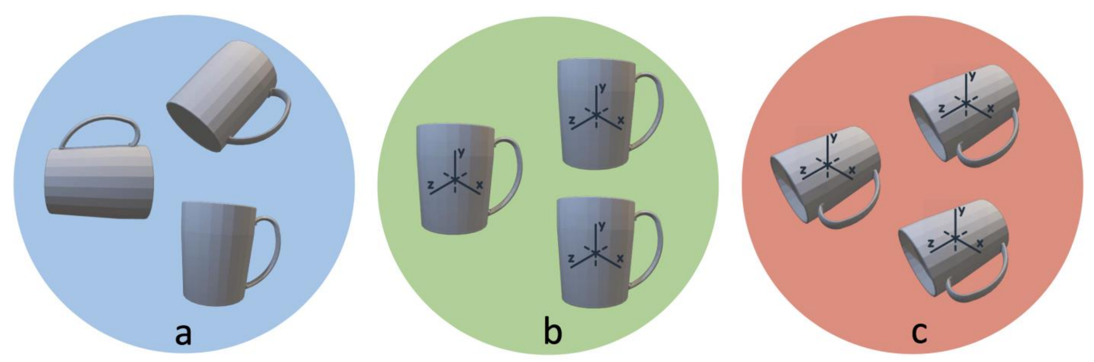
Learning to Orient Surfaces by Self-Supervised Spherical CNNs

Riccardo Spezialetti[1], Federico Stella[1], Marlon Marcon[2], Luciano Silva[3], Samuele Salti[1], Luigi Di Stefano[1]

[1]University of Bologna, Italy [2]Federal University of Technology - Paraná, Dois Vizinhos, Brazil [3]Federal University of Paraná, Curitiba, Brazil

Problem motivation and related work

Objects in the real world can appear with different orientations (a), and humans learn to neutralize such orientations for recognition and interaction purposes (b). Similarly, robotic and computer vision systems require orientation neutralization in many important tasks: grasping, navigation, surface matching, augmented reality, shape classification and more.



These systems pursue rotation-invariance in two ways:

- *By* ***rotation-invariant operators***
- *By estimating* ***a canonical orientation****, not necessarily natural to humans (c)*

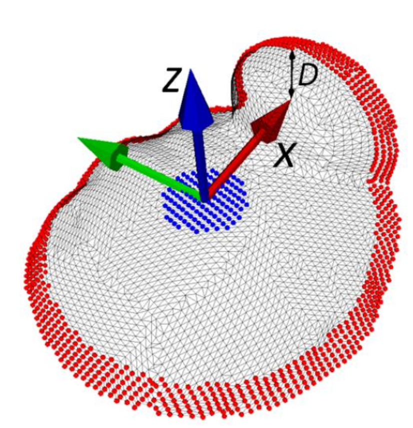
State of the art

Rotation-invariant operators:

- *PRIN [3], invariant spherical correlations*
- *Zhang et al. [4], low-level geometric features*

Canonical orientation:

- *SHOT [5], FLARE [6], TOLDI [7], GFrames [8], 3DSN[9], hand-crafted Local Reference Frames (LRFs) that perform at their best on specific datasets*
- *PointNets [1][2], limited generalization to unseen rotations*



Local Reference Frame, image from [6]

Open problems

- *Limited generalization to unseen rotations*
- *Limited generalization to unseen shapes or datasets*
- *Most LRFs are hand-crafted and work best on specific datasets*

Proposed solution - Compass

- *First approach to learn a canonical orientation*
- *Fully data-driven, no geometric assumptions or hand-crafted choices*
- *The key property of a canonical orientation is* ***equivariance to 3D rotations****. Compass achieves it by leveraging on Spherical CNNs [10] alongside a self-supervised training pipeline*

Spherical CNNs

Overview of Spherical CNNs (more details in [10]).

