Canonical Fields: Self-Supervised Learning of Pose-Canonicalized Neural Fields

Anonymous CVPR submission

Paper ID 9580

A. Preprocessing Density Field

Before querying NeRF for occupancy, we do not have the extremities of the object. We begin by querying a uniform grid within the unit cube to obtain a 32^3 grid density field. The density values are normalized to range $[0, 1]$ (see page 3 in the main paper). We perform an initial K-means clustering with $K = 2$ to separate the foreground and background clusters. Note that this clustering is on the entire scene and different from the clustering we perform while computing the losses. The clustering that we perform while computing the loss is within the object bounding box and not for the entire scene. The mean of query coordinate locations in the foreground cluster gives us the center of the object **c**. We then obtain the extremities from the foreground cluster and find its maximal diagonal length ($\boldsymbol{l}$). We re-sample the region within the bounding cube (at center $\boldsymbol{c}$ side length $\boldsymbol{l}$) and use it as input to CaFi-Net.

B. Equivariance Properties of CaFi-Net

B.1. Averaging Equivariant Signals is Equivariant

A tensor field of type ℓ (ℓ-field) is a map $f : \mathbb{R}^3 \to \mathbb{R}^{2\ell+1}$. We have an action of $\mathrm{SO}(3)$ on any ℓ-field f given by $(R.f)(x) := D^\ell(R)f(R^{-1}x)$ for any rotation $R \in \mathrm{SO}(3)$ and $x \in \mathbb{R}^3$ where $D^\ell(R) \in \mathrm{SO}(2\ell + 1)$ is the Wigner matrix of type ℓ. We say that transformation F transforming tensor fields of type p into tensor fields of type q if it commutes with the action of $\mathrm{SO}(3)$, i.e. for all type p-field f and rotation $R \in \mathrm{SO}(3)$ we have $F(R.f) = R.F(f)$. We will call such transformations (p, q)-equivariant transformations.

Lemma 1 *For any $r > 0$ the local average operator mean_r defined over p-fields by:*

$$\mathrm{mean}_r(f)(x) := \int_{B(x,r)} f(y)dy$$

where $B(x, r) \subset \mathbb{R}^3$ is the open ball of radius r centered at $x \in \mathbb{R}^3$ is a (p, p)-equivariant transform of fields.

Proof: Let $x \in \mathbb{R}^3$ and $R \in \mathrm{SO}(3)$ we have:

$$\begin{aligned}\mathrm{mean}_r(R.f)(x) &= \int_{B(x,r)} (R.f)(y)dy \\ &= \int_{B(x,r)} D^p(R)f(R^{-1}y)dy \\ &\underset{u=R^{-1}y}{=} D^p(R)\int_{R^{-1}B(x,r)} f(u)du \\ &= D^p(R)\int_{B(R^{-1}x,r)} f(u)du \\ &= (R.\mathrm{mean}_r(f))(x)\end{aligned}$$

B.2. Locally Averaged Density Weighted Equivariant Vectors are Equivariant

Lemma 2 *Scaling equivariant field with density is a (1, 1)-equivariant transformation.*

Proof: Let σ be a type-0 density field and f be a type-1 field at the corresponding location, weighing f by the average of σ is:

$$\mathrm{mean}_r(\sigma)(x) \cdot f(x) := \left(\int_{B(x,r)} \sigma(y)dy\right) f(x)$$

$$\begin{aligned}&\mathrm{mean}_r(R \cdot \sigma)(x) \cdot (R \cdot f)(x) \\ &= \left(\int_{B(x,r)} (R \cdot \sigma)(y)dy\right) D^1(R)(f)(R^{-1}x) \\ &= \left(\int_{B(x,r)} (\sigma)(R^{-1}y)dy\right) D^1(R)(f)(R^{-1}x) \\ &= \left(I \cdot \int_{R^{-1}B(x,r)} (\sigma)(u)du\right) D^1(R)(f)(R^{-1}x) \\ &= D^1(R) \cdot \mathrm{mean}_r(\sigma)(x) \cdot (f)(R^{-1}x)\end{aligned}$$

Thus, $\mathrm{mean}_r(R \cdot \sigma)(x) \cdot f(x) = D^1(R) \cdot \mathrm{mean}_r(\sigma)(x) \cdot f(R^{-1}x)$ proving the result that scaling equivariant features by average density do not break the equivariance property.

