Supplementary Material for Correspondence-Free Non-Rigid Point Set Registration Using Unsupervised Clustering Analysis

Mingyang Zhao[1] Jingen Jiang[2] Lei Ma[3] Shiqing Xin[2] Gaofeng Meng[1,4,5] Dong-Ming Yan[4,5]
[1]CAIR, HKISI, CAS [2]Shandong University [3]Peking University [4]MAIS, CASIA [5]UCAS

Abstract

This supplementary material contains additional content that supports our paper, including theoretical foundations, proofs, and experimental studies. Specifically, in Section 1, we provide a concise overview of the Nyström method to highlight its random approximation scheme. Theoretical proofs of the approximation bound for our Laplacian kernel are presented in Section 2. Furthermore, in Section 3, we empirically explore the advantages of our newly defined function, complementing the theoretical analysis. Finally, in Section 4, we present more qualitative comparison results and demonstrate the versatility of our method on shape transfer.

1. Nyström Low-Rank Approximation

The Nyström method is introduced by [2] to improve the efficiency of kernel machines. It was initially developed to address the integral equation given by:

$$\int p(y)K(x,y)\phi_i(y)dy = \lambda_i\phi_i(x), \tag{1}$$

where $p(y)$ represents the probabilistic density function, $K(x,y)$ is the kernel function, and ϕ_i and λ_i denote the unknown eigenfunctions and eigenvalues, separately. The integral equation can be approximated by taking the empirical average:

$$\frac{1}{q}\sum_{j=1}^{q} K(x,x_j)\phi_i(x_j) \cong \lambda_i\phi_i(x), \tag{2}$$

where $\{x_1, x_2, \cdots, x_q\}$ are sampled from $p(x)$. By substituting $x \in \{x_1, x_2, \cdots, x_q\}$ in Eq. (2), we get

$$\mathbf{KU} = \mathbf{U\Lambda}, \tag{3}$$

where $\mathbf{K} = K(x_i, x_j) \in \mathbb{R}^{q\times q}$ represents the kernel matrix, $\mathbf{U} \in \mathbb{R}^{q\times q}$ is the orthogonal matrix, and $\mathbf{\Lambda} \in \mathbb{R}^{q\times q}$ denotes the diagonal matrix containing the eigenvalues. Therefore, the eigenfunctions and eigenvalues in Eq. (1) can be estimated by [2]

$$\phi_i(x_j) \simeq \sqrt{q}U_{ji}, \quad \lambda_i \simeq \lambda_i^{(q)}/q. \tag{4}$$

This observation suggests that various subsets of size q provide approximate representations of the original eigenfunctions ϕ_i and eigenvalues λ_i [3]. Hence, the Nyström method involves randomly selecting a subset $Z = \{z_i\}_{i=1}^m$ from the input data $\{x_i\}_{i=1}^n$ in order to approximate the complete kernel matrix $\mathbf{K}$, as follows

$$\Phi_{\mathbf{K}} \simeq \sqrt{\frac{m}{n}}\mathbf{E}\mathbf{\Phi}_{\mathcal{Z}}{\mathbf{\Lambda}_{\mathcal{Z}}}^{-1}, \mathbf{\Lambda}_{\mathbf{K}} \simeq \frac{n}{m}\mathbf{\Lambda}_{\mathcal{Z}}. \tag{5}$$

Here $\mathbf{E} \in \mathbb{R}^{n\times m}$ and $\mathbf{\Phi}_{\mathcal{Z}}, \mathbf{\Lambda}_{\mathcal{Z}} \in \mathbb{R}^{m\times m}$ represent the corresponding eigenvector and eigenvalue matrices of $\mathbf{W} = K(z_i, z_j) \in \mathbb{R}^{m\times m}$, respectively. Ultimately, the low-rank approximation of the kernel matrix K can be expressed as follows [2]:

$$\begin{aligned}\mathbf{K} &\simeq \left(\sqrt{\frac{m}{n}}\mathbf{E}\mathbf{\Phi}_{\mathcal{Z}}{\mathbf{\Lambda}_{\mathcal{Z}}}^{-1}\right)\left(\frac{n}{m}\mathbf{\Lambda}_{\mathcal{Z}}\right)\left(\sqrt{\frac{m}{n}}\mathbf{E}\mathbf{\Phi}_{\mathcal{Z}}{\mathbf{\Lambda}_{\mathcal{Z}}}^{-1}\right)^T \\ &= \mathbf{E}\mathbf{W}^{-1}\mathbf{E}^T.\end{aligned} \tag{6}$$

As discussed above, the effectiveness of the Nyström method may be uncertain due to its random approximation scheme.