- **Spherical Signal**: *a continuous K-valued function* $f : S^2 \rightarrow \mathbb{R}^K$.
- **Rotation of Spherical Signals**: *the operator* L_R *rotates a function* f *by* $R \in \mathrm{SO}(3)$, *by composing its input with* R^{-1}, *i.e.* $[L_R f](x) = f(R^{-1}x)$.
- **Spherical Correlation**: *given a K-valued spherical signal* f *and a filter* ψ, $f, \psi : S^2 \rightarrow \mathbb{R}^K$:

$$[\psi \star f](R) = \langle L_R \psi, f \rangle = \int_{S^2} \sum_{k=1}^{K} \psi_k(R^{-1}x) f_k(x) dx. \tag{1}$$

Notice that the output is a signal on SO(3).

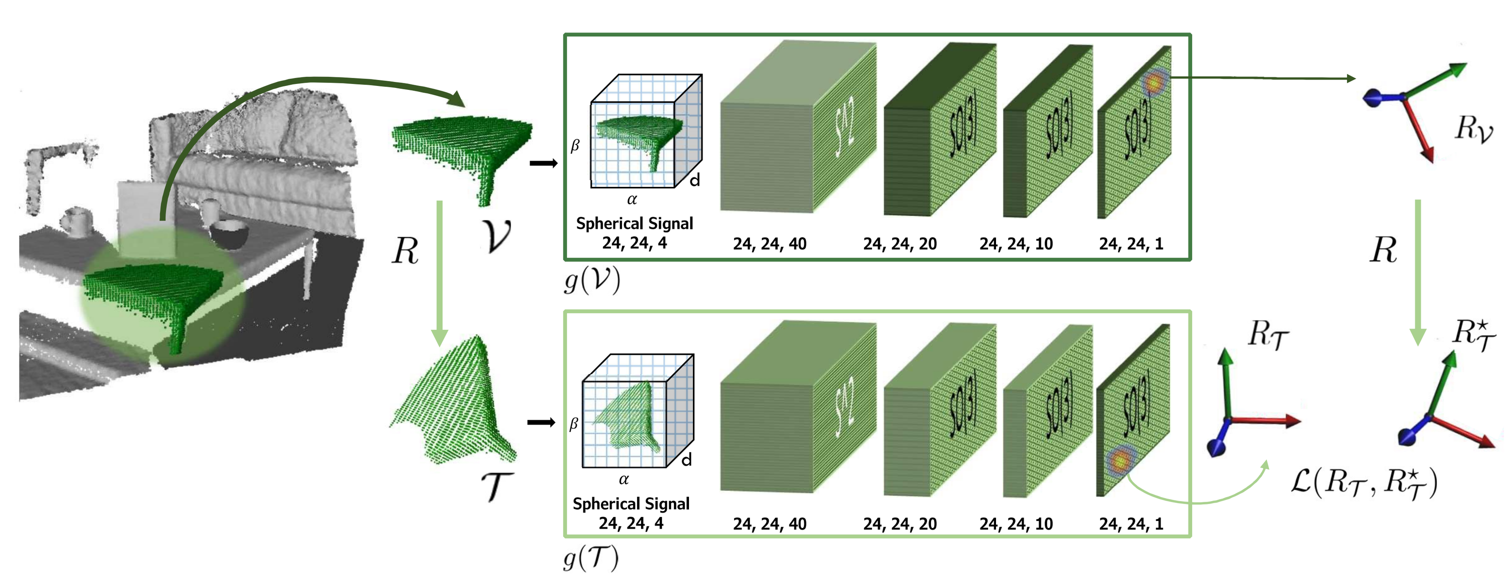
- **Rotation of SO(3) Signals**: *the operator* L_R *can be extended to work with an SO(3) signal* $h : \mathrm{SO}(3) \rightarrow \mathbb{R}^K : [L_R h](Q) = h(R^{-1}Q)$.
- **SO(3) Correlation**: *given a K-valued SO(3) signal* h *and a filter* ψ, $h, \psi : \mathrm{SO}(3) \rightarrow \mathbb{R}^K$:

$$[\psi * h](R) = \langle L_R \psi, h \rangle = \int_{\mathrm{SO}(3)} \sum_{k=1}^{K} \psi_k(R^{-1}Q) h_k(Q) dQ. \tag{2}$$

