LGAfford-Net: A Local Geometry Aware Affordance Detection Network for 3D Point Clouds

Ramesh Ashok Tabib Dikshit Hegde Uma Mudenagudi
Center of Excellence in Visual Intelligence, School of Electronics and Communication Engineering,
KLE Technological University, Vidyanagar, Hubballi, Karnataka, INDIA-580031
ramesh_t@kletech.ac.in, dikshithegde@gmail.com, uma@kletech.ac.in

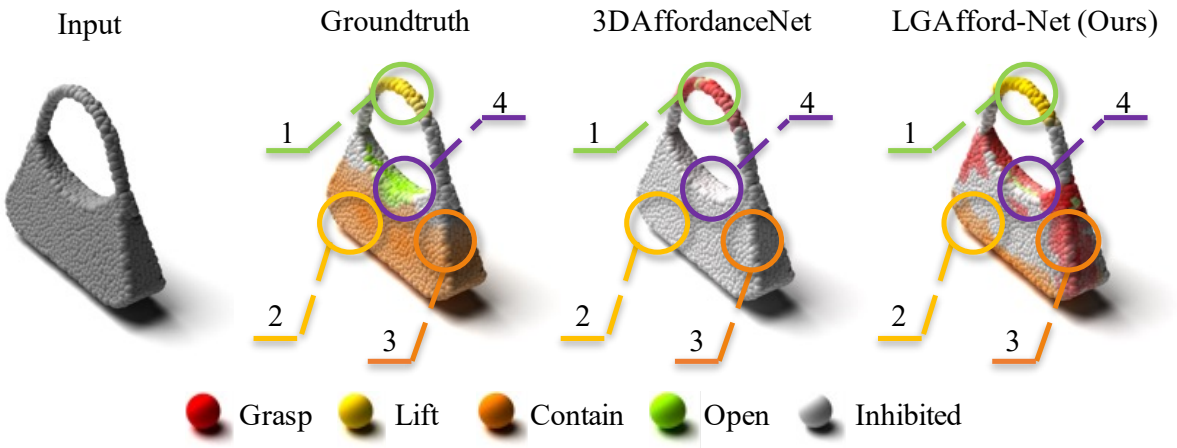

Figure 1. **LGAfford-Net** combines local geometry and semantic cues for affordance detection. Regions 1, 2, and 4 highlight instances where LGAfford-Net outperforms 3DAffordanceNet [6] in affordance detection. However, Region 3 stands out as exceptional region where LGAfford-Net predicts a *Grasp* affordance, the label unavailable in the groundtruth. This shows the generalizability of the LGAfford-network, showcasing its ability to identify affordances beyond explicitly annotated labels.

Abstract

In this paper, we introduce LGAfford-Net, a novel architecture tailored for affordance detection in 3D point clouds. Affordance, crucial for human-robot interaction, denotes regions on objects where interaction is possible. Understanding affordance demands perceiving 3D space akin to humans. Leveraging the ubiquity of point clouds in capturing 3D environments, our method addresses challenges posed by their sparse, unordered, and unstructured nature. Unlike prior approaches that overlook local context and semantic cues, we propose a Semantic Geometric Correlator (SGC) block, integrating Local Geometric Descriptor (LGD) for local understanding, and Edge Convolution for semantic awareness. The integration of SGC, LGD, and edge convolution within our network enhances its capability to perceive and understand affordances by leveraging both geometric and semantic information effectively. Additionally, we employ Class Specific Classifiers (CSC) to accommodate multiple affordance types per point. CSC effectively establish one to many relationship between point to affordance labels. We demonstrate the results of proposed architecture on 3DAffordanceNet a benchmark dataset and compare them with state-of-the-art methods. We demonstrate the effectiveness of the features learnt by our proposed architecture for the point cloud classification task using the ModelNet40 dataset.

1. Introduction

In this paper, we propose a novel architecture for detection of Affordance in 3D point cloud by leveraging local geometric information and name it as LGAfford-Net. Affordance, as introduced by Gibson [8], refers to the understanding of how humans interact with the environment. This understanding is pivotal for the development of intelligent

systems capable of navigating [11, 28, 40], assisting [14], and interacting with individual objects [17, 18] across diverse scenes [5, 10]. In recent years, the advent of visual sensors, particularly RGB [31, 34, 45] and depth cameras (RGBD) [23, 25, 33], aids the collection of data for affordance estimation. However, the existing 2D/2.5D datasets fail to capture the true geometry of objects and scenes, affecting the accuracy of affordance prediction.