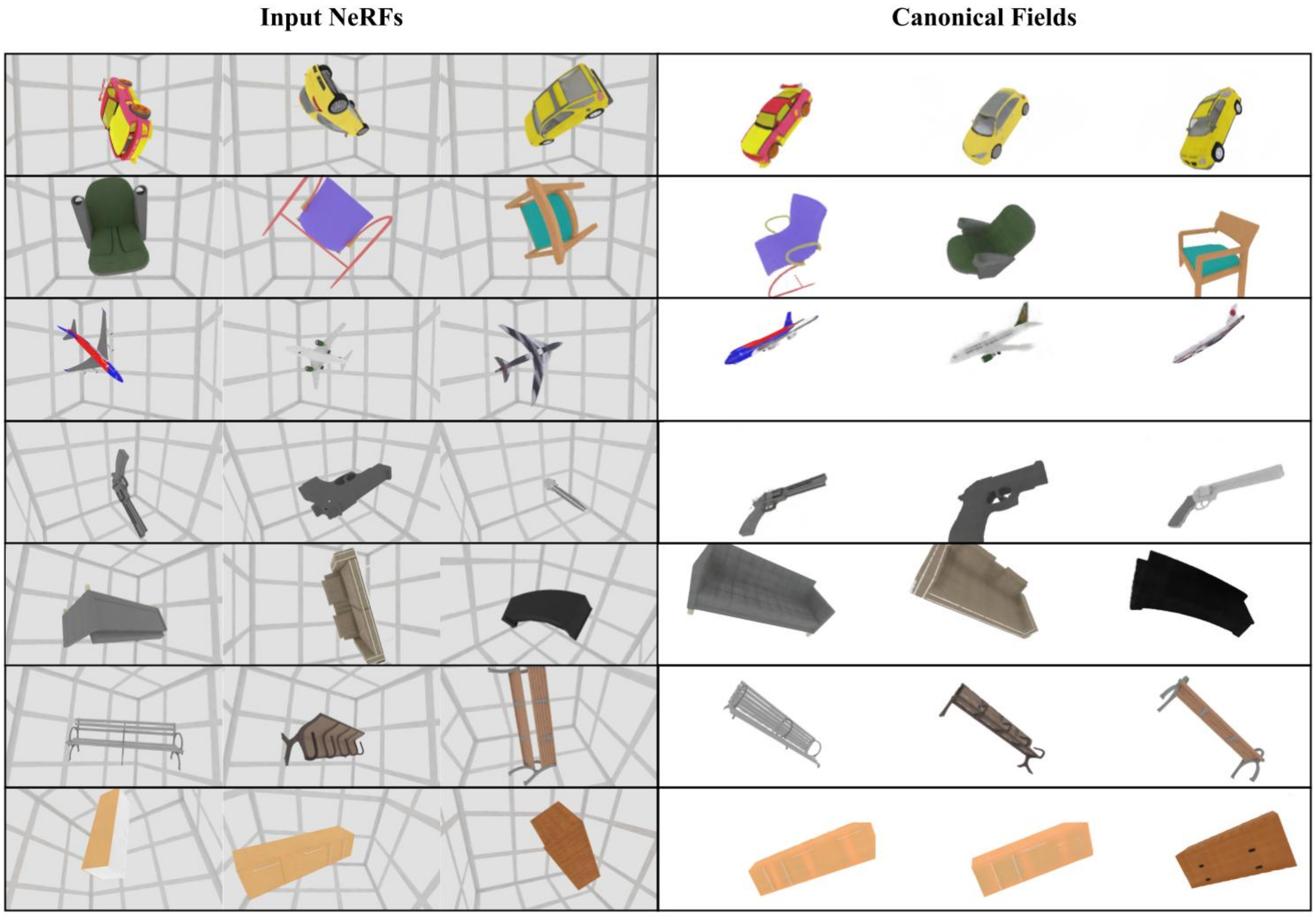

Figure 1. **CaFi-Net** qualitative canonicalization results for 6 categories (see following pages for more results). We omit the background in the canonical rendering for easier view.

B.3. Gradient is Type-1 Equivariant

Lemma 3 *The gradient operator is a* $(0, 1)$*-equivariant transformation.*

Proof: Let f be a 0-type field, by the chain rule of differentiation, for any $x, h \in \mathbb{R}^3$ and $R \in \mathrm{SO}(3)$ we have

$$\begin{aligned}
\langle \nabla(R.f), h \rangle &= D_x(R.f)(h) \\
&= D_x f \circ R^{-1}(h) \\
&= D_{R^{-1}x} f \circ D_x R^{-1}(h) \\
&= \langle \nabla_{R^{-1}x} f, R^{-1} h \rangle \\
&= \langle R \nabla_{R^{-1}x} f, h \rangle \\
&= \langle R \nabla_{R^{-1}x} f, h \rangle
\end{aligned}$$

thus $\nabla(R.f) = R.\nabla f$ which concludes the proof.