2. Theoretical Proof

In contrast to the conventional Nyström method mentioned above, our clustering-improved method provides a rigorous approximation bound specifically tailored to the Laplacian kernel utilized in our approach.

Lemma 1. *[3] Suppose* $(K(a,b) - K(c,d))^2 \leq C_{\mathcal{X}}^k\left(\|a-c\|^2 + \|b-d\|^2\right), \forall a,b,c,d$, *then the error of the general Nyström low-rank approximation* $\epsilon = \|\mathbf{L} - \mathbf{E}\mathbf{W}^{-1}\mathbf{E}^T\|_F$ *is bounded by*

$$\mathcal{E} \leq 4T\sqrt{mC_{\mathcal{X}}^k eT} + mC_{\mathcal{X}}^k Te\|\mathbf{W}^{-1}\|_F \tag{7}$$

where $T = \max_k |S_k|$, *and* $e = \sum_{i=}^{n} |x_i - z_{c(i)}\|^2$ *is the total quantization error of coding each sample* $x_i \in \mathcal{X}$ *with the closest landmark point* $z_j \in \mathcal{Z}$.

Proposition 1. *The low-rank approximation error* $\epsilon = \|\mathbf{L} - \mathbf{E}\mathbf{W}^{-1}\mathbf{E}^T\|_F$ *in terms of the Laplacian kernel is bounded by*

$$\epsilon \leq 4\sqrt{2}T^{3/2}\gamma\sqrt{C'q} + 2C'\gamma^2 Tq\|\mathbf{W}^{-1}\|_F, \quad (8)$$

where $\|\cdot\|_F$ *is the matrix Frobenious norm,* $T = \max_i |\mathbf{P}_i|$, $q = \sum_{j=1}^{C} \|\boldsymbol{y}_j - \boldsymbol{z}_{c'(j)}\|_2^2$ *is the clustering quantization error with* $c'(j) = \operatorname{argmin}_{i=1,\cdots,C'} \|\boldsymbol{y}_j - \boldsymbol{z}_i\|_2$, *and* γ *is the Laplacian kernel bandwidth defined in* $K(\boldsymbol{y}_i, \boldsymbol{y}_j) = \exp(-\gamma\|\boldsymbol{y}_i - \boldsymbol{y}_j\|_1)$.

Proof. Given four samples, $\boldsymbol{y}_i$, $\boldsymbol{y}_j$, $\boldsymbol{y}_m$, and $\boldsymbol{y}_n \in \mathbf{Y}$, according to triangular inequality, we have

$$\|\boldsymbol{y}_i - \boldsymbol{y}_j\|_1 \leq \|\boldsymbol{y}_m - \boldsymbol{y}_n\|_1 + \|\boldsymbol{y}_i - \boldsymbol{y}_m\|_1 + \|\boldsymbol{y}_j - \boldsymbol{y}_n\|_1 \quad (9)$$

and

$$\|\boldsymbol{y}_m - \boldsymbol{y}_n\|_1 \leq \|\boldsymbol{y}_i - \boldsymbol{y}_j\|_1 + \|\boldsymbol{y}_m - \boldsymbol{y}_i\|_1 + \|\boldsymbol{y}_n - \boldsymbol{y}_j\|_1. \quad (10)$$