In response to these challenges, researchers have turned their attention to 3D point clouds, as they provide a more comprehensive representations of the geometry inherent in objects and scenes. Despite its potential representations, processing point cloud data for affordance detection remains challenging due to its unstructured and unordered nature. Recent techniques for processing point cloud data include sharedMLP [4, 26, 42], hierarchical feature extraction [9, 27], and geometric feature analysis [1, 3, 15, 16, 29, 38] have made significant progress in the tasks such as classification [26, 42], segmentation [4, 22], upsampling [20, 21, 24] and refinement [12, 32, 35, 37, 39]. These techniques fail to determine the probabilistic outcome associated with the object due to unavailability of distinct probabilities for each point/regions. To address this challenge, recent research introduces novel dataset, called as 3D-AffordanceNet [6]. This dataset provides probability values for each point in the point cloud, allowing for the estimation of multiple affordances with varying confidence levels.

Despite these advancements, accurately estimating affordance in 3D point clouds remains an open problem. Traditional approaches, including segmentation-based methods [30], classify points into specific affordances with usage of conventional activation functions. However, these points may have multiple probabilistic affordance labels in overlapping region and demands establishment of one to many relational mapping between the point and considered affordance labels. In this context, we propose a supervised approach for affordance detection by considering semantic local geometry, and one to many relational mapping.

Considering local geometric features for affordance detection includes capturing the intricacies of the local neighbourhood which stems from inherent statistical properties. We capture statistical features through proposed Local Geometric Descriptor (LGD) to provide valuable insights into the distribution and properties of data points, and also aids in capturing the geometric characteristics necessary for detecting affordances accurately. The spatial arrangement of points and geometric properties of objects together facilitate affordance perception, making local geometric information necessary for robust detection. Towards this, we propose Semantic Geometric Correlator (SGC) in conjunction with LGD to extract semantic features by considering local geometry. Introducing SGC into our approach, facilitates capture of fine-grained geometric details essential for affordance detection.

Additionally, we introduce a Class Specific Classifier (CSC) to establish a direct relationship between semantic local features and their corresponding affordance labels. This classifier acts as an intermediary mechanism, bridging the gap between semantic local features and corresponding affordance labels. The CSC learns to discern subtle patterns and correlations within the semantic features, allowing for local geometry aware detection of affordances.

As discussed, local geometric features offer a more comprehensive representation of the underlying structure associated with 3D point clouds, allowing our model to learn subtle variations and patterns indicative of different affordance labels. Moreover, by retaining the annotated probabilities of affordance during training, our approach leverages both geometric and probabilistic information to enhance affordance detection. This fusion of semantic local geometry with probabilistic affordance annotations enhances the interpretability and robustness of our model, enabling it to make more informed decisions towards affordance detection. More specifically, the contribution of our work include,

- **LGAfford-Net**, a novel architecture specifically designed for the detection of affordances within point clouds, leveraging local geometric information which includes,
  - **Semantic Geometric Correlator (SGC)** block to facilitate the correlation of local geometric features with semantically similar regions within point clouds using (Section 2.1),
    - **Local Geometric Descriptor (LGD)** tailored for comprehending local surfacial information (Section 2.1.1), and
    - Edge Convolution [42] for considering both semantic and local information for more robust affordance detection.
  - A **Class Specific Classifier (CSC)** for affordance detection, considering the multiple affordance types associated with each point within a point cloud (Section 2.2).
- Demonstration of the proposed architecture through extensive experimentation on the 3D-AffordanceNet dataset [6], along with comparative analysis against state-of-the-art methods.
- Demonstration through classification of 3D objects using ModelNet40 dataset [43] to determine the robustness of the extracted features through LGAfford-Net.

In Section 2, we discuss the proposed LGAfford-Net, Local geometry aware affordance detection network. We discuss the results and comparison of the proposed methodology with state-of-the-art methods in Section 3, and conclude on the findings and shortcomings in Section 4.

