2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 14-18, 2024. Abu Dhabi, UAE

NRDF - Neural Region Descriptor Fields as Implicit ROI Representation for Robotic 3D Surface Processing

Anish Pratheepkumar[1,2], Markus Ikeda[1], Michael Hofmann[1], Fabian Widmoser[1], Andreas Pichler[1], Markus Vincze[2]

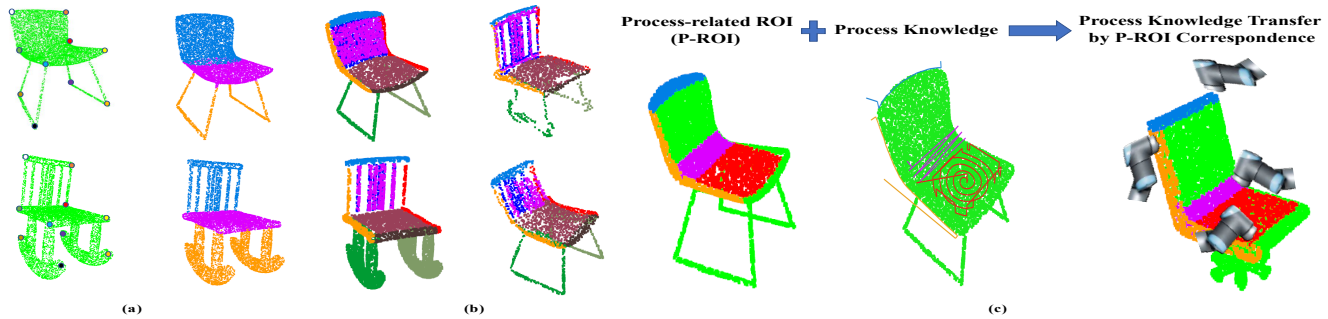

Fig. 1: (a) Conventional correspondence estimations: (left) keypoint correspondence, (right) semantic part correspondence. (b) Proposed arbitrary region of interest (ROI) correspondence: (left) arbitrary ROI on reference object and (right) the NRDF estimated corresponding ROI on target object. (c) Proposed concept of process knowledge transfer with P-ROI correspondence: (left) four P-ROI on a reference chair object, (middle) specific process strategy assigned to each P-ROI (trajectory illustrated in the same color of the sub regions), and (right) when a new instance of category-level target object is presented the same process strategy associated with the corresponding P-ROI is executed.

***Abstract*— To automate 3D surface processing across diverse category-level objects it is imperative to represent *process-related region of interest* (P-ROI), which is not obtained with conventional keypoint or semantic part correspondences. To resolve this issue, we propose Neural Region Descriptor Fields (NRDF) for achieving unsupervised dense 3D surface region correspondence such that arbitrary ROI is retrieved for a new instance of a known category of object. We utilize the NRDF representation as a medium to facilitate one-shot P-ROI level process knowledge transfer. Recent developments in implicit 3D object representations have focused on keypoint or part correspondences, which have resulted in applications like robotic grasping and manipulation. However, explicit one-shot P-ROI correspondence, and its application for 3D surface process knowledge transfer, is treated for the first time in this work, to the best of our knowledge. The evaluation results show that the proposed approach outperforms the dense correspondence baselines in implicit shape representation and the capacity to retrieve matching arbitrary ROIs. In addition, we validate the practicality of our proposed system in a real-world robotic surface processing application. Our code is available at https://github.com/Profactor/Neural-Region-Descriptor-Fields.**

I. Introduction

Surface processing is an indispensable aspect of manufacturing, spanning a multitude of industries such as furniture, automobile, aerospace, mold, etc., and encompassing operations such as polishing, oiling, cleaning, painting, etc., [1]–[3]. There is a rich history of research addressing multiple aspects of robotic surface processing, including process strategy [2]–[4], focusing on specific applications [5]–[7], concerning specific surface types [7], [8], etc. However, it is still challenging to generalize robotic surface processing, and the task continues to be performed manually by relying on skilled human workers [8]–[10].

The recent transition from mass production to mass customization leads to topology variations in category-level objects, resulting in a high-mix, low-volume scenario [11], [12]. Hence, we probe into the specific question of how to generalize robotic 3D surface processing across diverse category-level objects. There are multiple existing and currently researched methodologies focusing on how to perform a robotic surface process execution [2]–[4]. In contrast, our focus is on how to introduce a generalization approach that facilitates process knowledge transfer across diverse category-level objects. This avoids the need for reprogramming, encouraging flexible automation of robotic surface processing.

It is intuitive that a surface process is executed on sub regions or P-ROIs such that a particular P-ROI is associated with a specific process strategy. The process strategy involves multiple aspects such as process trajectory, direction, contact force, tool angle, etc., which we refer to as the process knowledge. From this perspective, we propose to approach generalizing category-level surface processing by identifying corresponding P-ROI which share a similar process strategy. Hence, by defining P-ROIs and its associated process strategy for one object, a region correspondence system will facilitate identifying corresponding P-ROIs on a new instance of the object, and enable transfer of the process knowledge, an example is shown in Fig. 1c.

Conventional correspondence estimations focus on dis-

*This work was accomplished within the Lighthouse project supported by the Austrian Institute of Technology (AIT).

[1]Profactor Gmbh, Im Stadtgut D1, 4407 Steyr-Gleink, Austria. {anish.pratheepkumar, markus.ikeda, michael.hofmann, fabian.widmoser, andreas.pichler} @profactor.at

[2]Vision for Robotics Laboratory, Automation and Control Institute, TU Wien, 1040 Vienna, Austria. vincze@acin.tuwien.ac.at

crete keypoints or object semantic parts as shown in Fig. 1a. It is evident that keypoints alone are insufficient for defining a surface process strategy. Also, on a semantic part of an object, it is challenging to define a single surface processing strategy as it is possible to have multiple P-ROIs within a semantic part. Moreover, the P-ROIs could span across multiple semantic parts, for example, the orange P-ROI in Fig. 1c is spanning across the two blue and pink semantic parts as seen in Fig. 1a. It could even be at the intersection of two semantic parts, for example, the pink P-ROI in Fig. 1c is at the intersection of the two blue and pink semantic parts of the chair as in Fig. 1a. These aspects justify the need for specific P-ROI correspondence where the boundaries of the region are not constrained to the component parts.

