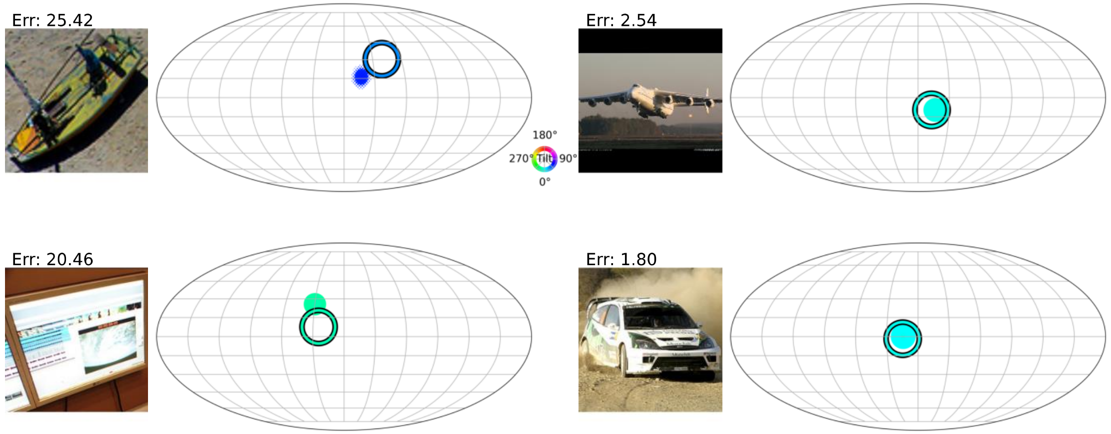
Loss Function	Acc@15°	Acc@30°	Rot Err.
MSE Loss	**0.6807**	**0.6956**	22.27°
L1 loss	0.6796	0.6933	22.12°
Huber loss	0.6710	0.6873	**19.26°**
Cosine loss	0.4414	0.4978	64.29°
Geodesic loss	0.0009	0.0071	132.65°

Comparison of **different loss functions.**

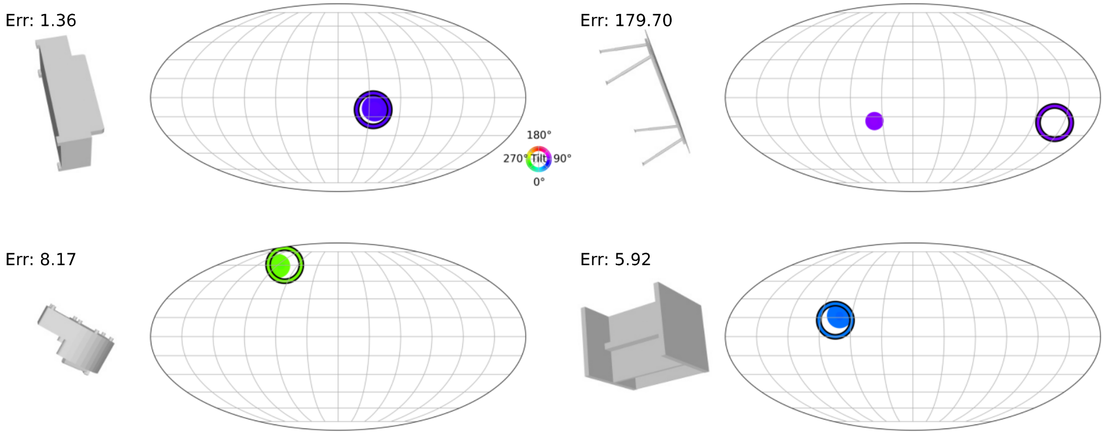
- (Left) Rotation parametrization in the frequency domain facilitates **accurate 3D rotation prediction**.
- (Left) SO(3)-equivariant modules using spherical CNNs are critical for **effective generalization**.
- (Right) MSE loss demonstrates optimal performance with **a simple yet effective approach**.

Results: Pose Visualization

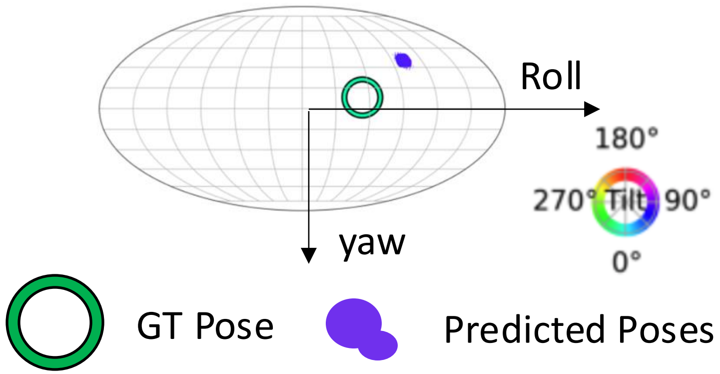
Results of Pascal3D+.



Results of ModelNet10-SO(3).



Visualization of a distribution over SO(3)[1].