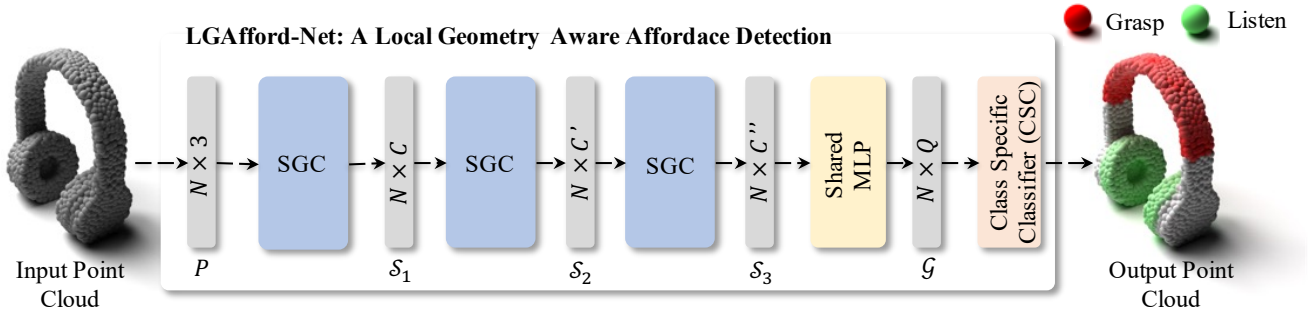

Figure 2. The proposed architecture of LGAfford-Net: A Local Geometry aware Affordance Detection network. Here, SGC represents Semantic Geometric Correlator, $\boldsymbol{P}$ represents input point cloud, $\boldsymbol{\mathcal{S}_i}$ represents Semantic local geometric features, $\boldsymbol{\mathcal{G}}$ represents the combined hierarchical geometric features considered for affordance detection.

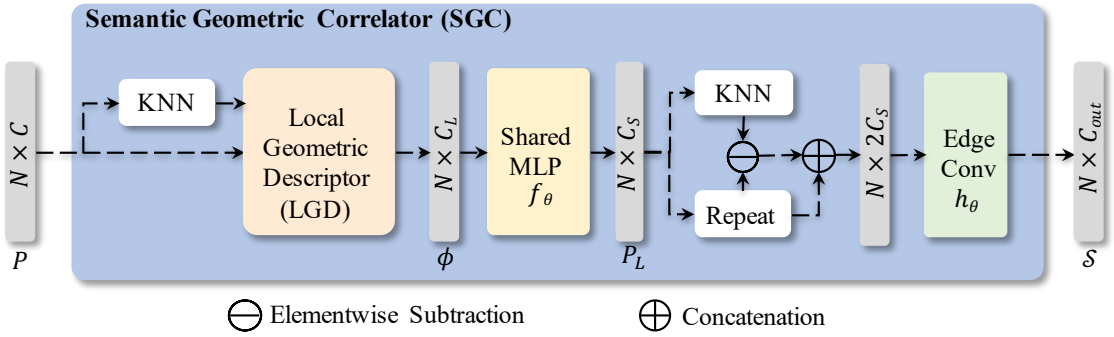

Figure 3. SGC: Semantic Geometric Correlator to capture the semantic local geometric features of the local neighbourhood. Here, ϕ represents Local Geometric Features, $\boldsymbol{P_L}$ represent learnt Local Geometry Features, $\boldsymbol{\mathcal{S}}$ represents the Semantic Local Geometric Features.

2. LGAfford-Net: Local Geometry Aware Affordance Detection Network

In this section, we introduce LGAfford-Net, a Local Geometry Aware network tailored for affordance detection in point clouds, as illustrated in Figure 2. Affordance detection in point clouds necessitates an understanding of both local geometry and semantic context. To address this challenge, we propose Semantic Geometric Correlator (SGC) block to extracts features by combining semantic and geometric information. We also include a Class Specific Classifier (CSC) to assign affordance labels to individual points, leveraging the features $\boldsymbol{\mathcal{G}}$ extracted by series of SGC blocks. This classifier enhances the level of details and accuracy of affordance detection by understanding the one to many relational mapping between the point and considered affordance labels. Thus, LGAfford-Net combines semantic and local geometric information, enabling accurate and context-aware affordance detection in point clouds.