However manually annotating and learning category-level P-ROIs for specific surface processing operation is time consuming, and the P-ROIs could vary based on the type of operation executed or with the type of tool adopted. In contrast, a more generalized approach would be to develop a dense correspondence system which is capable of retrieving arbitrary ROIs across category-level objects in a one-shot manner such that any P-ROI configuration could be defined and retrieved as needed.

Compared to conventional 3D representations such as voxels, meshes and point clouds, implicit object representations [13], [14] have recently emerged as an effective approach [15]–[17]. A recent work [18] introduce an implicit Neural Descriptor Field (NDF) representation, which enables category-level dense correspondence estimation via few-shot iterative optimization, and applies the technology for point based robotic manipulation. For this work we leverage the implicit descriptor [18] representation and advance it from few-shot iterative point correspondence estimations to one-shot non-iterative surface region level correspondence estimations for retrieving any random ROI on the object.

The following is a summary of our contributions:

1) We introduce one-shot implicit descriptor-based 3D object dense correspondence system capable of retrieving arbitrary ROIs on object surface, and then utilize it for process knowledge transfer with P-ROI correspondence across category-level objects.
2) To achieve the arbitrary ROI retrieval, we propose the new NRDF representations which essentially optimizes the implicit descriptors to explicitly consider the surface geometry variations in 3 dimensions. To this end, we design a novel inverse descriptor function capable of mapping descriptors back to the 3D space locations.
3) In addition, we introduce a new descriptor loss function that focuses on optimizing the descriptor space such that the descriptors of the corresponding object surface regions are as similar as possible.
4) The evaluations considering shape representation capacity and arbitrary ROI retrieval capacity show superiority of the proposed approach over the baselines. Furthermore, we demonstrate the proposed concept of process knowledge transfer with P-ROI correspondence in a real world surface processing experiment.

II. Related Work

To achieve the objective of category-level generalization in 3D object surface processing we leverage 3D object dense correspondence. Accordingly, we split the prior work into *Generalizeable 3D Object Surface Processing* and *3D Object Dense Correspondence*.

A. Generalizeable 3D Object Surface Processing

Multiple works have studied generalization approaches to robotic surface processing. Early works rely on known CAD models [19] with some focusing specially on mass production scenarios [20]. Recent works follow a Learning from Demonstration (LfD) approach which involve a skilled operator performing a surface processing operation, and utilizing it as a learning resource for the robot to later perform the task autonomously. An approach of kinesthetic teaching for transferring tool trajectory and contact forces of a grinding operation by ensuring the surface quality was addressed in [21], [22]. Even though the approaches mention transfer of skills to new geometries, the study is limited to transfer between simple shapes; planar surface to a cylindrical surface. Many LfD approaches [23]–[25] focusing on human skill transfer with an impedance control system for simultaneous force and position control are proposed. However the the approaches are evaluated on specific surfaces, and does not consider its scalability to category-level surface processing which is in demand with the current mass customization trend in manufacturing [11], [12].

To the best of our knowledge, there is no existing approaches which explicitly study the surface processing generalization on a category level. Meanwhile, category-level approaches are widely studied for robotic grasping and manipulation tasks. Transfer of category-based functional grasping skills by latent space non-rigid registration is studied in [26]. Florence et al. [27] propose using dense visual 2D descriptors as a representation for robotic manipulation, and Simenonov et al. [18] develop the concept of NDF descriptor for category-level 3D dense correspondence estimation for manipulation. Thomson et al. [28] propose manipulation skill transfer to novel objects, considering categorical shape variation. In this work, we propose to extend such category-level task generalization to robotic surface processing.

B. 3D Object Dense Correspondence

3D object dense correspondences are approached mainly from three perspectives; *keypoint-based*, *part-based* and *ROI-based*. Here we treat the related works from all three approaches since the techniques of correspondence estimations are potentially used interchangeably. Prior approaches target correspondence estimation in a supervised and unsupervised manner [29], here we focus on the unsupervised approaches.

The majority of the earlier *Keypoint-based* correspondence works follow an iterative approach [30], [31], and utilize either handcrafted feature descriptors or template deformations. A recent geometric deep learning approach [32] uses a set of discrete 3D structure point representation for *keypoint-based* and *part-based* correspondence. However, if

a denser correspondence is needed, retraining with modified structure point dimension is mandatory. A branched auto encoder approach [33] learns *part-based* correspondence by adopting a multiple branched output for the implicit decoder, such that each branch predicts the implicit boundary of a specific component part of the object. An implicit network for *keypoint-based* correspondence was recently introduced in [34] where the feature points on the object is modeled as a signed distance field (SDF) related implicit spheres. Another set of work [29], [35] uses a combination of self-reconstruction and cross-reconstruction losses to enable category-level correspondences with an implicit network using [33] as a base. The approach targets both *keypoint-based* and *part-based* correspondences. However, the prior approaches are either limited to discrete points or not inherently devised to provide arbitrary ROI correspondence, and the evaluations and validations are performed mainly on semantic part components (which are fixed regions and not arbitrary in nature) or on discrete point pairs.

ROI-based correspondence focuses on the sub regions, and is different with respect to semantic part-based correspondence [36]. Not many works have targeted this goal in the literature. A fuzzy correspondence computation to estimate similar ROI is proposed in [37]. Denitto et al. propose a biclustering approach [36] to compute ROI correspondences. A computation of stable ROI correspondences across non-isometric shapes are considered in [38]. However, the prior ROI correspondence works consider handcrafted feature vectors, and the correspondence computations are iterative and computationally expensive, hindering direct applications in real robotic systems. In contrast, we propose a dense implicit representation approach which encodes any ROI on a reference object as an implicit NRDF which is directly used to estimate the corresponding ROI on the target object by a simple forward pass on an inverse descriptor function.

