Semantic-Aware Implicit Template Learning via Part Deformation Consistency

Sihyeon Kim   Minseok Joo   Jaewon Lee
Juyeon Ko   Juhan Cha   Hyunwoo J. Kim*

Department of Computer Science and Engineering, Korea University

{sh_bs15, wlgkcjf87, 2j1ejyu, juyon98, hanchaa, hyunwoojkim}@korea.ac.kr

Abstract

Learning implicit templates as neural fields has recently shown impressive performance in unsupervised shape correspondence. Despite the success, we observe current approaches, which solely rely on geometric information, often learn suboptimal deformation across generic object shapes, which have high structural variability. In this paper, we highlight the importance of part deformation consistency and propose a semantic-aware implicit template learning framework to enable semantically plausible deformation. By leveraging semantic prior from a self-supervised feature extractor, we suggest local conditioning with novel semantic-aware deformation code and deformation consistency regularizations regarding part deformation, global deformation, and global scaling. Our extensive experiments demonstrate the superiority of the proposed method over baselines in various tasks: keypoint transfer, part label transfer, and texture transfer. More interestingly, our framework shows a larger performance gain under more challenging settings. We also provide qualitative analyses to validate the effectiveness of semantic-aware deformation. The code is available at https://github.com/mlvlab/PDC.

1. Introduction

Template learning is essential for shape analysis since the template, which is a compact representation of arbitrary shapes, provides a good prior for shape fitting so that we can establish dense correspondence and transfer key attributes such as texture or keypoints between shapes. Traditionally, handcrafted templates such as meshes, part elements with primitive shapes (*e.g.*, sphere, cylinder, cone), or parametric models have been widely used for human/animal body [3, 4, 5, 6, 7, 8, 9], human face [10, 11], human hand [12, 13, 14], and generic objects [15, 16, 17]. Defining templates for less complex shapes with consistent topologies and similar structures is relatively straightforward. However, for shapes like generic object shapes with high structural variability or
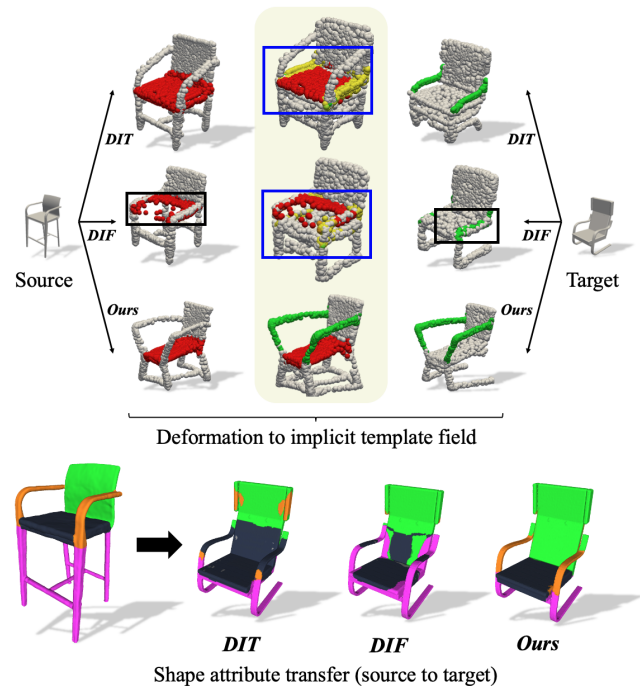

Figure 1. **Importance of part deformation consistency.** Given input shapes with high structural variability, only considering low-level geometry during template learning [1, 2] can lead to suboptimal deformation (*e.g.*, cannot distinguish between seat and arm where two parts are geometrically close while semantically distant: low seat-low arm, high seat-high arm). Ours succeeds in transferring shape attributes by encouraging part deformation consistency with semantic information.

non-rigid shapes with large-scale deformations, manually defining suitable templates becomes challenging.

To address this challenge, recent works [1, 2, 18, 19] have focused on learning implicit templates based on neural fields [20]. Neural fields, which are known for strong representation power, efficiency, and flexibility, have been successful in multiple 3D vision tasks, *e.g.*, shape/scene reconstruction [21, 22, 23, 24], neural rendering [25, 26, 27], and human digitization [28, 29, 30]. Likewise, neural fields-based template learning have shown superior perfor-

---

*Corresponding author.

mance on dense correspondence between complex shapes. Both [1] and [2] suggest decomposing the implicit representation of shapes into a deformation field and a template field. Here, the template is learned as a continuous function, which is expected to capture shared structures among the given shapes within a category (*e.g.*, airplane, chair), while the deformation depends solely on the conditioned geometry of each shape (*i.e.*, SDF value and surface normal). We observe current approaches often learn *suboptimal deformation* especially when a shape category has high structural variability like diverse local deformation scales and unseen/missing part structures; see Figure 1. That is, incorporating extra knowledge of part semantics is necessary to handle such cases.

In this paper, we propose a novel framework to learn semantically plausible deformation by distilling the knowledge of self-supervised feature extractor [31] and employing it as a semantic prior to satisfy *part deformation consistency* during the template learning process. Part deformation consistency is our inductive bias that the deformation result within the same part should be consistent. By leveraging semantic prior to understand part arrangements of shapes, we suggest local conditioning with semantic-aware deformation code and carefully designed regularizations to encourage semantically corresponding deformation. Semantic-aware deformation code is a point-wise soft assignment of part deformation priors, where part deformation priors are multiple latent codes to encode information on corresponding part deformation. We calculate assigned weights considering the part semantics of each point. In addition, we propose input space regularization and latent space regularization to encourage part-level distances to be close, providing an explicit way to control the deformation field so that the model can learn flexible deformation and successfully learn common structures under diverse shapes. Lastly, we suggest global scale consistency regularization to preserve the volume of the template against large-scale deformations. To validate the effectiveness of our framework, we conduct extensive surrogate experiments reflecting different levels of correspondence such as keypoint transfer, part label transfer, and texture transfer, and demonstrate competitive performance over baselines with qualitative analyses.

Our **contributions** are summarized as follows: ① We propose a new framework that learns global implicit template fields founded on part deformation consistency while understanding part semantics to enable semantically plausible deformation, ② We impose hybrid conditioning by combining global conditioning of geometry and local conditioning of part semantics for flexible deformation, ③ We design novel regularizations to successfully manipulate the template fields, ④ We show the effectiveness of our framework with both qualitative and quantitative experiments.

2. Related Works

Template learning for shape correspondence. Defining a template for a set of arbitrary shapes has been studied for a long time. A high-quality template provides a good initialization for shape alignment and various applications in vision and graphics. For shapes with consistent topologies and structures such as faces or bodies of humans and animals, handcrafted templates have been popular [3, 5, 7, 8, 9] and still widely used today. On the other hand, it was more challenging for generic object shapes with inconsistent topology and diverse part structures. Prior works suggest to predefine or learn local templates from data as in primitive parts [15] or part elements [32, 16, 33, 17, 34] to handle the inconsistent structure. Recent works have tried to learn a canonical shape of generic objects through implicit template field [1, 2, 35, 36], which are proven to be effective for representing shapes. However, existing approaches mostly lack explicit consideration of semantic knowledge, which enables semantically plausible deformation so that the learned template captures a common structure even with the various structured shapes.