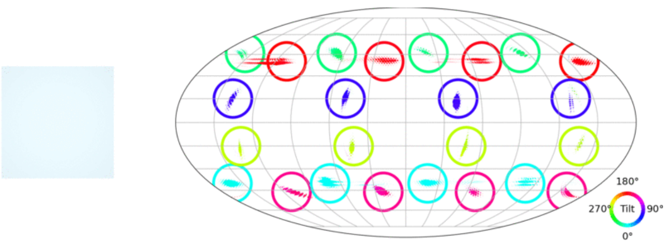
- Our method captures **a sharp modality of the pose distribution** in well-defined poses.
- Our method encounters **challenges with pose ambiguity caused by object symmetry**.

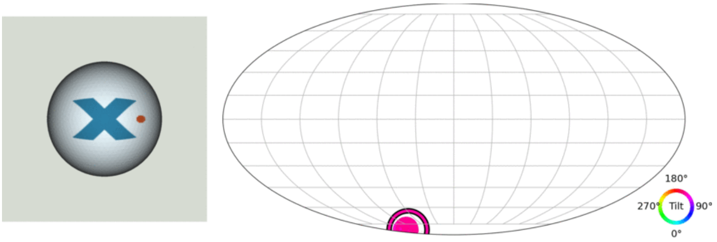
1. Implicit-PDF: Non-Parametric Representation of Probability Distributions on the Rotation Manifold (Murphy et al., ICML 2021)

Results: Joint Training for Symmetric Objects

Results on SYMSOL I and II: with distribution loss $\mathcal{L}_{\text{dist}}$[1]

	SYMSOL I						SYMSOL II			
	avg	cone	cyl	tet	cube	icosa	avg	sphereX	cylO	tetX
$\mathcal{L}_{\text{wigner}}$	2.54	2.42	2.68	2.93	2.67	1.99	-8.88	4.51	-7.64	-23.52
$\mathcal{L}_{\text{dist}}$ [35]	3.41	3.75	3.10	4.78	3.27	2.15	4.84	3.74	5.18	5.61
$\mathcal{L}_{\text{wigner}}+\mathcal{L}_{\text{dist}}$	**4.11**	**4.43**	**3.76**	**5.59**	**3.93**	**2.85**	**6.20**	**6.66**	**5.85**	**6.11**

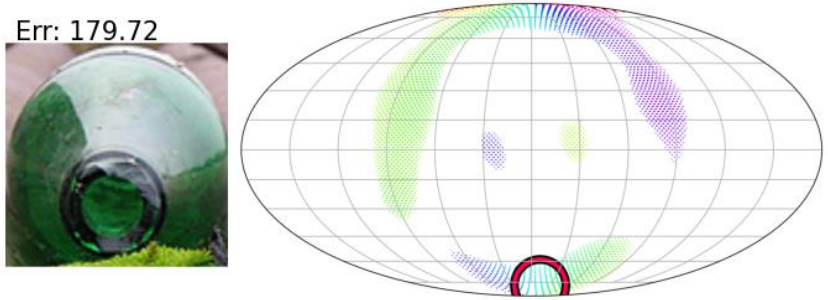




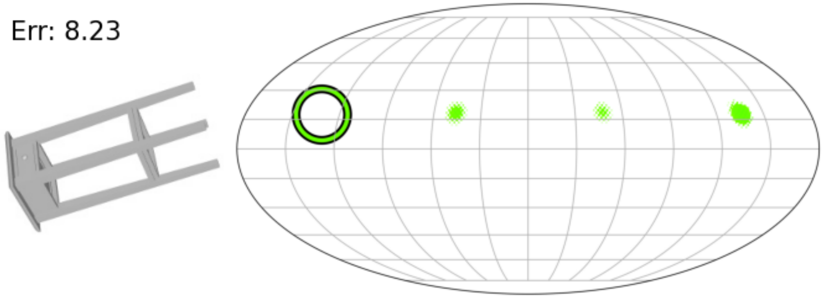
Handle 3D symmetry cases

With log-likelihood distribution loss[1]

Err: 179.72



Err: 8.23



1. Implicit-PDF: Non-Parametric Representation of Probability Distributions on the Rotation Manifold (Murphy et al., ICML 2021)

Conclusion

- Single-image pose estimation by 3D equivariant pose harmonics estimator
  - Problem 1: the **discontinuities and singularities in spatial domain**
    - Predict **Wigner-D coefficients in frequency-domain** for a 3D rotation
  - Problem 2: require **large-scale training data**
    - SO(3)-equivariant representations **for data sampling efficiency**
  - Problem 3: **pose ambiguity and 3D symmetry** in the world
    - Probabilistic prediction by **joint training with log-likelihood distribution loss**

- Future work
  - More effective rotation representation for 3D space
  - Improving computational efficiency
  - Out-of-distribution domain generalization

Poster Session 4

Thursday (12nd, Dec), 4:30pm - 7:30pm

See you soon!

NEURAL INFORMATION PROCESSING SYSTEMS

Thank you!

Project Page **Paper** **Code**

POSTECH

POHANG UNIVERSITY OF SCIENCE AND TECHNOLOGY

ComputerVision Lab.