III. BACKGROUND: NEURAL DESCRIPTOR FIELDS

The recently introduced NDF [18] representations encode point level dense correspondences across category-level objects. The approach involves a neural descriptor function F which maps any query point $p_i \in \mathbb{R}^3$ associated with a 3D object point cloud $P \in \mathbb{R}^{n\times 3}$ to a spacial descriptor $k_i \in \mathbb{R}^d$:

$$F : \mathbb{R}^3 \times \mathbb{R}^{n\times 3} \rightarrow \mathbb{R}^d \tag{1}$$

The function F is developed by training a neural shape encoder-decoder network with the objective of predicting the binary occupancy on the input object point cloud for any 3D query point p. Essentially, predicting whether a given query point is occupied within the object surface boundary or not. After training, the spacial descriptor k_i is extracted for a 3D query point p_i with respect to the object geometry by concatenating the intermediate activations of the multilayer perceptron (MLP) decoder network. The interesting property of such an extracted point level descriptor is that it is similar for corresponding points in the category-level objects, which facilitates dense correspondence.

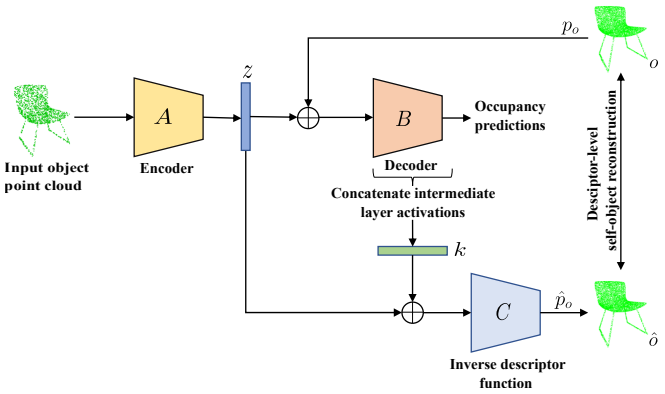

Fig. 2: NRDF architecture diagram: An object point cloud input to the shape encoder A gives a shape embedding z as output. The MLP decoder B in conjunction with A is trained to predict for any query 3D point $p \in \mathbb{R}^3$ the binary occupancy on the object surface, and the intermediate layer activations of B provides the query point level NDF descriptors k. The proposed inverse descriptor function C, with k and z as inputs, map the descriptors back to 3D space. Here we also illustrate the descriptor-level self-object reconstruction which recovers query points p_o sampled on the object o surface S.

IV. NEURAL REGION DESCRIPTOR FIELDS

In Section IV-A, we initially introduce the NRDF system architecture, and the design of novel inverse descriptor function. Subsequently, we define the adopted approach to optimize descriptors to explicitly consider the surface geometry variations in 3D. Inspired from [29], [32], we perform this optimization by adding an adapted self-reconstruction and cross-reconstruction loss to the NRDF training along with the newly introduced descriptor loss function. Then we explain how to facilitate arbitrary ROI correspondence with such optimized NRDF descriptors in Section IV-B. Finally, the proposed concept of process knowledge transfer with P-ROI correspondence is discussed in Section IV-C.

A. Approach

The proposed end to end network architecture is illustrated in Fig. 2, here we consider a point cloud $P \in \mathbb{R}^{n\times 3}$ input space for the object. In contrast to the NDF [18], we employ a parallel ensemble of Residual Network (ResNet) [39] integrated PointNet [40] and a 1D convolution integrated PointNet as the shape encoder A to generate a shape embedding z. The decoder B is an MLP which along with A trains on object boundary occupancy data, such that for any 3D query point $p_i \in \mathbb{R}^3$ its binary occupancy is predicted, mathematically:

$$A : \mathbb{R}^{n\times 3} \rightarrow z,$$
$$B : \mathbb{R}^3 \times z \rightarrow [0, 1]$$

The desired query point level adapted NDF [18] descriptor k_i for the object geometry is then extracted by the intermediate layer activation concatenations of B, defined as a descriptor function $\theta\,(z, p_i) = k_i$.

a) Inverse Descriptor Function: The inverse descriptor function C is designed to map the feature descriptors k back to the 3D space. As shown in Fig. 2, C takes the adapted NDF feature descriptors k of a set of surface query points p_o, and the shape embedding z as input. The concatenated result

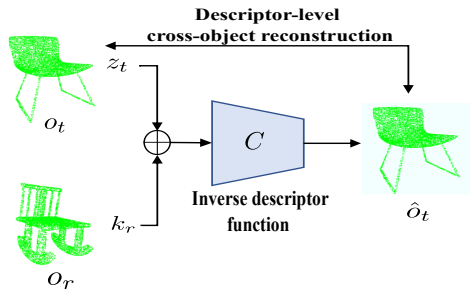

Fig. 3: Descriptor-level cross-object reconstruction. The inverse descriptor function C retrieves the target object $\hat{o_t}$ by taking the target object o_t embedding z_t and the reference object descriptors k_r as input.

of descriptors k and shape embedding z is passed to an MLP with 3 output neurons. Next, we discuss the approach and the related loss functions used to enable the prediction of the associated 3D coordinate points $\hat{p_o}$ on the object surface.

Descriptor-Level Self-Object Reconstruction Given a set of query points p_o sampled on the object o surface S boundary, we propose to optimize C such that it maps the implicit descriptors of p_o back to the 3D space. The function C inherently is not able to predict the associated point coordinates. Hence, to enable the desired state of $\hat{p_o} \approx p_o$ we introduce descriptor-level self-object reconstruction. This is realised by training C on a descriptor-level self reconstruction loss function which minimizes the mean squared error (MSE) between the actual input object surface points p_o and predicted points $\hat{p_o}$ defined as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(C\left(z_o, k_{oi}\right) - p_{oi}\right)^2, \tag{2}$$

where n indicates the number of points sampled on the object geometry surface, and o indicates the object considered. The network when trained with a combination of descriptor-level self-object reconstruction and the occupancy loss function (mean squared error in predicting the binary occupancies), it injects explicitly the spacial distribution of object geometry to the generated feature descriptors k.