Based on Lagrange's mean value theorem, for two Laplacian kernel functions $K(\boldsymbol{y}_i, \boldsymbol{y}_j)$ and $K(\boldsymbol{y}_m, \boldsymbol{y}_n)$, we have

$$\begin{aligned}
&(K(\boldsymbol{y}_i, \boldsymbol{y}_j) - K(\boldsymbol{y}_m, \boldsymbol{y}_n))^2 \\
&= (\exp(-\gamma\|\boldsymbol{y}_i - \boldsymbol{y}_j\|_1) - \exp(-\gamma\|\boldsymbol{y}_m - \boldsymbol{y}_n\|_1))^2 \\
&= (K'(\psi)(-\gamma(\|\boldsymbol{y}_i - \boldsymbol{y}_j\|_1 - \|\boldsymbol{y}_m - \boldsymbol{y}_n\|_1)))^2 \quad (11) \\
&\leq (K'(\psi)\gamma)^2(\|\boldsymbol{y}_i - \boldsymbol{y}_m\|_1 + \|\boldsymbol{y}_j - \boldsymbol{y}_n\|_1)^2 \\
&\leq 2(K'(\psi)\gamma)^2(\|\boldsymbol{y}_i - \boldsymbol{y}_m\|_1 + \|\boldsymbol{y}_j - \boldsymbol{y}_n\|_1)^2,
\end{aligned}$$

where $K'(\psi)$ is the derivative of the Laplacian kernel with $\psi \geq 0$. According to Lemma 1, we take $C_{\mathcal{Y}}^K = 2\gamma^2$, which concludes that

$$\epsilon \leq 4\sqrt{2}T^{3/2}\gamma\sqrt{C'q} + 2C'\gamma^2 Tq\|W^{-1}\|_F. \quad (12)$$

□

3. Study of the Newly Defined Function

In addition to the theoretical analysis (from both information theory and convex optimization perspectives) presented in the paper, we have conducted additional experiments to further demonstrate the effectiveness of our newly defined non-rigid registration approach. To this end, we vary $\lambda \in [0.06, 2.1]$ and register the source bunny to the deformed one.

The statistical results, summarized in Fig. 1, provide valuable insights into the performance of our method. These results demonstrate that our approach achieves higher-quality deformations and requires fewer iterations

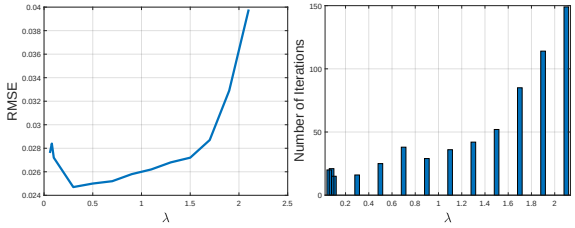
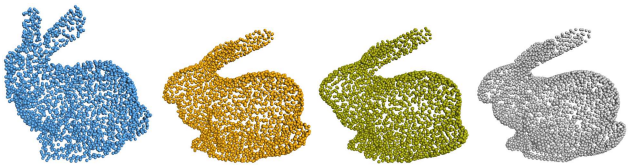


Figure 1. Investigation of the advantages of our newly defined registration function by varying λ.



(a) Source (b) Laplacian (c) Gaussian (d) Target

Figure 2. Qualitative comparisons between the Laplacian and Gaussian kernels with the noise level equating to 0.02. The Laplacian kernel delivers more accurate registration results.

when the algorithm converges within the range of $\lambda \in [0.3, 1]$. Deviation from this optimal range, either with smaller or larger values of λ, leads to less accurate registrations and increases the number of iterations required for convergence. This finding can be well explained by considering the impact of λ on the registration process. A smaller value of λ reduces the influence of entropy and convexity, potentially compromising the accuracy of the registration. On the other hand, a larger value of λ pulls the original objective function towards a completely convex one, introducing more deviations and necessitating additional iterations to achieve convergence, even under the same convergence threshold.