Neural fields for 3D representations. Recent studies have shown that representing 3D shapes or scenes as implicit representations with a neural network has several advantages: powerful representation, flexibility, and memory efficiency. These are called neural fields [20]. For instance, Occupancy networks [22] proposed a method to represent 3D geometry as a continuous occupancy function. DeepSDF [23] learns 3D complex surfaces with a neural network of auto-decoder framework for 3D modeling. Further, implicit representation has been introduced to various applications [37, 38]. To leverage the advantages, many applications based on neural fields have emerged such as neural rendering [25, 26, 27], human digitization [28, 29, 30], and generative modeling [39, 40, 41, 42]. They have also been applied to learn the template of complex shapes as mentioned above.

Utilizing self-supervised segmentation models. Recently, there has been a growing body of research on analyzing complex shapes in a self-supervised manner due to the high cost of annotation for 3D data. Unsupervised semantic segmentation/co-segmentation models like BAE-Net [31], RIM-Net [43], AdaCoSeg [44], and PartSLIP [45] have demonstrated the ability to comprehend part semantics without ground truth labels. For 2D and multi-modal tasks, using large-scale self-supervised segmentation models like DINO [46] as a feature extractor has been widely adopted, including pixelNeRF [47], Feature Field [48], and ClipNeRF [49]. Inspired by these recent works, we utilize pretrained unsupervised part segmentation models to harness part semantic information for template learning.
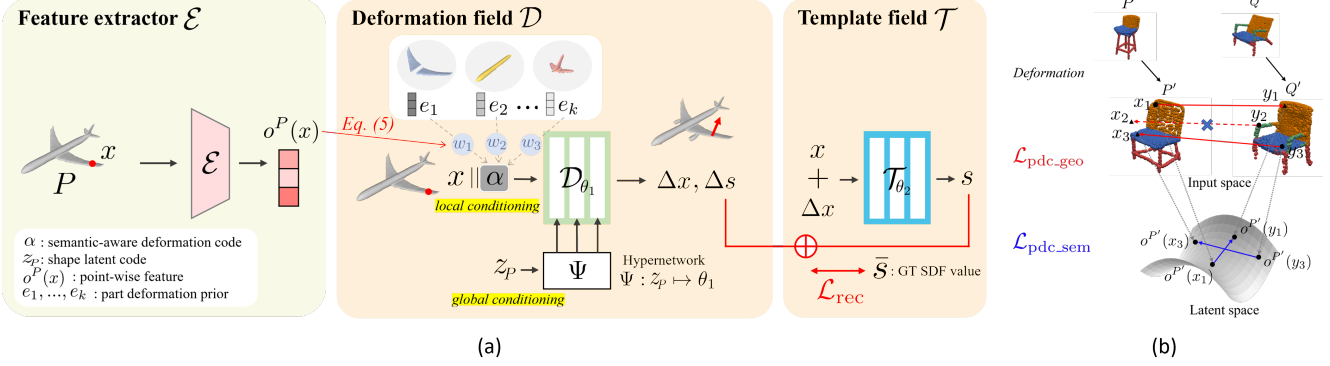
Figure 2. **Framework of semantic-aware implicit template learning.** To learn semantic-aware implicit template field, (a) we extract part semantics from pretrained feature extractor E and incorporate knowledge into semantic-aware deformation code ($\boldsymbol{\alpha}$) and (b) part deformation consistency regularizations. With optimizing with the reconstruction loss $\mathcal{L}_{\text{rec}}$ and several regularizers, our framework succeeds to learn semantically plausible deformation.

3. Preliminaries

Implicit templates have been studied with *neural fields*. Given a 3D shape P, a neural field, which is a neural network $\mathcal{F}_\theta$ parameterized by θ, approximates continuous functions such as an occupancy function [21, 22] or a Signed Distance Function (SDF) [23] as

$$\mathcal{F}_\theta(\boldsymbol{x}; \boldsymbol{z}_P) = s. \tag{1}$$

Herein, $\boldsymbol{x} \in \mathbb{R}^3$ is a query point and s is an occupancy/SDF function value. To represent multiple 3D shapes, for each shape, a shape latent code $\boldsymbol{z}_P \in \mathbb{R}^d$ is learned. In general, a surface is defined as the zero-level set of a SDF, *i.e.*, $\{x | \mathcal{F}_\theta(\boldsymbol{x}; \boldsymbol{z}_P) = 0\}$ as in DeepSDF [23].

The neural field can be further decomposed into an *implicit template field* $\mathcal{T}$ and a deformation field $\mathcal{D}$ [1, 2]:

$$\mathcal{F}_\theta(\boldsymbol{x}; \boldsymbol{z}_P) = \mathcal{T}_{\theta_2}(\mathcal{D}_{\theta_1}(\boldsymbol{x}; \boldsymbol{z}_P)). \tag{2}$$

The implicit template field $\mathcal{T}_{\theta_2}$ captures a common structure and outputs an occupancy/SDF value s in the template space. $\mathcal{D}_{\theta_1}$ maps a point x from a shape-specific space to the common template space by adding a deformation vector $\Delta\boldsymbol{x} \in \mathbb{R}^3$, *i.e.*, $\boldsymbol{x} + \Delta\boldsymbol{x}$. In DIF-Net [2], $\mathcal{D}_{\theta_1}$ additionally predicts a correction value $\Delta s \in \mathbb{R}$ to capture instance-specific local structures. In this case, the neural field $\mathcal{F}_\theta$ is defined as

$$\begin{aligned} \mathcal{F}_\theta(\boldsymbol{x}; \boldsymbol{z}_P) &= \mathcal{T}_{\theta_2}(\boldsymbol{x} + \Delta\boldsymbol{x}) + \Delta s \\ \Delta s, \Delta\boldsymbol{x} &= \mathcal{D}_{\theta_1}(\boldsymbol{x}; \boldsymbol{z}_P). \end{aligned} \tag{3}$$

The existing approaches [1, 2] learn the implicit template field by minimizing the distance between the template and the deformed individual shapes. The distance is measured by SDF values and surface normals. Also, several regularizers are imposed such as deformation smoothness (*e.g.*, penalizing the spatial gradient of $\mathcal{D}_{\theta_1}$ and restricting the deformation scale between points). However, the loss functions

and regularizers are computed by *low-level geometric* information. They are not effective enough to handle generic objects with large structural variability. Hence, in this paper, we take *part semantics* into account for template learning to achieve 'semantically plausible' deformation and dense correspondence.

4. Method

We present a semantic-aware implicit template learning framework based on part deformation consistency. In this section, we first delineate the overall architecture of our method equipped with part deformation priors and semantic-aware deformation code. Then, we introduce our deformation consistency regularizers including part deformation consistency regularizations. Lastly, we discuss the training procedure with several additional loss functions.

4.1. Semantic-aware implicit template field

We here present a novel semantic-aware implicit template learning framework. The overall pipeline of the proposed method is shown in Figure 2(a).

Template field and deformation field. We extend the implicit template learning by incorporating semantic information into the deformation field. Specifically, given a query point $\boldsymbol{x}$ of a shape P, the deformation field $\mathcal{D}_{\theta_1}$ predicts a deformation vector $\Delta\boldsymbol{x}$ and a correction value Δs conditioned on a point-wise *semantic-aware deformation code* (SDC) $\boldsymbol{\alpha}(\boldsymbol{x}; P)$ in addition to a shape latent code $\boldsymbol{z}_P$. Then, our framework is defined as

$$\begin{aligned} \mathcal{F}_\theta(\boldsymbol{x}; P, \mathcal{E}) &:= \mathcal{T}_{\theta_2}(\boldsymbol{x} + \Delta\boldsymbol{x}) + \Delta s, \\ \Delta\boldsymbol{x}, \Delta s &= \mathcal{D}_{\theta_1}(\boldsymbol{x}; \boldsymbol{z}_P, \boldsymbol{\alpha}(\boldsymbol{x}; P)), \end{aligned} \tag{4}$$

where $\mathcal{E}$ is a pretrained feature extractor. We assign $\boldsymbol{\alpha}(\boldsymbol{x}; P)$ based on a point-wise semantic feature, which is extracted from $\mathcal{E}$. For simplicity, we denote the neural field $\mathcal{F}_\theta(\boldsymbol{x}; P, \mathcal{E})$ by $\mathcal{F}_\theta(\boldsymbol{x})$ and SDC $\boldsymbol{\alpha}(\boldsymbol{x}; P)$ by $\boldsymbol{\alpha}(\boldsymbol{x})$.