Formally, we define the point cloud $\boldsymbol{P}$ as a set of points $\{\boldsymbol{p_1}, \boldsymbol{p_2}, ..., \boldsymbol{p_N}\}$, where each point $\boldsymbol{p_i} \in \mathbb{R}^3$ represents the Cartesian coordinates $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$ of a point in 3D space, and $\boldsymbol{N}$ denotes the total number of points in the point cloud.

In what follows, we discuss Semantic Geometric Correlator and Class Specific Classifier.

2.1. Semantic Geometric Correlator (SGC)

Semantic Geometric Correlator (SGC) block, exploits the intrinsic relationship between semantic and local geometric features, by correlating local geometric properties with semantic information as depicted in Algorithm 1.

Initially, we estimate the local neighbourhood of a point $\boldsymbol{p_i}$ using K-Nearest Neighbours [2, 19] to obtain $\boldsymbol{index_{Li}}$. Local Geometric features $\boldsymbol{\phi_i}$ are estimated using the local neighbours via Local Geometric Descriptor (LGD). Towards extracting learnt features ($\boldsymbol{P_{Li}}$) of the local neighbourhood, we convolve weights of shared mlp on $\boldsymbol{\phi_i}$. To estimate semantic local geometric features, we group the learnt features ($\boldsymbol{P_{Li}}$) via K-Nearest Neighbours and perform Edge Convolution (EdgeConv) [42] as shown in Equation 1.

$$\boldsymbol{\mathcal{S}_i} = \max_{\boldsymbol{j \in K_S}} [\boldsymbol{h_\theta}(\boldsymbol{x_j} - \boldsymbol{x_i}, \boldsymbol{x_i})] \tag{1}$$

where, $\boldsymbol{x_j}$ represent the local neighbours and $\boldsymbol{x_i}$ represents the query point.

EdgeConv [42] helps to capture features of regions with

Algorithm 1: Semantic Geometric Correlator

Input: Point Cloud $\rightarrow P$; // B, N, C

Output: Semantic Geometric Correlator's weights (f_θ, h_θ), Semantic Local Geometric Features $\mathcal{S}$;

Initialize $K_L, K_S, N, C_L, C_S, C_{out}$

$index_L$ = KNN (P, K_L)

/* B, N, K_L */

ϕ = LGD $(P, index_L)$

/* B, N, C_L */

P_L = SharedMLP $(C_L, C_S)(\phi)$

/* B, N, C_S */

$index_S$ = KNN (P_L, K_S)

/* B, N, K_S */

P_{neigh} = gatherOp $(index_S, P_L)$

/* B, N, K_S, C_S */

$\mathcal{S}$ = max (EdgeConv (P_{neigh}, P_L))

/* B, N, C_{out} */

similar geometry but different spatial locations, facilitating easier affordance prediction. This integration of semantic and local geometric information enhances the model's ability to infer meaningful patterns necessary for accurate affordance detection.

2.1.1 Local Geometric Descriptor (LGD)

Local Geometric Descriptors enables understanding of the spatial arrangement of points within the point cloud. To represent the concept of triangular faces in mesh processing, we construct triangles using the two nearest neighbors of each point p_i in the point cloud P, denoted as (p_{j1}, p_{j2}).

$$\phi_i = \begin{cases} p_i = x, y, z; & p_i \in \mathbb{R}^n, \\ \overrightarrow{e_1} = p_{j1} - p_i; & \overrightarrow{e_1} \in \mathbb{R}^n, \\ \overrightarrow{e_2} = p_{j2} - p_i; & \overrightarrow{e_2} \in \mathbb{R}^n, \\ |\overrightarrow{e_1}| = \ell_2(\overrightarrow{e_1}); & |\overrightarrow{e_1}| \in \mathbb{R}^1, \\ |\overrightarrow{e_2}| = \ell_2(\overrightarrow{e_2}); & |\overrightarrow{e_2}| \in \mathbb{R}^1, \\ \hat{n} = \overrightarrow{e_1} \times \overrightarrow{e_2}; & \hat{n} \in \mathbb{R}^3, \\ \mu_i = mean(p_j); & p_j \in \mathbb{R}^n, \\ \sigma_i = std(p_j); & p_j \in \mathbb{R}^n. \end{cases} \quad (2)$$

These triangles serve as local patches, allowing us to capture geometric characteristics effectively. We then estimate the Local Geometric Descriptor (LGD) using the generated triangles, as defined in Equation 2.