4. More Qualitative Results

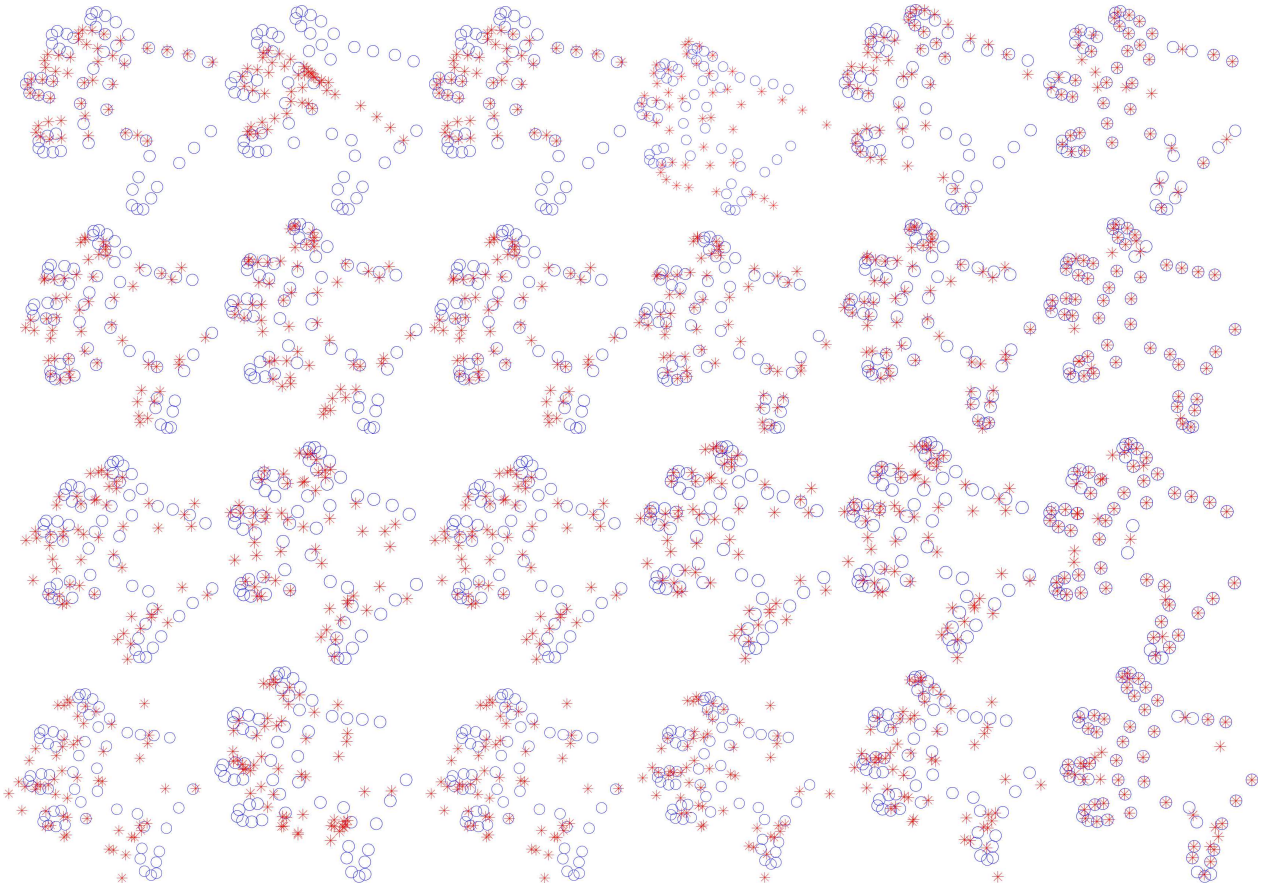
We provide further qualitative comparison results from Fig. 2 to Fig. 4. Moreover, we showcase the versatility of our proposed method by demonstrating additional shape transfer results in Fig. 5 utilizing models sourced from the ShapeNet dataset [1].

References

[1] Angel X Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, et al. Shapenet: An information-rich 3d model repository. *arXiv preprint arXiv:1512.03012*, 2015. 2, 5

[2] Christopher Williams and Matthias Seeger. Using the nyström method to speed up kernel machines. *Proc. Int. Conf. Neural Inf. Process. Syst.*, 13, 2000. 1

[3] Kai Zhang, Ivor W Tsang, and James T Kwok. Improved Nyström low-rank approximation and error analysis. In *Proc. Int. Conf. Mach. Learn.*, pages 1232–1239, 2008. 1



(a) Input (b) MR-RPM (c) BCPD (d) GMMReg (e) ZAC (f) Ours

Figure 3. Qualitative comparisons on the hand pose dataset with occlusion (first row) and noise (the remaining rows) disturbances.