Feature extractor. In this paper, for feature extraction, we mainly use BAE-Net [31] as an encoder $\mathcal{E}$, which is proven effective in capturing part semantics. Note that we pretrained BAE-Net on selected categories of ShapeNetV2 [50] by self-supervision without any extra labels or data for a fair comparison. The features from BAE-Net are k dimensional vectors, *i.e.*, $\boldsymbol{o}^P(\boldsymbol{x}) = \mathcal{E}(\boldsymbol{x}, P) \in \mathbb{R}^k$. k is a hyperparameter, which is closely related to the number of semantic parts. Since the number of semantic parts is unknown in self-supervised learning, we simply chose sufficiently large k.

Semantic-aware deformation code. Our framework assumes that different semantic parts necessitate different deformations. To achieve this, the proposed method learns a deformation prior for each semantic part. The deformation priors are encoded as latent codes $\boldsymbol{e}_1, \cdots, \boldsymbol{e}_k \in \mathbb{R}^{d'}$. A point $\boldsymbol{x}$ is softly assigned to k semantic parts using the softmax values of semantic features $\boldsymbol{o}^P(\boldsymbol{x})$. Combining the soft assignment with deformation priors, *semantic-aware deformation code* (SDC) denoted as $\boldsymbol{\alpha}(\boldsymbol{x}) \in \mathbb{R}^{d'}$ is defined as

$$\boldsymbol{\alpha}(\boldsymbol{x}) = \sum_{i=1}^{k} \frac{\exp(\boldsymbol{o}_i^P(\boldsymbol{x}))\boldsymbol{e}_i}{\sum_{j=1}^{k} \exp(\boldsymbol{o}_j^P(\boldsymbol{x}))}, \forall i \in \{1, \ldots, k\}, \quad (5)$$

where $\boldsymbol{o}_i^P(\boldsymbol{x})$ denotes the i-th dimension value of the semantic features. The SDC at $\boldsymbol{x}$ is computed by a linear combination of deformation priors. We learn the point's probability of belonging to each of the k parts, which is served as part semantics.

Remarks. For comparison, we also applied hard assignment of part deformation priors for SDC, *i.e.*, $\boldsymbol{\alpha}(\boldsymbol{x}) = \boldsymbol{e}_{i^*}$, where $i^* = \operatorname{argmax}_i(\boldsymbol{o}_i^P(\boldsymbol{x}))$. However, as points are completely separated with local conditioning of the corresponding part deformation prior $\boldsymbol{e}_{i^*}$, the final deformation result tends to break the shape geometry (see Figure 3), leading to performance degradation as shown in Table 1. Thereby, we choose the soft assignment to enable flexible deformation while preserving the shape geometry.

We train the deformation field $\mathcal{D}_{\theta_1}$ and $\mathcal{T}_{\theta_2}$ in an end-to-end fashion. In our experiments, $\mathcal{D}_{\theta_1}$ is a 6-layer MLP conditioned on both *global geometry* $\boldsymbol{z}_P$ via hypernetwork Ψ [51, 52] and *local part semantics* $\boldsymbol{\alpha}(\boldsymbol{x})$ via concatenation with input coordinates. The shape latent code $\boldsymbol{z}_P$ and part deformation priors $\{\boldsymbol{e}_i\}$ are randomly initialized and jointly optimized with neural fields. $\mathcal{T}_{\theta_2}$ is a 5-layer MLP.

Table 1. **Ablations of different assignment strategies for SDC.** We conduct a quantitative comparison between hard and soft assignments with 5-shot part label transfer on ShapeNet chair.

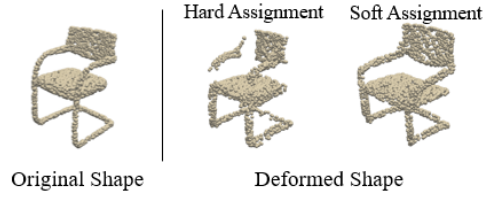| Assignment strategies	Hard	Soft
chair (5-shot PLT)	77.0	**84.2**

Figure 3. **Effects of different assignment strategies for SDC.** Hard assignment breaks the shape geometry, while soft assignment preserves the geometry while learning flexible deformations.

4.2. Part deformation consistency regularizations

In addition to conditioning neural fields, leveraging semantic information, we propose novel part deformation consistency regularizations to encourage semantically plausible deformation. Semantically similar points should be close after deformation in both an input space and a latent space. That is, the deformation of points from the same part should be consistent. Figure 2(b) illustrates the concept of our regularizations. Given a pair of deformed shapes P', Q', part deformation consistency regularizations are defined as

$$\mathcal{L}_{\text{pdc\_geo}} = \sum_{i=1}^{k} \left(\frac{1}{|\mathcal{P}_i'|} \sum_{\boldsymbol{x} \in \mathcal{P}_i'} \|\boldsymbol{x} - C(\boldsymbol{x}, \mathcal{Q}_i')\|_2^2 \right.$$
$$\left. + \frac{1}{|\mathcal{Q}_i'|} \sum_{\boldsymbol{y} \in \mathcal{Q}_i'} \|\boldsymbol{y} - C(\boldsymbol{y}, \mathcal{P}_i')\|_2^2 \right), \text{ and} \quad (6)$$

$$\mathcal{L}_{\text{pdc\_sem}} = \sum_{i=1}^{k} \left(\frac{1}{|\mathcal{P}_i'|} \sum_{\boldsymbol{x} \in \mathcal{P}_i'} \|\boldsymbol{o}^{P'}(\boldsymbol{x}) - \boldsymbol{o}^{Q'}(C(\boldsymbol{x}, \mathcal{Q}_i'))\|_2^2 \right.$$
$$\left. + \frac{1}{|\mathcal{Q}_i'|} \sum_{\boldsymbol{y} \in \mathcal{Q}_i'} \|\boldsymbol{o}^{Q'}(\boldsymbol{y}) - \boldsymbol{o}^{P'}(C(\boldsymbol{y}, \mathcal{P}_i'))\|_2^2 \right), \quad (7)$$