Descriptor-Level Cross-Object Reconstruction We propose to further optimize the descriptors to provide the capacity to blend smoothly across diverse objects within a category, facilitating correspondence mapping between object ROIs. To this end, we introduce the descriptor-level cross-object reconstruction wherein given any two category-level objects in the training data, C predicts the 3D points corresponding to the locations on the object, irrespective from which object the descriptors were extracted. As shown illustratively in Fig. 3, we propose to optimize the descriptors to converge such that given a shape embedding z_t of any target object o_t and the optimized descriptors k_r of reference object o_r surface S_r, C approximate the corresponding target object o_t surface S_t. For this optimization we exploit the continuous mapping between the descriptor space and category-level object surface points. This is enforced by aligning reference and target objects at a descriptor level by application of reconstruction loss. The cross reconstruction loss includes a combination of multiple losses, similar to the approach in [29], we consider minimizing losses pertaining to the following three standard shape similarity enforcing measures such as Chamfer Distance (CD) [41], Earth Movers Distance (EMD) [41], [42] and Normal Consistency Distance (NCD); each targeting specific aspects of the geometry as explained below. Also, an adapted form of smoothness loss [29] as a Descriptor Smoothness (DS) loss is adopted to reinforce the descriptor-level correspondence across the category-level object surface. In contrast to component part embedding level reconstruction in [29], we perform a descriptor-level reconstruction to align the corresponding object surface points in a category, targeting a new objective of arbitrary ROI correspondence estimation.

The CD is computed as mean distance between the actual target object surface points p_t and the nearest neighbour of p_t in the predicted surface points $\hat{p_t}$ which ensures similarity accuracy; its consistency is ensured by computing the same in the reverse direction i.e., from $\hat{p_t}$ to p_t, defined as:

$$CD = \frac{1}{n}\sum_{i=1}^{n}\left\|p_{ti} - N_{\hat{p_t}}\left(p_{ti}\right)\right\|_2^2 + \frac{1}{n}\sum_{i=1}^{n}\left\|\hat{p_{ti}} - N_{p_t}\left(\hat{p_{ti}}\right)\right\|_2^2, \tag{3}$$

where $\hat{p_{ti}}$ indicate the surface point predicted by C and is defined as $C\left(z_t, k_{ri}\right)$, the nearest neighbour of $\hat{p_{ti}}$ in the target object surface points set p_t is indicated as $N_{p_t}\left(\hat{p_{ti}}\right)$, and $N_{\hat{p_t}}\left(p_{ti}\right)$ indicates the nearest neighbour of target object surface point in the predicted points set.

The EMD distance performs a finer similarity computation considering the local density distribution of the two finite point sets having same cardinality. This is ensured by initially solving an assignment problem that accounts for a one to one correspondence between the predicted surface points $\hat{p_t}$ and the target surface points p_t essentially resulting in a bijective mapping $\phi : \hat{p_t} \rightarrow p_t$. The EMD distance is defined as:

$$EMD = \frac{1}{n}\sum_{i=1}^{n}\left\|\hat{p_{ti}} - \phi\left(\hat{p_{ti}}\right)\right\|_2 \tag{4}$$

The NCD distance compliments the system by providing discriminative knowledge which supports C in distinguishing between surface points with respect to the relative local positions. For example, the normals of the surface points on a chair object's seat, back rest front portion and backrest back portion are all different. We compute NCD based on the cosine similarity distance between neighbouring point normal vectors of the predicted $\hat{p_t}$ and the target surface points p_t, and vice versa, defined as:

$$NCD = \frac{1}{n}\sum_{i=1}^{n}\left(1 - \frac{n_{\hat{p_{ti}}} \cdot N_{p_t}\left(n_{\hat{p_{ti}}}\right)}{\left\|n_{\hat{p_{ti}}}\right\|_2 \left\|N_{p_t}\left(n_{\hat{p_{ti}}}\right)\right\|_2}\right) + \frac{1}{n}\sum_{i=1}^{n}\left(1 - \frac{n_{p_{ti}} \cdot N_{\hat{p_t}}\left(n_{p_{ti}}\right)}{\left\|n_{p_{ti}}\right\|_2 \left\|N_{\hat{p_t}}\left(n_{p_{ti}}\right)\right\|_2}\right), \tag{5}$$

where the $n_{p_{ti}}$ and $n_{\hat{p_{ti}}}$ indicate the normals of surface point p_{ti} and $\hat{p_{ti}}$ respectively.

The DS loss is computed by estimating consistency in the difference between the reference object surface points p_r and the predicted target surface points $\hat{p}_t$ corresponding to the reference descriptor k_r in random patch neighbourhood of the reference object, defined as:

$$DS = \sum_{v,v' \in \Omega(u)} \left(1 - d_{cos}\left(\hat{p_t}^{(v)} - {p_r}^{(v)}, \hat{p_t}^{(v')} - {p_r}^{(v')}\right)\right), \tag{6}$$

where u is the multiple random patch centers and $\Omega(u)$ indicate the neighbourhood point indices in the patches, with v' representing the permuted neighbours in v. The DS loss ensures that the descriptors in the local neighbourhood of reference object o_r is responsible for reconstructing the corresponding local neighbourhood of the target object $\hat{o}_t$ by minimising the distance based on the cosine similarity d_{cos} between the local neighbourhood difference vectors. It ensures that during the cross-reconstruction from the reference to the target object, the reference point indices do not drift too far from its neighbourhood. This allows for an index level correspondence between reference and target by enforcing a smoother optimization, for example, preventing the descriptors from the left part of reference object to predict the right part of the target object, which similarity loss measures alone do not guarantee.

In association with prior mentioned loss functions, we introduce a novel *descriptor loss function* which promotes the descriptor space to blend smoothly between diverse objects in a category such that the resulting optimised descriptor distribution is consistent with corresponding regions on the object surface. For this purpose, we minimize the descriptor distances between the surface points on a reference object and corresponding points on the target object. To this end, we exploit the point correspondence inference scheme of [29] with which for each point p_{ti} on the target object we approximate the corresponding point $\tilde{p}_{ri}$ on the reference object. Which is given by the index of the nearest neighbour of the point p_{ti} in the points $\hat{p}_t$ predicted by the inverse descriptor function C. Once the corresponding point descriptors are computed we minimize the distance between them as:

$$DD = \frac{1}{n}\sum_{i=1}^{n} \left\| \theta\left(z_r, \tilde{p_{ri}}\right) - \theta\left(z_t, p_{ti}\right) \right\|_2^2 \tag{7}$$

The combined minimization of the binary occupancy, self-object reconstruction and cross-object reconstruction losses results in an optimized descriptor function which facilitates the descriptor space to be distributed in a manner that the resulting descriptors k of corresponding surface regions on the category-level objects are as identical as feasible.

