Columbia Artificial Intelligence and Robotics Lab

# GarmentNets: Category-Level Pose Estimation for Garments via Canonical Space Shape Completion

Cheng Chi and Shuran Song

Code & Data: garmentnets.cs.columbia.edu

## Task



1. Pickup & Observe  2. Map to Canonical Pose  3. Completion with WNF  4. Pose in Task Space

**Category-level Pose Estimation for Garments.** Given a partially occluded point cloud of an unseen garment instance, the goal of the algorithm is to infer full configuration of the garment including both observed and unobserved surface.
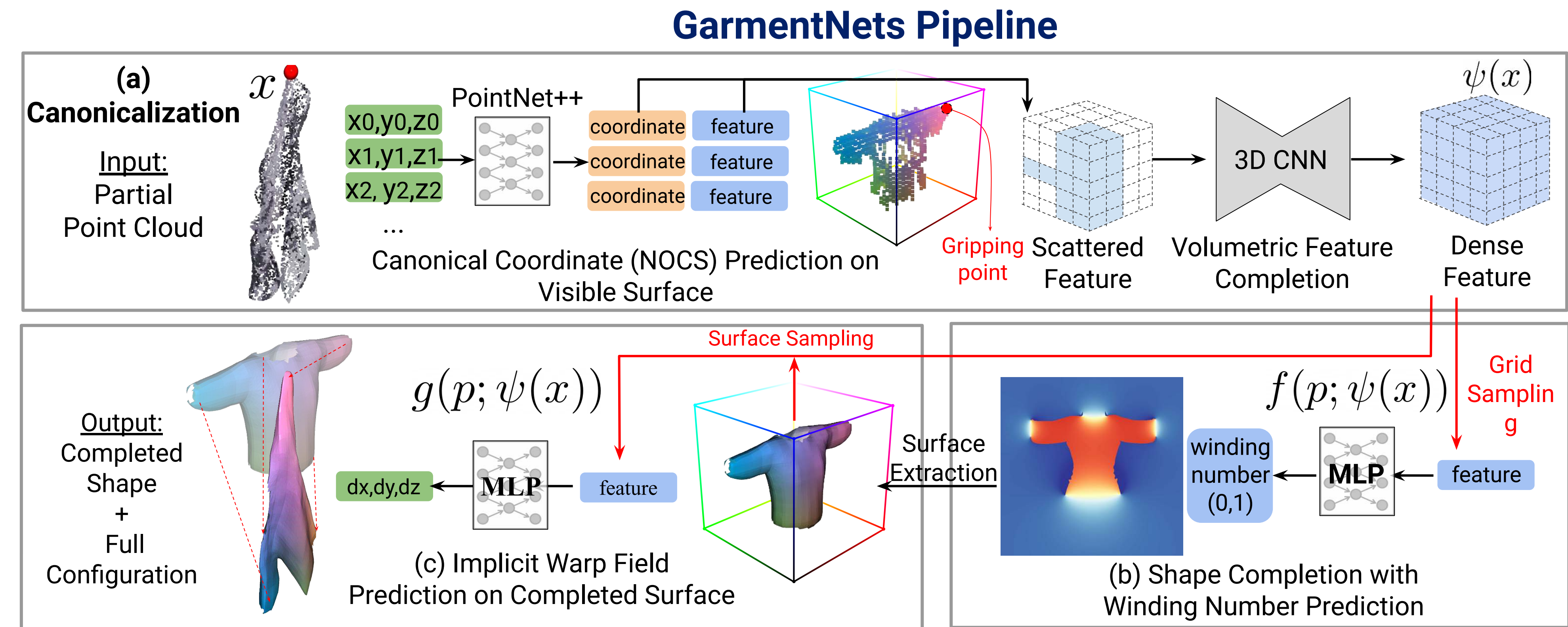
## Challenges for Garment Perception