where C is a correspondence matching function for a point $\boldsymbol{x}$ and its nearest point in a set $\mathcal{Q}$, *i.e.*, $C(\boldsymbol{x}, \mathcal{Q}) = \operatorname{argmin}_{\boldsymbol{y} \in \mathcal{Q}} \|\boldsymbol{x} - \boldsymbol{y}\|_2^2$. $\mathcal{P}_i'$ and $\mathcal{Q}_i'$ denote sampled point sets of ith deformed part shapes of P' and Q'. They are defined as $\mathcal{P}_i' = \{\boldsymbol{x} + \mathcal{D}_{\theta_1}(\boldsymbol{x}; \boldsymbol{z}_P, \boldsymbol{\alpha}(\boldsymbol{x}; P)) \,|\, i{=}\operatorname*{argmax}_k(\boldsymbol{o}_k^P(\boldsymbol{x})), \boldsymbol{x} \in \Omega_P\}$ and $\mathcal{Q}_i' = \{\boldsymbol{x} + \mathcal{D}_{\theta_1}(\boldsymbol{x}; \boldsymbol{z}_Q, \boldsymbol{\alpha}(\boldsymbol{x}; Q)) \,|\, i = \operatorname*{argmax}_k(\boldsymbol{o}_k^Q(\boldsymbol{x})), \boldsymbol{x} \in \Omega_Q\}$, where Ω_P and Ω_Q are sampled query point sets for shape P and Q. Note that each deformed part shape is extracted based on part semantics from *original* shapes P and Q. Here, we explicitly guide the part-level deformation among various shapes to be consistent. Moreover, since we understand part arrangement and
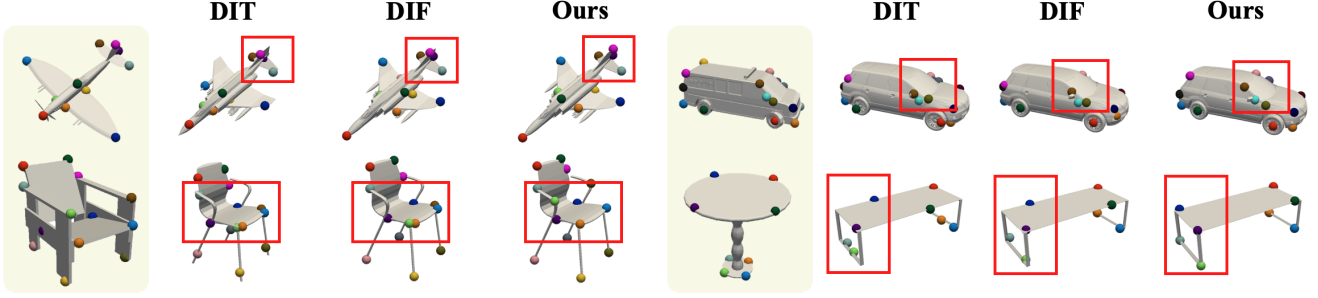
Figure 4. **Keypoint transfer result on ShapeNet.** We transfer keypoint labels from source shapes (colored in ivory) to target shapes and compare ours with DIT [1] and DIF [2]. For additional visualizations including results on ScanObjectNN [53], see the supplement.
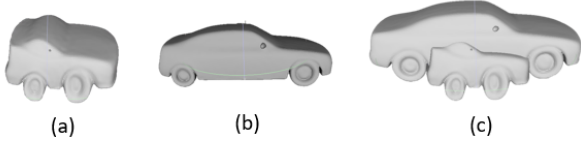

Figure 5. **Efficacy of global scaling consistency** $\mathcal{L}_{\text{scale}}$.

only penalize matched parts, our method does not forcibly deform "unseen/missing match" cases.

We partly observed spatial distortion when part deformation consistency was too strongly imposed. This is because using external semantics for deformation could break the continuity of geometric fields. In addition, we propose *global scale consistency* and *global deformation consistency* to address the concern. For the former one, we disentangle a scaling operation from the entire non-rigid deformation $\mathcal{D}_{\theta_1}$ to calculate the global scale of the learned implicit template field. That is, we want to approximate $\mathcal{D}$ with the global scaling of the scaling factor of r, *i.e.*, $\mathcal{D} \approx rI$. Then the problem becomes finding r from

$$\min_{r} \sum_{s} \sum_{i} \|\boldsymbol{x}_i^s + \Delta \boldsymbol{x}_i^s - r \boldsymbol{x}_i^s\|_2^2, \qquad (8)$$

where s and i are an index for shape and point, respectively. A closed-form solution of (8) is $r = \sum_{s} \sum_{i} \frac{\boldsymbol{x}_i^{s\top}(\boldsymbol{x}_i^s + \Delta \boldsymbol{x}_i^s)}{\boldsymbol{x}_i^{s\top} \boldsymbol{x}_i^s}$. See the supplement for proof. Since we want the learned $\mathcal{D}$ to be insensitive to global scale operation, *i.e.*, the implicit template to keep its scale by the average scale of input shapes, we penalize the expected $\mathbb{E}[r]$ over input shapes $P_1 \cdots P_N$ in a single batch as:

$$\mathcal{L}_{\text{scale}} = \left| \mathbb{E}[r] - 1 \right| \qquad (9)$$

Figure 5 illuminates the effect of $\mathcal{L}_{\text{scale}}$. For the same input shapes, Figure 5(a) is a car template learned without $\mathcal{L}_{\text{scale}}$, Figure 5(b) is a car template learned with $\mathcal{L}_{\text{scale}}$, and Figure 5(c) is the result of placing two shapes in the same space. It shows $\mathcal{L}_{\text{scale}}$ helps the model to learn a more desirable template by penalizing the change in the global scale of the learned template fields.

For the latter one, we add global deformation consistency for a pair of deformed shapes P', Q':

$$\mathcal{L}_{\text{geo}} = \mathbf{CD}(P', Q'), \qquad (10)$$

where $\mathbf{CD}$ denotes a conventional Chamfer loss. Thus, the deformation consistency regularization can be written as:

$$\mathcal{L}_{\text{dc}} = \gamma_1 \mathcal{L}_{\text{pdc\_geo}} + \gamma_2 \mathcal{L}_{\text{pdc\_sem}} + \gamma_3 \mathcal{L}_{\text{scale}} + \gamma_4 \mathcal{L}_{\text{geo}}. \quad (11)$$

4.3. Optimization

We train our deformation and template fields with a reconstruction loss and several regularizers including our deformation consistency regularization. Following [2], the reconstruction loss $\mathcal{L}_{\text{rec}}$ for shape P is

$$\mathcal{L}_{\text{rec}} = \sum_{\boldsymbol{x} \in \Omega_P} \left| \mathcal{F}_\theta(\boldsymbol{x}) - \bar{s} \right| + \sum_{\boldsymbol{x} \in P} (1 - \langle \nabla \mathcal{F}_\theta(\boldsymbol{x}), \bar{n} \rangle)$$
$$+ \sum_{\boldsymbol{x} \in \Omega_P} \left| \|\nabla \mathcal{F}_\theta(\boldsymbol{x})\|_2 - 1 \right| + \sum_{\boldsymbol{x} \in \Omega_P \setminus P} \exp(-\delta \cdot |\mathcal{F}_\theta(\boldsymbol{x})|),$$
$$(12)$$

where $\bar{s}$ denotes the ground-truth SDF value of each query point, $\bar{n}$ is the ground truth surface normal vector of each surface point and $\nabla \mathcal{F}_\theta$ the spatial derivative of a 3D field. The second term of (12) is derived from the property of the SDF function that the gradient of an SDF is identical to the surface normal for surface points and the third is implicit geometric regularization [54] derived from the Eikonal equation. The last term pertains to penalizing off-surface points to be away from the surface points, which is controlled by $\delta \gg 1$.