B. Arbitrary ROI Correspondence with NRDF

With the optimized descriptors k generated by the combined self-object and cross-object reconstruction settings, we propose to represent any ROI on the object surface as an implicit NRDF $\Re$. If the arbitrary ROI $s_r \subsetneq S_r$ on the reference object is represented as a set of surface points $\bar{p}_r$,

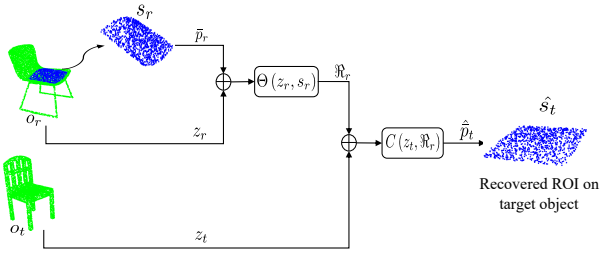

Fig. 4: Illustration of arbitrary ROI correspondence estimation with NRDF.

then $\Re_r$ is defined as the concatenation $\bigoplus$ of the optimized descriptors of $\bar{p}_r$, defined as:

$$\Re_r = \Theta\left(z_r, s_r\right) = \bigoplus_{i=1}^{m} \theta\left(z_r, \bar{p}_{ri}\right), \tag{8}$$

where m indicates the number of points in $\bar{p}_r$. The region descriptor field has an interesting property that it is consistent across the category-level objects and hence we directly recover the corresponding ROI $\hat{s}_t$ on the target object by:

$$\hat{s}_t = C\left(z_t, \Re_r\right), \tag{9}$$

where $\hat{s}_t$ consists of a set of surface points $\hat{\bar{p}}_t \approx \bar{p}_t$. Hence, with only the reference NRDF $\Re_r$ and the target shape embedding z_t the target ROI $\hat{s}_t$ is recovered as shown in Fig. 4. The exact 3D points in an object point cloud could vary each time when sampled from a mesh surface or captured with a 3D sensor; we ensure the predictions are in alignment with the geometric mean center of the target object, and a simple nearest neighbour associates the ROI prediction points $\hat{\bar{p}}_t$ directly to the currently captured target object points.

C. Process Knowledge Transfer with P-ROI Correspondence

For a robotic 3D surface processing operation there are multiple process parameters $\mathbb{P}$ that define a process strategy for a particular P-ROI. The parameters could include the process trajectory, tool angle, contact force, process speed, process direction, etc., which we collectively refer to as the process knowledge. We propose to perform the surface region level process knowledge transfer across category-level objects by a one time definition of the P-ROIs and its associated process knowledge on a reference object. Such a definition is then recorded as a knowledge dictionary D that maps each P-ROI to its associated process knowledge:

$$D : \{s_r^j\}_{j=1}^{K} \rightarrow \mathbb{P}, \tag{10}$$

where K indicate the number of reference P-ROI s_r^j considered on the object, and $\mathbb{P}$ indicate the list of process parameters according to a predefined order. Then for a new instance of the category-level object we estimate the corresponding target P-ROI $\hat{s}_t^j$ with NRDF and execute the similar process as per the definition in the knowledge dictionary D. A pictorial illustration of the proposed concept is illustrated in Fig. 1c, where a specific process trajectory type is adapted and transferred to corresponding regions on a target object, we also demonstrate such an NRDF assisted real world robotic surface process execution, as discussed in Section V-D.

TABLE I: Shape representation capacity evaluation results.

Model	Chair		Car		Plane	
	CD-L1	MSE	CD-L1	MSE	CD-L1	MSE
NDF [18]	0.083	0.051	0.065	0.015	0.070	0.012
IMDC [29]	0.062	0.041	0.060	0.010	0.036	0.006
NRDF	**0.054**	**0.040**	**0.058**	**0.008**	**0.030**	**0.004**

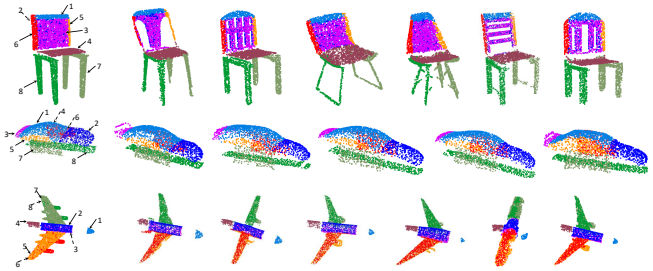

Fig. 5: Qualitative results for arbitrary ROI retrieval with NRDF. The first column shows the reference object and the 8 arbitrary ROIs (dashed lines indicate regions behind the current view). The corresponding ROI retrieved on different target objects are shown to the right.

V. Evaluation

In this section, we evaluate the effectiveness of our NRDF dense ROI correspondence system. Initially, the shape representation capacity is evaluated to verify how well proposed system has abstracted the category-level geometry. Then the arbitrary ROI retrieval capacity is evaluated quantitatively to verify how well the NRDF system estimates category-level arbitrary ROI correspondences, along with a qualitative visualization of the results. We further perform ablation analysis to investigate the effect of various loss functions adopted in the NRDF model development.

Dataset: We evaluate the proposed approach from a proof of concept basis by training on the shapenet dataset [13], [43] object categories such as Chair (furniture), Car (automobile), and Plane (aerospace). Each category has objects with similar part constituency; on average 500 in the training set and 50 in the test set. For the arbitrary ROI retrieval evaluation there is no existing dataset, and there is very little prior work that have explicitly targeted surface region correspondence. When it comes to overall dense correspondence itself there is no existing dataset with ground truths [29]. Hence we manually generated a dataset namely, BSRC (Benchmark for Surface Region Correspondence), with multiple consistent corresponding arbitrary ROIs, for 10 objects in each object category (the correspondences are approximations as a correspondence between topology varying objects in a category is not a concretely established concept and there is no explicit ground truths). A visualization of the multiple ROIs considered on each object category are shown in the first column of Fig. 5. For motivation of further research and experimentation we provide (https://github.com/aprath1/bsrc) the generated dataset.