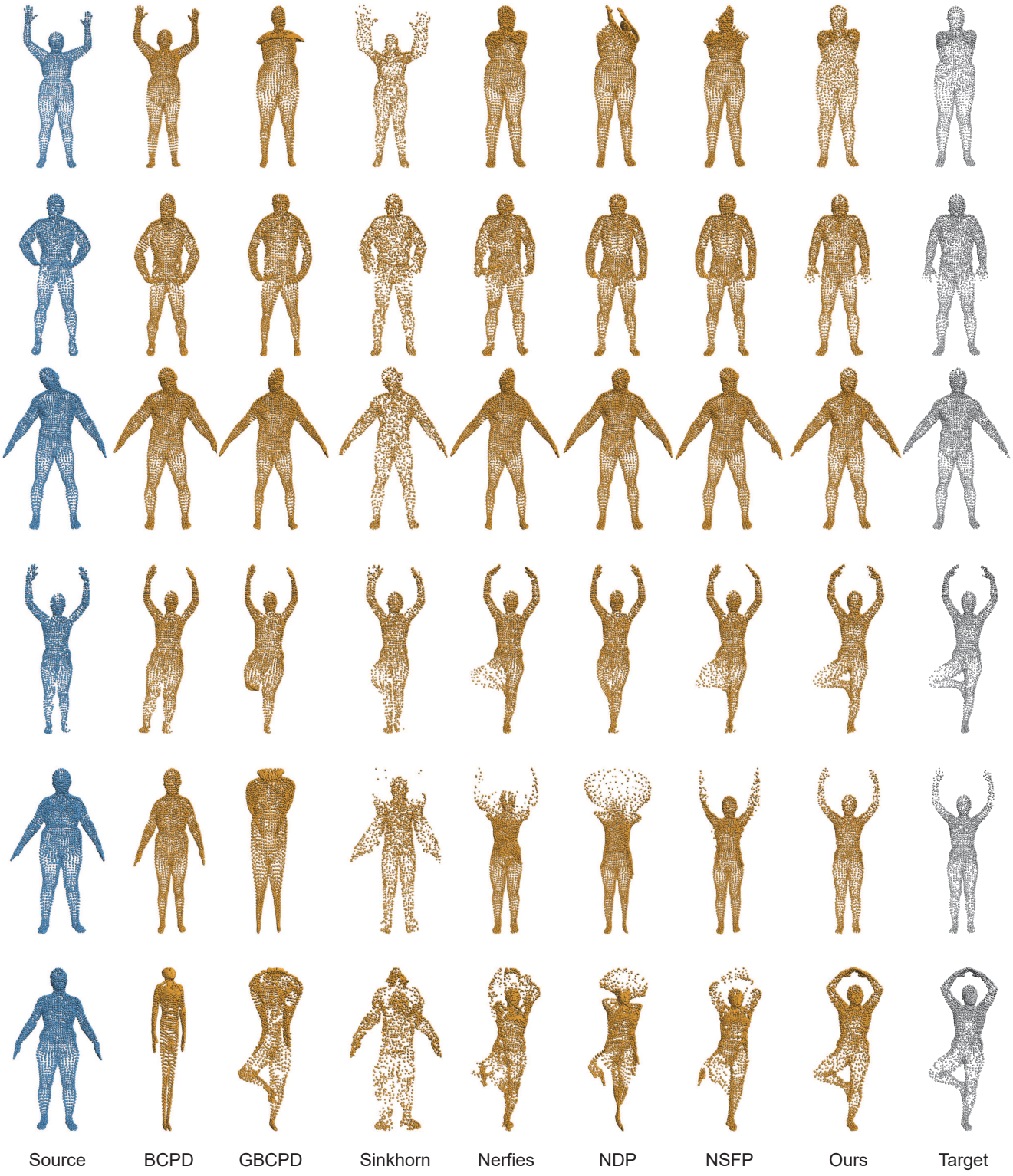


Figure 4. More qualitative comparison results on the FAUST dataset.