Since we do not know ground truth mapping between the template and input shapes, we manipulate the deformation field through multiple regularizers. First, we employ our proposed deformation consistency regularization (11). For further regularizers, we adopt [2] such as deformation smoothness $\mathcal{L}_{\text{smooth}}$, normal consistency $\mathcal{L}_{\text{normal}}$ and minimal correction $\mathcal{L}_{\text{c}}$, which are defined as

$$\mathcal{L}_{\text{smooth}} = \sum_{\boldsymbol{x} \in \Omega_P} \sum_{d \in \{X, Y, Z\}} \|\nabla \mathcal{D}_{\theta_1}^v(\cdot, \boldsymbol{z}_P)|_d(\boldsymbol{x})\|_2^2, \quad (13)$$
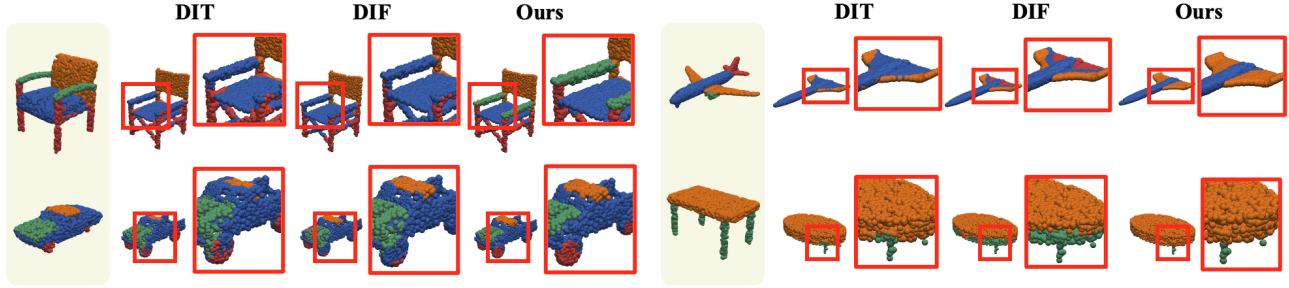
Figure 6. **Part label transfer result on ShapeNet.** We transfer part segment labels from source shapes (colored in ivory) to target shapes and compare ours with DIT [1] and DIF [2]. For additional visualizations, please refer to the supplement.

$$\mathcal{L}_{\text{normal}} = \sum_i \sum_{\boldsymbol{x} \in \mathcal{S}_i} \left(1 - \left\langle \nabla \mathcal{T}_{\theta_2}\left(\boldsymbol{x} + D_{\theta_1}^v(\boldsymbol{x})\right), \bar{n} \right\rangle\right), \tag{14}$$

$$\mathcal{L}_c = \sum_i \sum_{\boldsymbol{x} \in \Omega_P} |\mathcal{D}_{\theta_1}^{\Delta s}(\boldsymbol{x})|. \tag{15}$$

We also apply ℓ_2 regularization on shape latent codes and part deformation priors

$$\mathcal{L}_{\text{emb}} = \sum_i^N \|\boldsymbol{z}_i\|_2^2 + \sum_j^k \|\boldsymbol{e}_j\|_2^2. \tag{16}$$

The regularizers are collectively written as

$$\mathcal{L}_{\text{reg}} = \gamma_5 \mathcal{L}_{\text{smooth}} + \gamma_6 \mathcal{L}_{\text{normal}} + \gamma_7 \mathcal{L}_c + \gamma_8 \mathcal{L}_{\text{emb}}. \tag{17}$$

Hence, the final loss becomes

$$\mathcal{L} = \mathcal{L}_{\text{rec}} + \mathcal{L}_{\text{dc}} + \mathcal{L}_{\text{reg}}. \tag{18}$$

Evaluating dense correspondence. Given two pairs of seen or unseen shapes P, Q, we can evaluate the dense correspondence between P and Q by deforming each shape to the common template space. Moreover, the uncertainty of dense correspondence performance between two points $\boldsymbol{p} \in P$ and $\boldsymbol{q} \in Q$ can be measured as below:

$$u(p, q) = 1 - \exp(-\gamma \|(p + \Delta p) - (q + \Delta q)\|_2^2). \tag{19}$$

γ is a hyperparameter for modulation and $\Delta p, \Delta q$ are output from $\mathcal{D}_{\theta_1}$. Note that since each part deformation prior $\boldsymbol{e}$ is shared based on part semantics across entire training shapes, we do not need a test-time optimization on $\boldsymbol{e}$ for an unseen shape at inference time.

5. Experiments

We demonstrate the effectiveness of our proposed framework and provide both quantitative and qualitative results. In Section 5.1, we conduct multiple label transfer

Table 2. **Keypoint correspondence accuracy in KeypointNet [55].** PCK scores with threshold of 0.05 / 0.1 are reported.

PCK	airplane	car	chair	table
AtlasNetV2 [34]	23.9 / 45.3	33.5 / 64.0	25.6 / 50.2	25.3 / 48.0
DIT [2]	13.6 / 34.2	58.9 / 87.7	33.3 / 66.7	26.2 / 55.5
DIF [1]	16.5 / 40.5	67.8 / 90.4	34.3 / 57.6	37.9 / 62.9
Ours	**24.2 / 52.3**	**69.3 / 91.8**	**61.2 / 83.2**	**50.8/77.9**

Table 3. **5-shot part label transfer in four categories of ShapeNet.** The reported numbers are overall mIoU.

model	airplane	car	chair	table
AtlasNetV2 [34]	57.8	59.3	66.9	70.6
DIT [1]	64.2	66.1	80.4	86.0
DIF [2]	73.9	69.6	82.1	86.6
Ours	**75.1**	**70.4**	**84.2**	**86.9**

tasks for dense correspondence evaluation. In Section 5.2, we measure shape reconstruction accuracy to evaluate the representational power for unknown shapes. In Section 5.3, we present advanced dense correspondence analysis under more challenging settings. Lastly, in Section 5.4, we perform ablation studies to analyze the proposed regularization and the sensitivity over feature extractor [31].

Datasets and Baselines. We train and validate our model and baselines on four categories (*e.g.*, chair, table, airplane and car) of ShapeNetV2 [50]. We follow the training split of DIF [2] for each category. For label transfer tasks, we use the labels from ShapeNet-Part [56] and KeypointNet [55]. DIT [1] and DIF [2] are the most relevant baseline models as they both globally learn implicit template fields, and we select AtlasNetV2 [34] as an additional baseline. Dense correspondence is measured with local patch composition in [34]. For more implementation details, analyses, and visualizations, see the supplement.

5.1. Dense correspondence

To evaluate the quality of learned dense correspondence, we conduct various surrogate tasks including keypoint transfer, part label transfer, and texture transfer due to
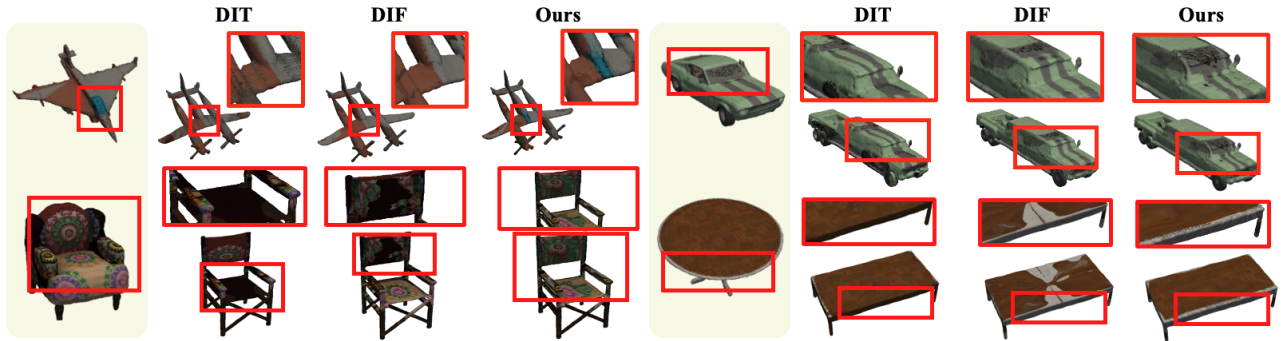
Figure 7. **Texture transfer result on ShapeNet.** We transfer part segment labels from source shapes (colored in ivory) to target shapes and compare ours with DIT [1] and DIF [2]. For additional visualizations, please refer to the supplement.