Both correlations are ***equivariant to rotations****, as proven in [10]:*

$$[\psi \star [L_Q h]](R) = [L_Q[\psi \star h]](R). \tag{3}$$

Architecture



Loss function is the geodesic distance between rotations on the SO(3) manifold:

$$\mathcal{L}(R_{\mathcal{T}}, R^{\star}_{\mathcal{T}}) := \cos^{-1}\left(\frac{(tr({R_{\mathcal{T}}}^T R^{\star}_{\mathcal{T}}) - 1)}{2}\right). \tag{4}$$

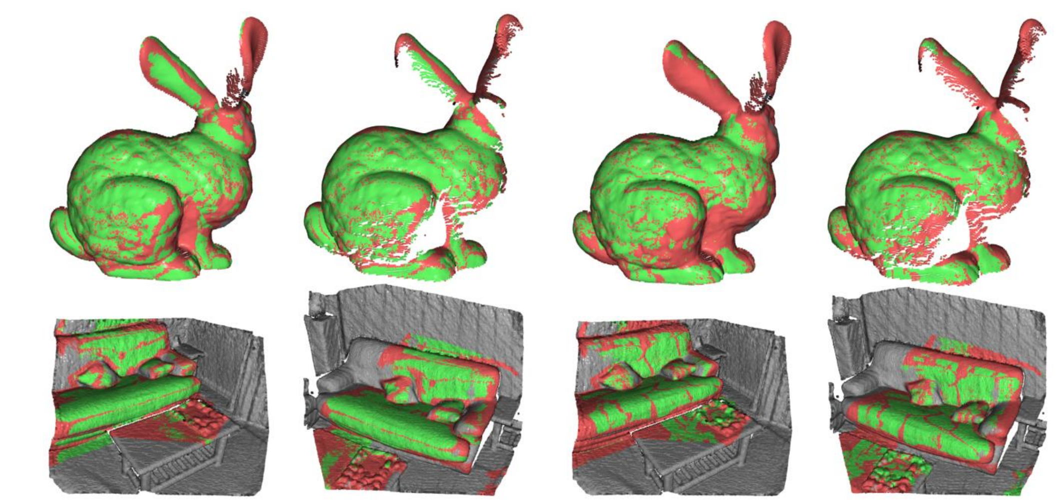
Reference frame selection from the last SO(3) layer:

$$C_R(\mathcal{V}) = \text{soft-argmax}(\tau\Phi(f_{\mathcal{V}})) = \sum_{i,j,k} \text{softmax}(\tau\Phi(f_{\mathcal{V}}))_{i,j,k}(i,j,k). \tag{5}$$

Applications and results

Surface patches

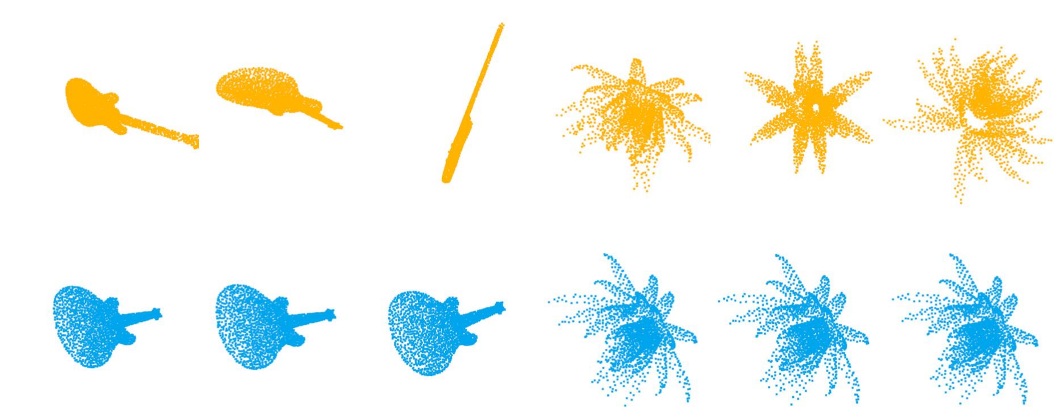
	LRF Repeatability ↑						
Dataset	SHOT	FLARE	TOLDI	3DSN	GFrames	Compass	Compass (Adapted)
3DMatch	0.212	0.360	0.215	0.220	n.a	**0.375**	n.a.
ETH	0.273	0.264	0.185	0.202	n.a	0.308	**0.317**
Stanford	0.132	0.241	0.197	0.173	0.256	0.361	**0.388**