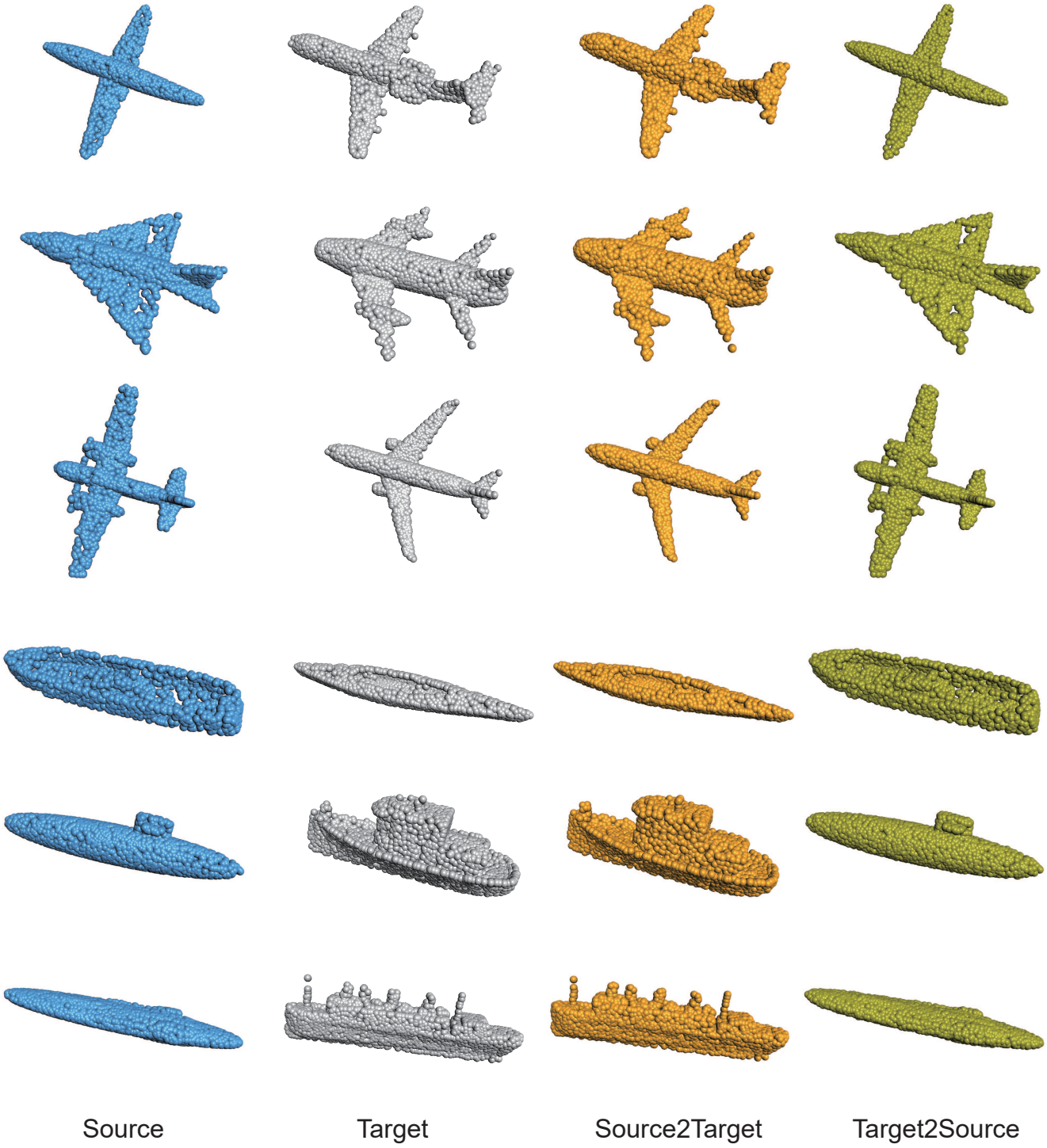


Figure 5. Application of the proposed method to shape transfer. The models `airplane` and `boat` are from the ShapeNet dataset [1].