Table 4. **Reconstruction accuracy for unseen shapes in four categories of ShapeNet.** The reported numbers are chamfer distance (CD) between ground truth and reconstructed meshes at a resolution of 256^3.

CD ($\times 10^3$)↓	airplane	car	chair	table
DIT [1]	0.2261	0.4082	0.3063	0.6187
DIF [2]	0.2093	0.3157	0.4229	**0.2640**
Ours	**0.1859**	**0.3150**	**0.2265**	0.2796

the lack of labeled datasets for 3D dense correspondence. For the following tasks, we transfer respective labels from source shapes to target shapes in a common template space.

Keypoint label transfer. Table 2 presents the quantitative result in a 5-shot keypoint transfer task. We measure the performance by a percentage of correct keypoints (PCK) [57]. We compute the distance between the transferred keypoints and the ground truth points to record the PCK score with the threshold of 0.05/0.1. Our method achieves the best performance in all four categories with a significant improvement. Especially for categories with high structural variability (*e.g.*, chair and table), a large performance gain is observed. For instance, even compared to the best baseline, on the chair category with the threshold of 0.1, our method (83.2) outperforms DIT [1] (66.7) by 16.5%, and on the table category with the threshold of 0.1, our method (77.9) outperforms DIF [2] (62.9) by 15%. We additionally provide the visualization of keypoint transfer in Figure 4, which also support the importance of understanding part semantics for shape correspondence. In particular, two keypoints of pea-green and orange from the source chair are geometrically close but semantically different since one is for the arm part and the other is for the seat part. Among several methods, only our framework succeeds in disentangling two keypoints in the target chair. We provide more qualitative results regarding keypoint transfer results in real-world data [53] to validate the robustness of our method in the supplement.

Part label transfer. For few-shot part label transfer tasks, we transfer part segment labels from source shapes to target shapes. In detail, for each (target) point of the deformed target shape, we find the n nearest points of the deformed source shapes and perform weighted voting using their labels and distances to the target point. In our experiments, we used $n = 10$ nearest points. For a 5-shot part label transfer task in Table 3, five source shapes were used just the same as [2] for every category. Table 3 shows our framework achieves the best performance over all four categories. Moreover, Figure 6 illustrates the qualitative results of one-shot part label transfer on ShapeNet [50]. Similar to keypoint transfer, Figure 6 also proves the effectiveness of semantic knowledge in dense correspondence.

Texture transfer. Figure 7 shows texture transfer results on ShapeNetV2 [50] to compare established dense correspondence between baselines and ours with higher-frequency label transfer. We observe notable texture transfer results. For instance, our framework successfully transfers the source car's texture of the windshield ($gray$) to the corresponding part of the target cars, whereas baselines transfer the same textures to the roof part of the target cars.

5.2. Shape reconstruction

We report the reconstruction quality of unseen shapes with Chamfer Distance (CD) to validate the representational power of our framework in Table 4. A smaller CD means more accurate reconstruction. As in previous work, we report the CD $\times 10^3$ for readability. Also, we consistently discuss the CD and performance gain below at this scale. We optimize the corresponding shape latent code for a given unseen shape. We measure CD as in DIF [2], where reconstructed meshes are extracted at a resolution of 256^3, and CD is calculated with $10K$ sampled points. Our framework on average achieved the best reconstruction results surpassing DIF by a margin of 0.0512. On average, our method and DIF exhibit 0.2518 and 0.3030 reconstruction errors, respectively. In three out of four categories, our method achieved the best reconstruction accuracy.

5.3. More challenging setups

To further validate the effectiveness of our method, we perform additional experiments and compare the results with DIF [2]. Overall, the results below imply current methods that lack part semantics and do not consider part deformation consistency may fail to learn semantically plausible deformation in challenging cases.

2-shot part label transfer within different subcategories. Our objective is to assess dense correspondence performance under difficult settings. We conduct 2-shot part label transfer within *two* different subcategories belonging to the same category, where shapes from different subcategories tend to have different geometry for same semantics. Since ShapeNet [50] includes incorrect subcategory labels [58], we manually cleaned subcategory labels for four categories (*e.g.*, airplane: airliner/jet, car: sedan/jeep, chair: straight chair/sofa, table: pedestal table/short table). Figure 8 displays examples of shapes from each subcategory. For robust evaluations, we randomly select the source shape 5 times for each subcategory. Table 5 shows our framework achieves a significant performance gain in all cases by over 7.2% on average compared to DIF [2].

Unsupervised dense correspondence in DFAUST [59]. Furthermore, we evaluate unsupervised dense correspondence with non-rigid shapes [59]. For more challenging settings, we train both DIF [2] and our model for two different subjects with different sizes and heights (*e.g.*, subject ID of 50002, 50026) and four distinct actions (*e.g.*, shake arms, chicken wings, running on spot, and jumping jacks). Colormap in Figure 9 demonstrates the correspondence quality, where we transfer color over the up-axis. In the source shape, the arms are labeled in red. After transferring these color labels to the target shapes, our method preserves the red/orange labeling of the arms across different motions, whereas the baseline fails to capture the same labeling (*i.e.*, yellow hands) since it lacks semantic knowledge during training. Additionally, we conduct keypoint transfer tasks and measure the PCK score as identical as in Section 5.1. Our model achieves 74.9/85.2 with a threshold of 0.05/0.1, whereas DIF [2] achieves 67.0/80.7 with the same threshold, showing our superior performance in the challenging settings (see the supplement for the full keypoint transfer result). Note that we can also learn implicit templates for non-rigid shapes with supervision as in [19]. We conduct experiments in unsupervised settings since our focus is to validate the effectiveness of utilizing additional semantic information for template learning.

5.4. Ablations

Ablation on regularizations. We conduct ablation studies regarding the proposed regularizations including part deformation consistency regularizations, global scale consistency regularization, and global deformation consistency
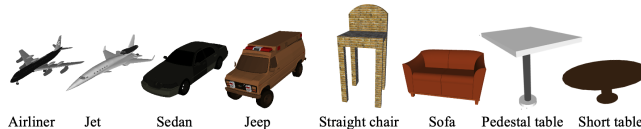

Airliner   Jet   Sedan   Jeep   Straight chair   Sofa   Pedestal table   Short table

Figure 8. **Types of source shapes for 2-shot part label transfer.**

Table 5. **2-shot part label transfer in subcategories of ShapeNet**. The reported numbers are overall mIoU over 5 runs.

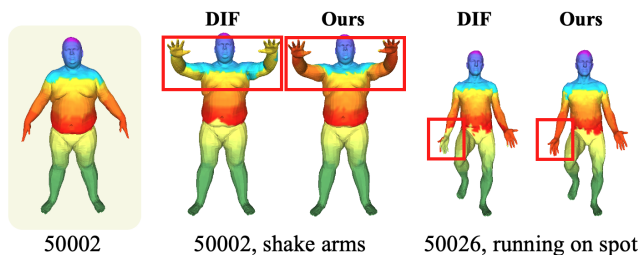| subcategories	DIF [2]	Ours
airplane(airliner/jet)	$64.5_{\pm1.61}$	$\mathbf{72.8}_{\pm0.99}$ (+8.3↑)
car(sedan/jeep)	$65.1_{\pm2.34}$	$\mathbf{69.1}_{\pm1.46}$ (+4.0↑)
chair(straight/sofa)	$66.3_{\pm6.06}$	$\mathbf{77.5}_{\pm5.81}$ (+11.2↑)
table(pedestal/short)	$74.4_{\pm10.1}$	$\mathbf{79.7}_{\pm6.26}$ (+5.3↑)

Figure 9. **Color transfer on DFAUST [59].** We transfer different color labels over the up-axis from source shapes (colored in ivory) to target shapes and compare ours with DIF [2].