- Infinite DoF   • Severe self-occlusion   • Thin structure

## Key idea

The key idea of GarmentNets is to formulate the pose estimation problem as a shape completion task in the canonical space. This canonical space is defined across garment instances within a category, therefore, specifies the shared category-level pose.
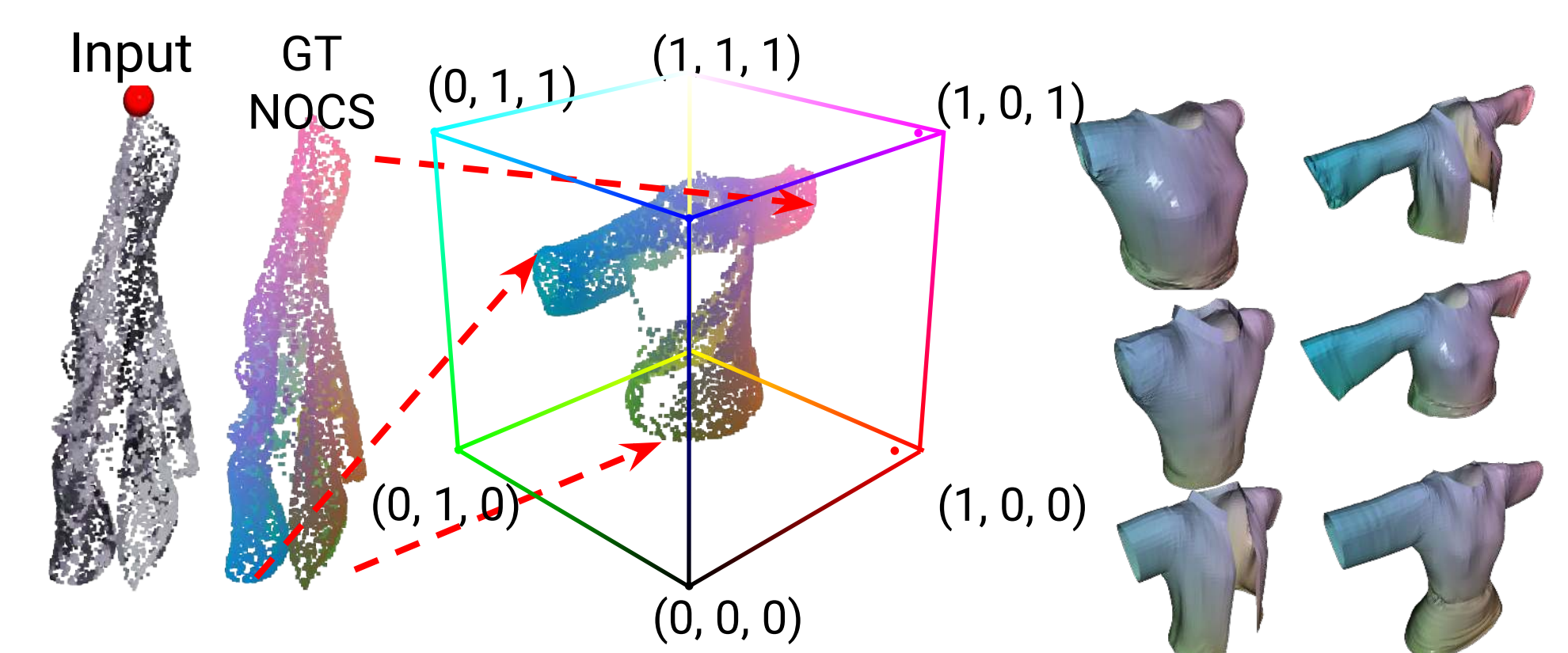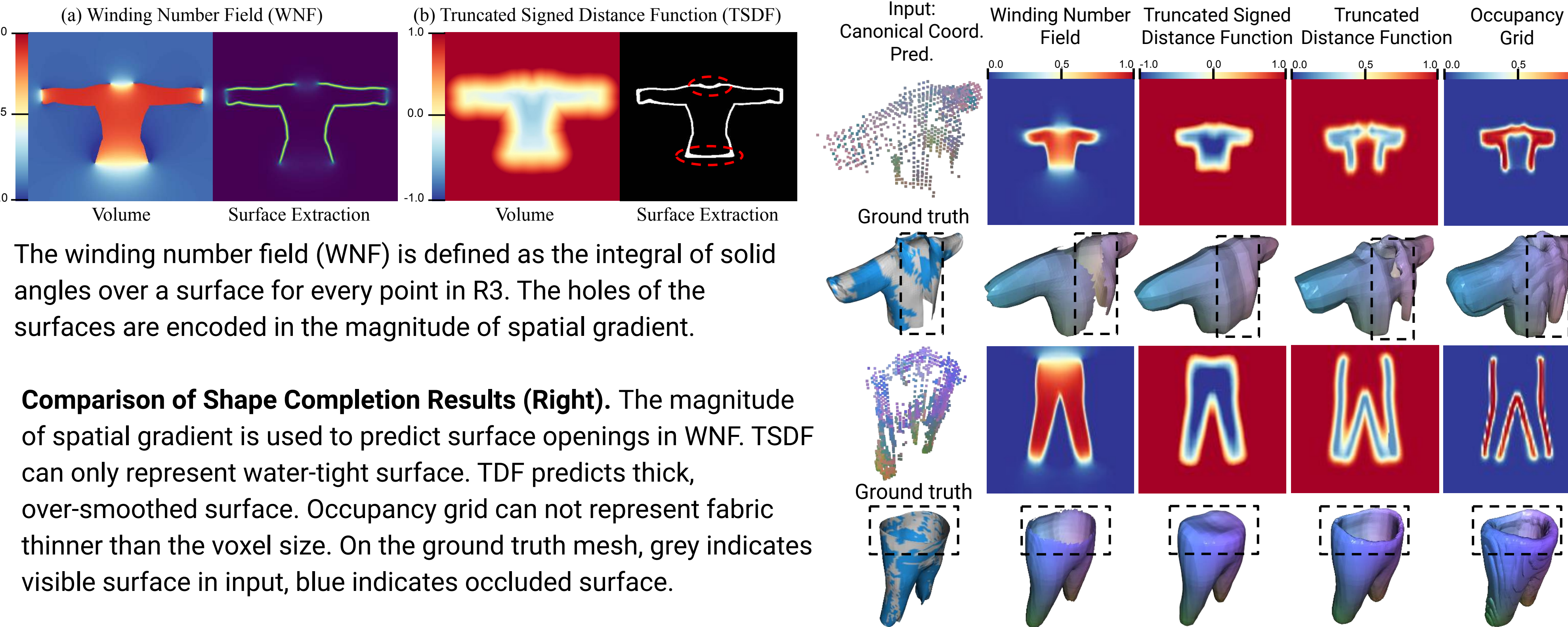