Training Details: We perform the model training in a progressive manner [29], where initially the model is trained on occupancy loss alone which optimizes the encoder A and decoder B networks. Followed by this the combined system model (A, B and C) is trained for descriptor-level self-object reconstruction with equal weights for Occupancy and MSE losses. Finally, we introduce the descriptor-level cross-object reconstruction with joint minimization of occupancy, descriptor-level self-object reconstruction and cross-object reconstruction losses. However, [29] considers two random objects at a time in each iteration of their part embedding cross reconstruction training, in contrast, we consider a batch of objects in each iteration optimizing the descriptors within a collection of objects. We realize this by setting each object in a batch against a permuted order of the objects in the same batch, which also results in an improved training speed. Here we use an empirically chosen weighting scheme where losses for Occupancy, MSE, CD, EMD, NCD, DS and DD are set with weights 10, 1, 10, 1, 0.01, 0.1 and 0.1 respectively.

TABLE II: Arbitrary ROI retrieval evaluation for 8 different ROIs on each object category measured with the CD-L1 metric.

	Chair								
Model	1	2	3	4	5	6	7	8	Average
NDF [18]	0.138	0.141	0.131	0.124	0.361	0.326	0.305	0.312	0.230
IMDC [29]	0.024	0.068	0.046	0.042	0.067	0.074	0.094	0.090	0.063
NRDF	**0.025**	**0.029**	**0.042**	**0.029**	**0.051**	**0.047**	**0.048**	**0.044**	**0.039**
	Car								
Model	1	2	3	4	5	6	7	8	Average
NDF [18]	0.088	0.350	0.084	0.210	0.221	0.233	0.272	0.094	0.194
IMDC [29]	0.019	0.025	**0.031**	0.024	**0.021**	0.022	0.026	0.022	0.024
NRDF	**0.018**	**0.014**	**0.031**	**0.022**	**0.021**	**0.014**	**0.014**	**0.017**	**0.019**
	Plane								
Model	1	2	3	4	5	6	7	8	Average
NDF [18]	0.100	0.098	0.065	0.095	0.070	0.220	0.238	0.231	0.142
IMDC [29]	0.028	0.040	0.063	0.098	**0.033**	0.037	0.034	0.039	0.047
NRDF	**0.023**	**0.030**	**0.039**	**0.054**	0.034	**0.035**	**0.031**	**0.031**	**0.035**

Baselines: We compare the performance of NRDF against two dense correspondence baselines. The NDF [18] baseline represent dense point correspondence, and for ROI retrieval capacity evaluation, we adapt thier point inference scheme to predict ROIs. The IMDC [29] baseline is capable of both dense point and semantic part correspondence, and their correspondence inference scheme is directly used to evaluate its capacity to retrieve arbitrary ROI.

A. Shape Representation Capacity

The shape representation capacity is evaluated by estimating the 0.75 level set of the implicit field captured by the implicit decoder B, which shows how well the network abstracts the geometry. Then as a quantitative evaluation the resulting point cloud is compared with the actual input object point cloud using the Chamfer Distance L1 metric i.e., L1 version of (3). Table I shows the results of comparison between the NDF [18] and the IMDC [29] baselines, and the proposed NRDF model. We observe that the proposed model has consistent improvement in the shape representation for all three tested object categories. In essence, it is intuitive that the shape representation improves if the model has improved accuracy in predicting the boundary occupancy of the geometry. Hence, to validate this we examine the MSE error in the binary occupancy predictions for the trained models and the results are shown in Table I. Overall,

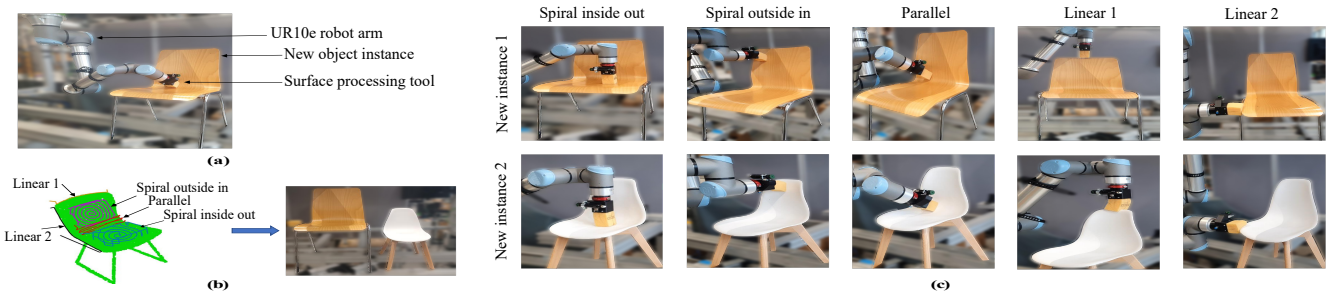

Fig. 6: (a) Robot environment setup. (b) Experiment: (left) 5 P-ROIs and the corresponding process trajectories defined on a virtual chair point cloud, (right) two new real chair instances to which the process knowledge need to be transferred. (c) Resulting surface process execution on the new object instances with NRDF assisted P-ROI correspondence based process knowledge transfer. Similar process states for each P-ROI on both the objects are shown.

from Table I we evidently observe that the proposed NRDF approach outperforms NDF and IMDC baselines in the shape representation capacity based on the CD-L1 metric. Furthermore, the MSE error rate in binary occupancy prediction is consistent with the CD-L1 which shows the effectiveness of our proposed approach.

B. Arbitrary ROI Retrieval Capacity

We evaluate the arbitrary ROI retrieval using the manually generated BSRC dataset. The quantitative evaluation is performed in a pairwise manner where multiple ROIs on category-level reference object is retrieved on a target object. Hence we compute for 10 objects in each category all possible 90 pair wise correspondence estimations. Furthermore, in each object we consider 8 arbitrary ROIs. It has to be noted that the proposed system is designed and intended to predict any corresponding ROI across category-level objects and is not limited to the 8 specific ROIs in the dataset. The arbitrary ROIs show cased in the dataset is only a representative of possible sub regions from all parts of the object to facilitate an evaluation of the ROI retrieval capacity.

To evaluate the ROI retrieval capacity, we compute CD-L1 distance error on each of the recovered region with respect to the ground truth, and a mean value is reported in Table II. Similar to prior shape representation evaluation, we compare the performance of the proposed NRDF approach with respect to the NDF [18] and IMDC [29] baselines. Here we observe the consistent superior performance of the proposed NRDF model with respect to the baseline NDF [18], for each region across all tested object categories. In comparison with the IMDC [29] model the proposed approach, on average, shows consistently improved error rates on ROI retrieval. Additionally, we show some qualitative results of arbitrary ROI retrieval with NRDF in Fig. 5.