Table 6. **Ablations for every proposed regularization**. We report mIoU for 5-shot part label transfer on ShapeNet chair. Note that $\mathcal{L}_{pdc}$ denotes $\mathcal{L}_{pdc\_geo} + \mathcal{L}_{pdc\_sem}$.

	$\mathcal{L}_{pdc}$	$\mathcal{L}_{geo}$	$\mathcal{L}_{scale}$	chair
(a)	✓			83.4
(b)	✓	✓		83.9
(c)	✓	✓	✓	**84.2**

Table 7. **BAE-Net k ablation.** We report mIoU for 5-shot part label transfer on ShapeNet chair over 3 runs.

	$k=2$	$k=4$	$k=6$	$k=8$	$k=10$
chair	$83.8_{\pm0.51}$	$83.5_{\pm1.48}$	$82.8_{\pm1.27}$	$83.6_{\pm0.85}$	$83.7_{\pm1.33}$

regularization. We conduct 5-shot part label transfer within a chair category, as same as Section 5.1. Table 6 shows that each component of our proposed regularizers contributes to performance improvement for dense correspondence. Note that by only using $\mathcal{L}_{pdc}$, *i.e.*, using $\mathcal{L}_{pdc\_geo}$ and $\mathcal{L}_{pdc\_sem}$, it still outperforms than baselines (80.4 [1] and 82.1 [2]). The best performance was achieved by incorporating all proposed regularizations.

BAE-Net [31] with different k. We evaluate the effect of k for BAE-Net [31] with mean/standard deviation over 3 runs for ShapeNet [50] chair in We train BAE-Net [31] with a range of $k = (2, 4, 6, 8, 10)$, where $k = 4$ is used in the main paper, and report mIoU for 5-shot part label

transfer task. Table 7 shows the efficacy of semantic priors, where our framework achieves the average mIoU of 83.5, which shows higher performance than baselines. More importantly, we observed that the performance is fairly robust over k.

6. Conclusion

We have proposed a new framework that learns a semantic-aware implicit template field. Our framework extracts semantic knowledge from a self-supervised feature extractor given unlabelled point clouds. By leveraging the semantic information, we propose novel conditioning with semantic-aware deformation code and part deformation consistency regularizations to encourage semantically plausible deformation. Furthermore, we suggest global scaling consistency and global deformation consistency regularizations to enhance the learning process. Through quantitative and qualitative evaluation, we demonstrate the superiority of our proposed method and verified the effect of all proposed components in our framework.

References

[1] Zerong Zheng, Tao Yu, Qionghai Dai, and Yebin Liu. Deep implicit templates for 3d shape representation. In *CVPR*, 2021. 1, 2, 3, 5, 6, 7, 8

[2] Yu Deng, Jiaolong Yang, and Xin Tong. Deformed implicit field: Modeling 3d shapes with learned dense correspondence. In *CVPR*, 2021. 1, 2, 3, 5, 6, 7, 8

[3] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. SMPL: a skinned multiperson linear model. *TOG*, 2015. 1, 2

[4] Brett Allen, Brian Curless, and Zoran Popović. Articulated body deformation from range scan data. *TOG*, 2002. 1

[5] Thibault Groueix, Matthew Fisher, Vladimir G. Kim, Bryan C. Russell, and Mathieu Aubry. 3d-coded: 3d correspondences by deep deformation. In *ECCV*, 2018. 1, 2

[6] Hanbyul Joo, Tomas Simon, and Yaser Sheikh. Total capture: A 3d deformation model for tracking faces, hands, and bodies. In *CVPR*, 2018. 1

[7] Wang Zeng, Wanli Ouyang, Ping Luo, Wentao Liu, and Xiaogang Wang. 3d human mesh regression with dense correspondence. In *CVPR*, 2020. 1, 2

[8] Sing Chun Lee and Misha Kazhdan. Dense point-to-point correspondences between genus-zero shapes. In *Computer Graphics Forum*, 2019. 1, 2

[9] Silvia Zuffi, Angjoo Kanazawa, David W. Jacobs, and Michael J. Black. 3d menagerie: Modeling the 3d shape and pose of animals. In *CVPR*, 2017. 1, 2

[10] Pascal Paysan, Reinhard Knothe, Brian Amberg, Sami Romdhani, and Thomas Vetter. A 3d face model for pose and illumination invariant face recognition. In *AVSS*, 2009. 1

[11] Tianye Li, Timo Bolkart, Michael J Black, Hao Li, and Javier Romero. Learning a model of facial shape and expression from 4d scans. *TOG*, 2017. 1

[12] Luca Ballan, Aparna Taneja, Jürgen Gall, Luc Van Gool, and Marc Pollefeys. Motion capture of hands in action using discriminative salient points. In *ECCV*, 2012. 1

[13] Srinath Sridhar, Antti Oulasvirta, and Christian Theobalt. Interactive markerless articulated hand motion tracking using rgb and depth data. In *ICCV*, 2013. 1

[14] Iason Oikonomidis, Manolis IA Lourakis, and Antonis A Argyros. Evolutionary quasi-random search for hand articulations tracking. In *CVPR*, 2014. 1

[15] Shubham Tulsiani, Hao Su, Leonidas J. Guibas, Alexei A. Efros, and Jitendra Malik. Learning shape abstractions by assembling volumetric primitives. In *CVPR*, 2017. 1, 2

[16] Vignesh Ganapathi-Subramanian, Olga Diamanti, Soeren Pirk, Chengcheng Tang, Matthias Niessner, and Leonidas Guibas. Parsing geometry using structure-aware shape templates. In *3DV*, 2018. 1, 2

[17] Thibault Groueix, Matthew Fisher, Vladimir G. Kim, Bryan Russell, and Mathieu Aubry. AtlasNet: A Papier-Mâché Approach to Learning 3D Surface Generation. In *CVPR*, 2018. 1, 2

[18] Pablo Palafox, Nikolaos Sarafianos, Tony Tung, and Angela Dai. Spams: Structured implicit parametric models. In *CVPR*, 2022. 1

[19] Ramana Sundararaman, Gautam Pai, and Maks Ovsjanikov. Implicit field supervision for robust non-rigid shape matching. In *ECCV*, 2022. 1, 8

[20] Yiheng Xie, Towaki Takikawa, Shunsuke Saito, Or Litany, Shiqin Yan, Numair Khan, Federico Tombari, James Tompkin, Vincent Sitzmann, and Srinath Sridhar. Neural fields in visual computing and beyond. In *Computer Graphics Forum*, 2022. 1, 2

[21] Zhiqin Chen and Hao Zhang. Learning implicit fields for generative shape modeling. 2019. 1, 3

[22] Lars Mescheder, Michael Oechsle, Michael Niemeyer, Sebastian Nowozin, and Andreas Geiger. Occupancy networks: Learning 3d reconstruction in function space. In *CVPR*, 2019. 1, 2, 3

[23] Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. Deepsdf: Learning continuous signed distance functions for shape representation. In *CVPR*, 2019. 1, 2, 3

[24] Songyou Peng, Michael Niemeyer, Lars Mescheder, Marc Pollefeys, and Andreas Geiger. Convolutional occupancy networks. In *ECCV*, 2020. 1