The **normalized coordinate** is obtained by scaling garments of the same category (worn by T-pose human) to a unit cube.

## GarmentNets Pipeline



(a) **Canonicalization**
Input: Partial Point Cloud
Canonical Coordinate (NOCS) Prediction on Visible Surface
Gripping point  Scattered Feature  Volumetric Feature Completion  Dense Feature

Output: Completed Shape + Full Configuration
$g(p; \psi(x))$
dx,dy,dz
(c) Implicit Warp Field Prediction on Completed Surface

$f(p; \psi(x))$
winding number (0,1)
(b) Shape Completion with Winding Number Prediction

**Network Overview.** Given a point cloud of a grasped garment in the task space, the (a) NOCS network predicts canonical coordinates for the observed points. The predicted coordinates are used to scatter the point-wise features into a 3D volume by providing the position index in the target volume. Then the (b) shape completion network infers the garment's full 3D geometry by predicting a winding number field for sampled positions p. Finally, the (c) warp field network predicts an implicit warp field that maps the completed surface (in canonical pose) back to original task space. The output mesh then encodes the garment's full configuration using per-vertex canonical coordinate label.

## 3D Shape Representation with 3D Winding Number Fields



(a) Winding Number Field (WNF)   (b) Truncated Signed Distance Function (TSDF)

The winding number field (WNF) is defined as the integral of solid angles over a surface for every point in R3. The holes of the surfaces are encoded in the magnitude of spatial gradient.