C. Ablation Analysis

We also perform an ablation analysis as shown in Table III, to examine the impact of the adopted loss functions in optimizing the descriptors and achieving a descriptor-level dense ROI correspondence. Here the implicit models were trained separately omitting each loss, indicated as *w/o*. Then the average CD-L1 error for region retrieval on the complete data is evaluated. Here we see the best performance is achieved by a combination of the losses, with the geometrical similarity ensured by CD, EMD and NC distance measures, and the descriptor consistency and similarity enforced by the DS and DD losses respectively.

TABLE III: Average CD-L1 error rates in arbitrary ROI retrieval with different loss measures ablated.

Model	Chair	Car	Plane
$NRDF_{w/oCD}$	0.048	0.023	0.047
$NRDF_{w/oEMD}$	0.223	0.275	0.351
$NRDF_{w/oNC}$	0.049	0.024	0.037
$NRDF_{w/oDS}$	0.047	0.022	0.037
$NRDF_{w/oDD}$	0.042	0.020	0.038
NRDF	**0.039**	**0.019**	**0.035**

D. Robotic Demonstration

We validate the practicality of the proposed approach by transferring P-ROI level process knowledge from a virtual chair to new instance of unseen chairs. The robot environment setup is shown in Fig. 6a. Initially the desired process knowledge is defined on 5 arbitrary P-ROIs on the virtual chair point cloud as shown in Fig. 6b. The process knowledge in this specific demonstration include the process trajectory, process direction, and orientation with respect to the 5 P-ROIs. Then NRDF assisted region correspondence is applied to estimate corresponding P-ROIs on the new chair instance point clouds to repeat the same process, in our experiments the trajectory is adapted to estimated corresponding P-ROI using a CAM software. Example execution images are shown in Fig. 6c, and the video execution of the same is provided as a supplementary material (https://youtu.be/YiEGInDQT-o).

VI. Conclusion

We present a contribution to solve the robotic problem of surface processing. Ideally, the robot is shown how to process a particular surface region and learns with this how to treat similar object surfaces. In this work we show that with the proposed implicit NRDF representation along with the inverse descriptor function, we are able to recover similar ROI on the target category-level object given only the reference object and an arbitrary ROI on it. The shape representation results and ROI retrieval tests on the BSRC dataset demonstrates the effectiveness of the proposed approach. Furthermore, our work opens an interesting research direction for point-based manipulation currently performed with NDF, as we clearly see that the shape representation capability is improving during the proposed training procedure

for NRDF. With respect to region recovery current system limitations include the challenges with respect to geometrical parts that are thin, for example the wings of a plane where the discriminative ability between the top surface and bottom surface is challenging. Future work will include research in this direction to add such a discriminative ability.

REFERENCES

[1] Y. Wen, J. Hu, and P. R. Pagilla, "A novel robotic system for finishing of freeform surfaces," in *Int. Conf. Robot. Automat. (ICRA)*. IEEE, 2019, pp. 5571–5577.

[2] Z.-Y. Liao, J.-Z. Wu, H.-M. Wu, H.-L. Xie, Q.-H. Wang, and X.-F. Zhou, "Profile error estimation and hierarchical compensation method for robotic surface machining," *Robot. Automat. Lett. (RA-L)*, 2024.

[3] W. Ng, H. Chan, W. K. Teo, and I.-M. Chen, "Programming robotic tool-path and tool-orientations for conformance grinding based on human demonstration," in *IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*. IEEE, 2016, pp. 1246–1253.

[4] C. W. Ng, K. H. Chan, W. K. Teo, and I.-M. Chen, "A method for capturing the tacit knowledge in the surface finishing skill by demonstration for programming a robot," in *Int. Conf. Robot. Automat. (ICRA)*. IEEE, 2014, pp. 1374–1379.

[5] M. Jin, S. Ji, L. Zhang, Q. Yuan, X. Zhang, and Y. Zhang, "Material removal model and contact control of robotic gasbag polishing technique," in *Conf. Robot. Automat. Mechatronics*. IEEE, 2008, pp. 879–883.

[6] B. Hazel, J. Cote, P. Mongenot, M. Sabourin, and F. Paquet, "Robotic polishing of turbine runners," in *Int. Conf. Appl. Robot. Power Industry (CARPI)*. IEEE, 2012, pp. 50–51.

[7] S. Hähnel, F. Pini, F. Leali, O. Dambon, T. Bergs, and T. Bletek, "Reconfigurable robotic solution for effective finishing of complex surfaces," in *Int. Conf. Emerg. Technologies and Factory Automat. (ETFA)*, vol. 1. IEEE, 2018, pp. 1285–1290.

[8] S. Schneyer, A. Sachtler, T. Eiband, and K. Nottensteiner, "Segmentation and coverage planning of freeform geometries for robotic surface finishing," *Robot. Automat. Lett. (RA-L)*, 2023.

[9] Y. Takeuchi, D. Ge, and N. Asakawa, "Automated polishing process with a human-like dexterous robot," in *Int. Conf. Robot. Automat. (ICRA)*. IEEE, 1993, pp. 950–956.

[10] K. Hayashi, H. Ueno, and H. Murakami, "Automation of precision finishing," in *World Automation Congress (WAC)*. IEEE, 2014, pp. 252–257.

[11] M. Mayr, F. Rovida, and V. Krueger, "Skiros2: A skill-based robot control platform for ros," in *IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*. IEEE, 2023, pp. 6273–6280.

[12] Z. L. Gan, S. N. Musa, and H. J. Yap, "A review of the high-mix, low-volume manufacturing industry," *Applied Sciences*, vol. 13, no. 3, p. 1687, 2023.

[13] L. Mescheder, M. Oechsle, M. Niemeyer, S. Nowozin, and A. Geiger, "Occupancy networks: Learning 3d reconstruction in function space," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 4460–4470.

[14] J. Chibane and G. Pons-Moll, "Neural unsigned distance fields for implicit function learning," *Adv. Neural Inf. Process. Syst.*, vol. 33, pp. 21 638–21 652, 2020.