[25] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *ECCV*, 2020. 1, 2

[26] Ricardo Martin-Brualla, Noha Radwan, Mehdi SM Sajjadi, Jonathan T Barron, Alexey Dosovitskiy, and Daniel Duckworth. Nerf in the wild: Neural radiance fields for unconstrained photo collections. In *CVPR*, 2021. 1, 2

[27] Keunhong Park, Utkarsh Sinha, Jonathan T Barron, Sofien Bouaziz, Dan B Goldman, Steven M Seitz, and Ricardo Martin-Brualla. Nerfies: Deformable neural radiance fields. In *CVPR*, 2021. 1, 2

[28] Shunsuke Saito, Zeng Huang, Ryota Natsume, Shigeo Morishima, Angjoo Kanazawa, and Hao Li. Pifu: Pixel-aligned implicit function for high-resolution clothed human digitization. In *ICCV*, 2019. 1, 2

[29] Tarun Yenamandra, Ayush Tewari, Florian Bernard, Hans-Peter Seidel, Mohamed Elgharib, Daniel Cremers, and Christian Theobalt. i3dmm: Deep implicit 3d morphable model of human heads. In *CVPR*, 2021. 1, 2

[30] Shunsuke Saito, Tomas Simon, Jason Saragih, and Hanbyul Joo. Pifuhd: Multi-level pixel-aligned implicit function for high-resolution 3d human digitization. In *CVPR*, 2020. 1, 2

[31] Zhiqin Chen, Kangxue Yin, Matthew Fisher, Siddhartha Chaudhuri, and Hao Zhang. Bae-net: Branched autoencoder for shape co-segmentation. In *ICCV*, 2019. 2, 4, 6, 8

[32] Vladimir G Kim, Wilmot Li, Niloy J Mitra, Siddhartha Chaudhuri, Stephen DiVerdi, and Thomas Funkhouser. Learning part-based templates from large collections of 3d shapes. *TOG*, 2013. 2

[33] Kyle Genova, Forrester Cole, Daniel Vlasic, Aaron Sarna, William T. Freeman, and Thomas A. Funkhouser. Learning shape templates with structured implicit functions. In *ICCV*, 2019. 2

[34] Theo Deprelle, Thibault Groueix, Matthew Fisher, Vladimir Kim, Bryan Russell, and Mathieu Aubry. Learning elementary structures for 3d shape generation and matching. In *NeurIPS*, 2019. 2, 6

[35] Jiahui Lei and Kostas Daniilidis. Cadex: Learning canonical deformation coordinate space for dynamic surface representation via neural homeomorphism. In *CVPR*, 2022. 2

[36] Ka-Hei Hui, Ruihui Li, Jingyu Hu, and Chi-Wing Fu. Neural template: Topology-aware reconstruction and disentangled generation of 3d meshes. In *CVPR*, 2022. 2

[37] Zekun Hao, Hadar Averbuch-Elor, Noah Snavely, and Serge Belongie. Dualsdf: Semantic shape manipulation using a two-level representation. In *CVPR*, 2020. 2

[38] Yueqi Duan, Haidong Zhu, He Wang, Li Yi, Ram Nevatia, and Leonidas J. Guibas. Curriculum deepsdf. In *ECCV*, 2020. 2

[39] Jiatao Gu, Lingjie Liu, Peng Wang, and Christian Theobalt. Stylenerf: A style-based 3d aware generator for high-resolution image synthesis. In *ICLR*, 2022. 2

[40] Michael Niemeyer and Andreas Geiger. Giraffe: Representing scenes as compositional generative neural feature fields. In *CVPR*, 2021. 2

[41] Michael Niemeyer and Andreas Geiger. Campari: Camera-aware decomposed generative neural radiance fields. In *3DV*, 2021. 2

[42] Miguel Angel Bautista, Pengsheng Guo, Samira Abnar, Walter Talbott, Alexander Toshev, Zhuoyuan Chen, Laurent Dinh, Shuangfei Zhai, Hanlin Goh, Daniel Ulbricht, et al. Gaudi: A neural architect for immersive 3d scene generation. In *NeurIPS*, 2022. 2

[43] Chengjie Niu, Manyi Li, Kai Xu, and Hao Zhang. Rim-net: Recursive implicit fields for unsupervised learning of hierarchical shape structures. In *CVPR*, 2022. 2

[44] Chenyang Zhu, Kai Xu, Siddhartha Chaudhuri, Li Yi, Leonidas J Guibas, and Hao Zhang. Adacoseg: Adaptive shape co-segmentation with group consistency loss. In *CVPR*, 2020. 2

[45] Minghua Liu, Yinhao Zhu, Hong Cai, Shizhong Han, Zhan Ling, Fatih Porikli, and Hao Su. Partslip: Low-shot part segmentation for 3d point clouds via pretrained image-language models. In *CVPR*, 2023. 2

[46] Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *ICCV*, 2021. 2

[47] Alex Yu, Vickie Ye, Matthew Tancik, and Angjoo Kanazawa. pixelnerf: Neural radiance fields from one or few images. In *CVPR*, 2021. 2

[48] Sosuke Kobayashi, Eiichi Matsumoto, and Vincent Sitzmann. Decomposing nerf for editing via feature field distillation. In *NeurIPS*, 2022. 2

[49] Can Wang, Menglei Chai, Mingming He, Dongdong Chen, and Jing Liao. Clip-nerf: Text-and-image driven manipulation of neural radiance fields. In *CVPR*, 2022. 2

[50] 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:1512.03012*, 2015. 4, 6, 7, 8

[51] David Ha, Andrew M. Dai, and Quoc V. Le. Hypernetworks. In *ICLR*, 2017. 4

[52] Vincent Sitzmann, Julien N.P. Martel, Alexander W. Bergman, David B. Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. In *NeurIPS*, 2020. 4

[53] Mikaela Angelina Uy, Quang-Hieu Pham, Binh-Son Hua, Duc Thanh Nguyen, and Sai-Kit Yeung. Revisiting point cloud classification: A new benchmark dataset and classification model on real-world data. In *ICCV*, 2019. 5, 7

[54] Amos Gropp, Lior Yariv, Niv Haim, Matan Atzmon, and Yaron Lipman. Implicit geometric regularization for learning shapes. In *ICML*, 2020. 5

[55] Yang You, Yujing Lou, Chengkun Li, Zhoujun Cheng, Liangwei Li, Lizhuang Ma, Cewu Lu, and Weiming Wang. Keypointnet: A large-scale 3d keypoint dataset aggregated from numerous human annotations. In *CVPR*, 2020. 6

[56] Li Yi, Vladimir G. Kim, Duygu Ceylan, I-Chao Shen, Mengyan Yan, Hao Su, Cewu Lu, Qixing Huang, Alla Sheffer, and Leonidas J. Guibas. A scalable active framework for region annotation in 3d shape collections. *TOG*, 2016. 6

[57] Li Yi, Hao Su, Xingwen Guo, and Leonidas Guibas. Syncspeccnn: Synchronized spectral cnn for 3d shape segmentation. In *CVPR*, 2017. 7

[58] Xinhai Liu, Zhizhong Han, Yu-Shen Liu, and Matthias Zwicker. Fine-grained 3d shape classification with hierarchical part-view attention. *IEEE Transactions on Image Processing*, 2021. 8

[59] Federica Bogo, Javier Romero, Gerard Pons-Moll, and Michael J. Black. Dynamic FAUST: Registering human bodies in motion. In *CVPR*, 2017. 8