C. Additional Training Details

To train CaFi-Net, we augment NeRF density fields with random rotations R_{rand} and canonicalize them at each training iteration. We can easily do this by sampling the NeRF model at location $R_{rand}^{-1}x$ instead of x in a differentiable manner using [1]. Our method is built over equivariant layers and non-linearities from [2]. Table 1 shows the

Equivariant Convolution Non-Linearities
InverseSphericalHarmonicTransform(sphere_samples=64)
BatchNorm(momentum=0.75)
ReLU
MLP(F_{in}, F_{out})
ForwardSphericalHarmonicTransform(sphere_samples=64)

Table 1. **Equivariant Convolution Non-Linearities** - We use the result in [2] and apply non-linearities after performing an inverse spherical harmonic transform to avoid breaking the equivariance of the each layer. The *sphere_samples* is the sphere sampling resolution to perform the Spherical Harmonic Transform and F_in, F_out are the input and output feature dimensions respectively.

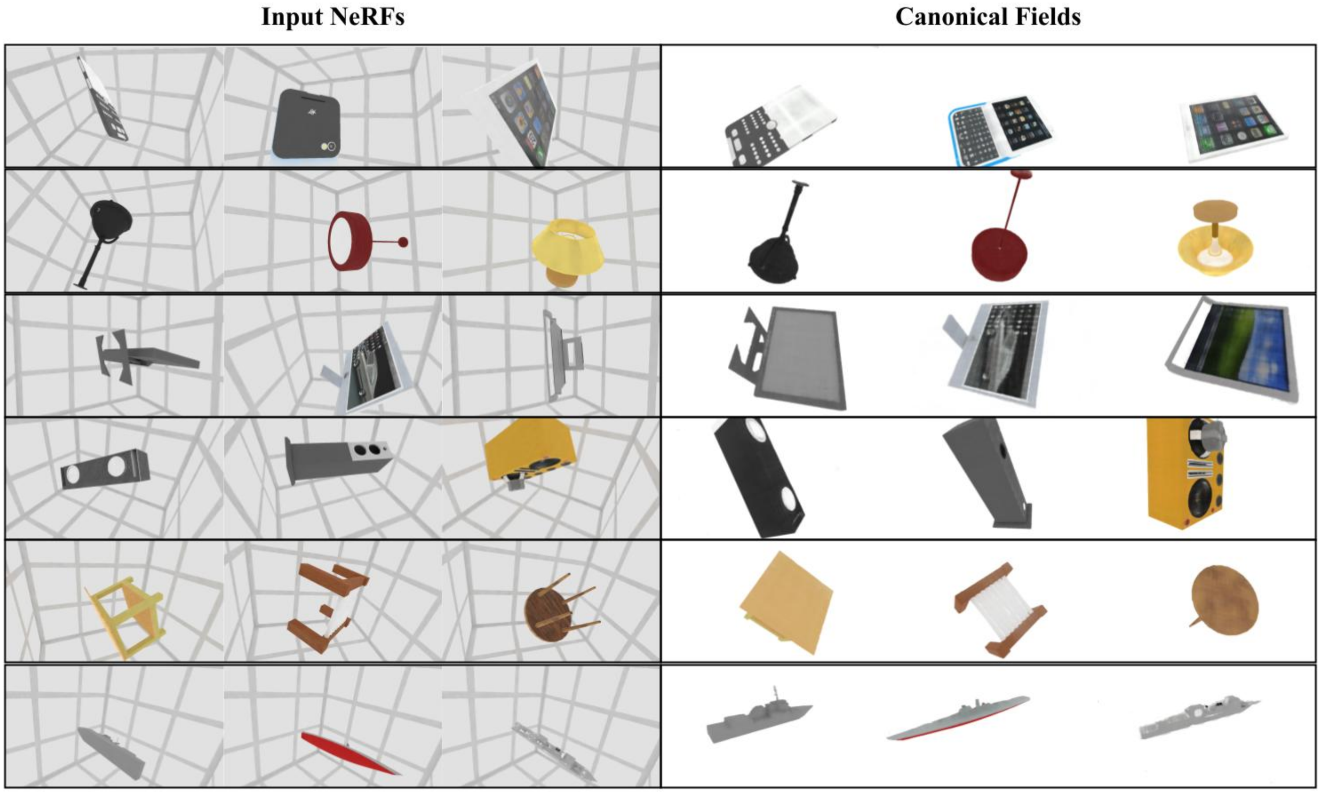

Figure 2. **CaFi-Net** qualitative canonicalization results for the remaining 7 categories. We omit the background in the canonical rendering for easier view.

non-linearities and learning layers that are used after each equivariant convolution described in the main manuscript.

D. Qualitative Results

We illustrate qualitative results by rendering objects in the canonical frame for all the 13 categories in Figure 1 and Figure 2. Here, we fix a camera position and viewing direction in the canonical frame and render all objects from the same camera. We have masked out the background by reducing the far-point for better view.

References

[1] Max Jaderberg, Karen Simonyan, Andrew Zisserman, and koray kavukcuoglu. Spatial transformer networks. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015. 2

[2] Adrien Poulenard and Leonidas J Guibas. A functional approach to rotation equivariant non-linearities for tensor field networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13174–13183, 2021. 2