[15] J. Fu, Y. Du, K. Singh, J. B. Tenenbaum, and J. J. Leonard, "Robust change detection based on neural descriptor fields," in *IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*. IEEE, 2022, pp. 2817–2824.

[16] E. Chun, Y. Du, A. Simeonov, T. Lozano-Perez, and L. Kaelbling, "Local neural descriptor fields: Locally conditioned object representations for manipulation," in *Int. Conf. Robot. Automat. (ICRA)*, 2023, pp. 1830–1836.

[17] L. Yen-Chen, P. Florence, J. T. Barron, A. Rodriguez, P. Isola, and T.-Y. Lin, "inerf: Inverting neural radiance fields for pose estimation," in *IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*. IEEE, 2021, pp. 1323–1330.

[18] A. Simeonov, Y. Du, A. Tagliasacchi, J. B. Tenenbaum, A. Rodriguez, P. Agrawal, and V. Sitzmann, "Neural descriptor fields: Se (3)-equivariant object representations for manipulation," in *Int. Conf. Robot. Automat. (ICRA)*. IEEE, 2022, pp. 6394–6400.

[19] M. C. Lee, S. J. Go, J. Y. Jung, and M. H. Lee, "Development of a user-friendly polishing robot system," in *IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, vol. 3. IEEE, 1999, pp. 1914–1919.

[20] J. Ming *et al.*, "Development of automatic mold polishing system," *Int. Con. Robot. Automat. (ICRA)*, pp. 14–19, 2003.

[21] Y. Kim, C. Sloth, and A. Kramberger, "A framework for transferring surface finishing skills to new surface geometries," in *IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*. IEEE, 2022, pp. 7650–7655.

[22] Y. Kim, C. Sloth, and A. Kramberger, "Skill transfer for surface finishing tasks based on estimation of key parameters," in *Conf. Automat. Sci. Eng. (CASE)*. IEEE, 2022, pp. 2148–2153.

[23] Y. Wang, C. Chen, F. Peng, Z. Zheng, Z. Gao, R. Yan, and X. Tang, "Al-promp: Force-relevant skills learning and generalization method for robotic polishing," *Robot. Comput.-Integr. Manuf.*, vol. 82, p. 102538, 2023.

[24] Y. Huo, P. Li, D. Chen, Y.-H. Liu, and X. Li, "Model-free adaptive impedance control for autonomous robotic sanding," *IEEE Trans. Automat. Sci. Eng. (T-ASE)*, vol. 19, no. 4, pp. 3601–3611, 2022.

[25] H. Ochoa and R. Cortesão, "Impedance control architecture for robotic-assisted mold polishing based on human demonstration," *IEEE Trans. Ind. Electronics*, vol. 69, no. 4, pp. 3822–3830, 2022.

[26] D. Rodriguez, C. Cogswell, S. Koo, and S. Behnke, "Transferring grasping skills to novel instances by latent space non-rigid registration," in *Int. Conf. Robot. Automat. (ICRA)*. IEEE, 2018, pp. 4229–4236.

[27] P. R. Florence, L. Manuelli, and R. Tedrake, "Dense object nets: Learning dense visual object descriptors by and for robotic manipulation," in *Conf. Robot Learn.* PMLR, 2018, pp. 373–385.

[28] S. Thompson, L. P. Kaelbling, and T. Lozano-Perez, "Shape-based transfer of generic skills," in *Int. Conf. Robot. Automat. (ICRA)*. IEEE, 2021, pp. 5996–6002.

[29] F. Liu and X. Liu, "Learning implicit functions for topology-varying dense 3d shape correspondence," *Adv. Neural Inf. Process. Syst.*, vol. 33, pp. 4823–4834, 2020.

[30] V. G. Kim, Y. Lipman, and T. Funkhouser, "Blended intrinsic maps," *ACM Trans. Graph. (TOG)*, vol. 30, no. 4, pp. 1–12, 2011.

[31] V. G. Kim, W. Li, N. J. Mitra, S. Chaudhuri, S. DiVerdi, and T. Funkhouser, "Learning part-based templates from large collections of 3d shapes," *ACM Trans. Graph. (TOG)*, vol. 32, no. 4, pp. 1–12, 2013.

[32] N. Chen, L. Liu, Z. Cui, R. Chen, D. Ceylan, C. Tu, and W. Wang, "Unsupervised learning of intrinsic structural representation points," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 9121–9130.

[33] Z. Chen, K. Yin, M. Fisher, S. Chaudhuri, and H. Zhang, "Bae-net: Branched autoencoder for shape co-segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 8490–8499.

[34] X. Zhu, D. Du, H. Huang, C. Ma, and X. Han, "3d keypoint estimation using implicit representation learning," *arXiv preprint arXiv:2306.11529*, 2023.

[35] F. Liu and X. Liu, "Learning implicit functions for dense 3d shape correspondence of generic objects," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2023.

[36] M. Denitto, S. Melzi, M. Bicego, U. Castellani, A. Farinelli, M. A. Figueiredo, *et al.*, "Region-based correspondence between 3d shapes via spatially smooth biclustering," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 4260–4269.

[37] V. G. Kim, W. Li, N. J. Mitra, S. DiVerdi, and T. Funkhouser, "Exploring collections of 3d models using fuzzy correspondences," *ACM Trans. Graph. (TOG)*, vol. 31, no. 4, pp. 1–11, 2012.

[38] V. Ganapathi-Subramanian, B. Thibert, M. Ovsjanikov, and L. Guibas, "Stable region correspondences between non-isometric shapes," in *Computer Graphics Forum*, vol. 35, no. 5. Wiley Online Library, 2016, pp. 121–133.

[39] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[40] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "Pointnet: Deep learning on point sets for 3d classification and segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 652–660.

[41] H. Fan, H. Su, and L. J. Guibas, "A point set generation network for 3d object reconstruction from a single image," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 605–613.

[42] M. Liu, L. Sheng, S. Yang, J. Shao, and S.-M. Hu, "Morphing and sampling network for dense point cloud completion," in *Proc. AAAI Conf. Artif. Intell.*, vol. 34, no. 07, 2020, pp. 11 596–11 603.

[43] A. X. Chang, T. Funkhouser, L. Guibas, *et al.*, "ShapeNet: An Information-Rich 3D Model Repository," *arXiv:1512.03012*, 2015.