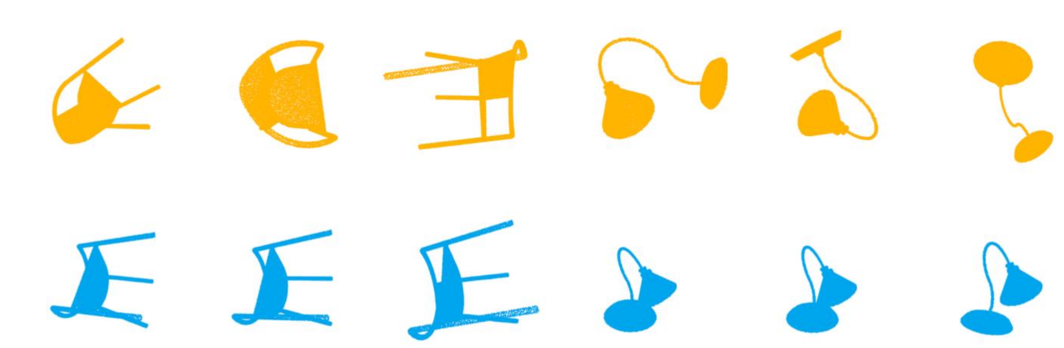
Compass *FLARE [6]*

Global shapes

	Classification Accuracy %							
	PointNet	PointNet++	Point2Seq	Spherical CNN	LDGCNN	SO-Net	PRIN	Compass + PointNet
NR	88.45	89.82	92.60	81.73	92.91	**94.44**	80.13	80.51
AR	12.47	21.35	10.53	55.62	17.82	9.64	70.35	**72.20**



ModelNet40 [11]



ShapeNet [12] (transfer learning)

References

[1] C. R. Qi et al. (2017). "Pointnet: Deep learning on point sets for 3d classification and segmentation". In: CVPR.

[2] C. R. Qi et al. (2017). "Pointnet++: Deep hierarchical feature learning on point sets in a metric space". In: NeurIPS.

[3] Y. You et al. (2019). "Prin: Pointwise rotation invariant network". In: AAAI.

[4] K. Zhang et al. (2019). "Linked dynamic graph cnn: Learning on point cloud via linking hierarchical features". In: arXiv.

[5] S. Salti et al. (2014). "Shot: Unique signatures of histograms for surface and texture description". In: CVIU.

[6] A. Petrelli et al. (2012). "A repeatable and efficient canonical reference for surface matching". In: 3DIMPVT.

[7] J. Yang et al. (2017). "Toldi: An effective and robust approach for 3d local shape description". In: Patt. Rec.

[8] S. Melzi et al. (2019). "Gframes: Gradient-based local reference frame for 3d shape matching". In: CVPR.

[9] Gojcic, Zan, et al. (2019). "The perfect match: 3d point cloud matching with smoothed densities." In: CVPR.

[10] T. S. Cohen et al. (2018). "Spherical CNNs". In: ICLR.

[11] Wu, Zhirong, et al. (2015). "3d shapenets: A deep representation for volumetric shapes." In: CVPR.

[12] Chang, Angel X., et al. (2015). "Shapenet: An information-rich 3d model repository." In: *arXiv*.