**Comparison of Shape Completion Results (Right).** The magnitude of spatial gradient is used to predict surface openings in WNF. TSDF can only represent water-tight surface. TDF predicts thick, over-smoothed surface. Occupancy grid can not represent fabric thinner than the voxel size. On the ground truth mesh, grey indicates visible surface in input, blue indicates occluded surface.
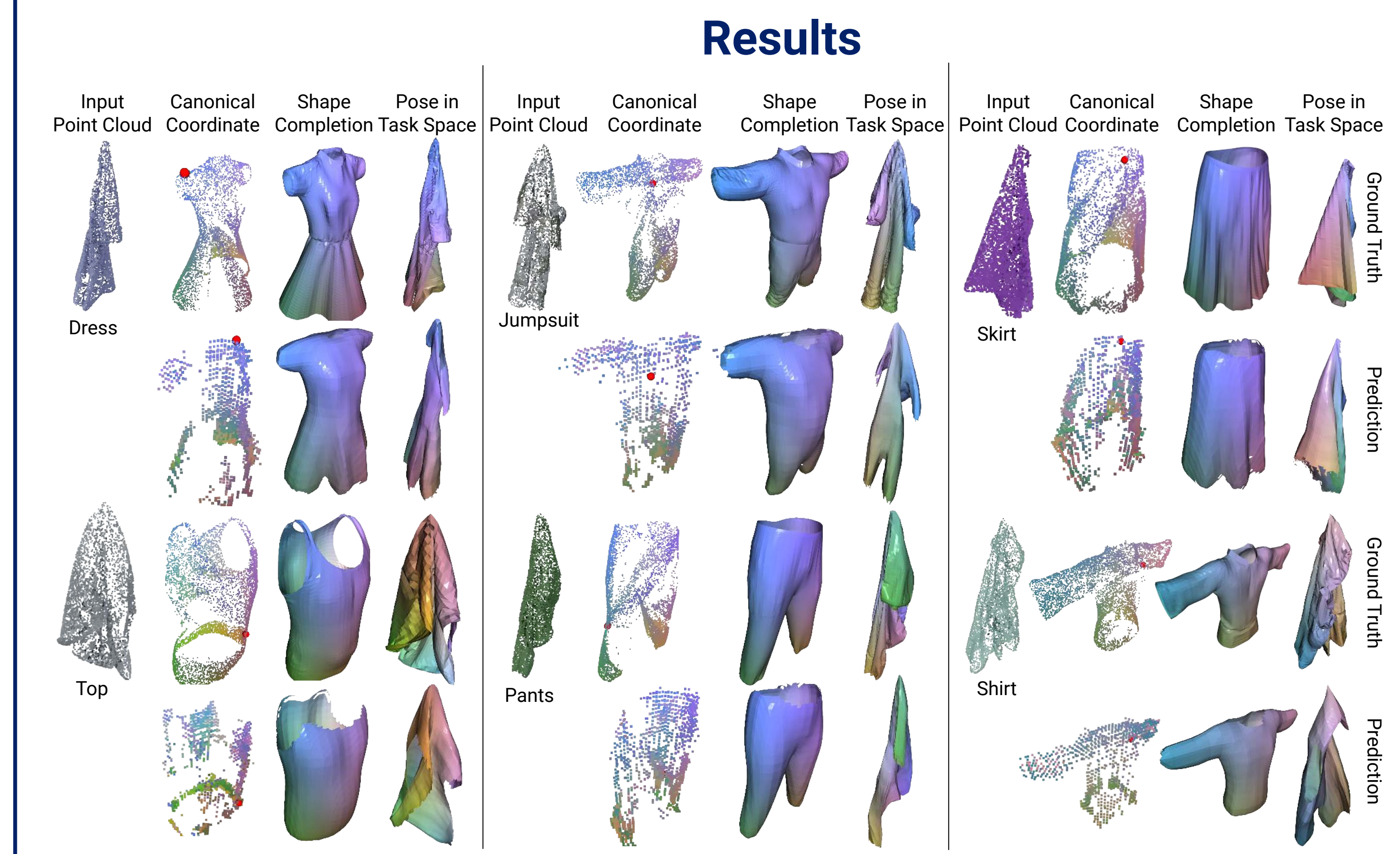
## Results



**Qualitative Results on Unseen Garment Instances (Simulation).**



**Qualitative Results on Unseen Garment Instances (Realworld Data).** We validate our algorithm on realworld garments where the garments is lifted by the robot arm and the point cloud captured with iPhone 12 Pro Max.

| Method | Dress | Jumpsuit | Skirt | Top | Pants | Shirt |
|---|---|---|---|---|---|---|
| OCC | 2.94 | 3.00 | 2.44 | 1.43 | 2.03 | 2.50 |
| TSDF | 2.45 | 1.76 | 3.03 | 2.38 | 1.44 | 1.98 |
| TDF | 2.55 | 2.18 | 2.08 | **1.22** | 1.67 | 2.11 |
| Ours | **1.94** | **1.45** | **2.00** | 1.30 | **1.03** | **1.70** |

**Tab 1. Shape Completion Error.** The error is measured using Chamfer distance (cm) under the canonical pose. Occ: occupancy grid, TSDF: truncated signed distance function, TDF: truncated unsigned distance function and Ours: winding number field.

| | Method | Dress | Jump. | Skirt | Top | Pants | Shirt |
|---|---|---|---|---|---|---|---|
| $D_c$ | NN | **2.09** | 1.89 | 2.18 | 1.82 | **1.39** | 1.69 |
| | Direct | 13.62 | 63.81 | 12.55 | 9.63 | 11.12 | 9.40 |
| | Ours | 2.12 | **1.82** | **2.14** | **1.54** | 1.41 | **1.63** |
| $D_n$ | NN | 12.74 | 13.38 | 20.55 | 11.57 | 12.43 | 12.11 |
| | Direct | 48.32 | 81.79 | 39.73 | 31.19 | 36.44 | 43.43 |
| | Ours | **6.63** | **6.06** | **7.34** | **4.47** | **4.37** | **4.94** |

**Tab 2. Pose Estimation Error.** While NN achieves comparable Chamfer distance (Dc), the pose estimation error (Dn) is significantly higher, indicates that the retrieved mesh does not share similar configuration as the input.