Here, p_j is the nearest neighbours gathered using $index_{Li}$ to capture local geometric information. We compute $\overrightarrow{e_1}$ and $\overrightarrow{e_2}$ representing edge vector for (p_{j1}, p_{j2}) relative to p_i respectively. ℓ_2 represents the L2 Norm of the vector, informing about the displacement of the first two neighbours with respect to the point p_i. Unlike [3, 29], we also consider standard deviation σ_i [1] and mean μ_i [37] of the local neighbours p_j. The mean (μ_i) denotes the placement of the query point p_i in the point cloud (boundary or planar point) and standard deviation (σ_i) denotes the density of the local neighbours along with the directional vector pointing towards mean. LGD encapsulates essential local geometric information, enabling us to extract meaningful features for affordance detection.

Affordance class labels, represent distinct actions or interactions within a given environment. Therefore, it is reasonable to assume that these classes are independent of each other. Multiple affordance classes may coexist, each with its own probability of occurrence. However, traditional classifiers, such as those employing softmax activation, tend to merge the probabilities of different affordance classes, potentially leading to ambiguity and inaccuracies, especially in case of overlapping affordances. To address this issue, we encorporate the methodology similar to [36] and develop a Class Specific Classifier (CSC) for the preservation of the independent nature of each affordance class, thereby retaining the one to many relationship between the point and each affordance class. In what follows, we discuss the Class Specific Classifier.

2.2. Class Specific Classifier (CSC)

To develop a class specific classifiers for affordance detection, we assign a dedicated classifier to each affordance class as shown in Figure 4.

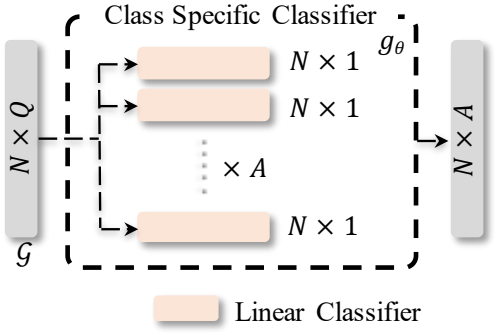

Figure 4. Illustration of the Class Specific Classifier architecture for affordance detection. Each affordance class is associated with its own dedicated classifier as depicted in the figure above, allowing for independent predictions of the probability associated with each affordance based on input features. Here A represents the number of affordance classes.

Each classifier is trained independently to predict the probability of its corresponding affordance given the input features $\mathcal{G}$. By decoupling the classifiers for different affordance classes, we ensure the probability predictions are not influenced by the presence of other affordances, thus

preserving the independence of each class. The motivation behind developing class specific classifiers for affordance detection stems from the need of maintaining the decoupled relationship between the point and the corresponding affordance classes.

By treating each affordance class as distinct and independent, we can better capture the relationships between affordances and their corresponding environmental cues. This approach allows for robust handling of overlapping affordances, as each class specific classifier focuses solely on predicting the probability of its designated affordance class without being influenced by other classes.

2.3. Loss Function

Towards fine-tuning of semantic local geometric features and the affordance prediction, we optimise the learning parameter (f_θ, h_θ, g_θ) as shown in Figure 3 and Figure 4 respectively. We use a combination [6] of Binary Cross Entropy loss ($\mathcal{L}_{BCE}$) and Dice loss ($\mathcal{L}_{DICE}$) as shown in Equation 3.

$$\mathcal{L}_{total} = \mathcal{L}_{BCE} + \mathcal{L}_{DICE} \tag{3}$$

We include Binary Cross Entropy loss as shown in Equation 4, to handle the prediction of the affordance along with fine-tuning of feature extraction module.

$$\begin{aligned}\mathcal{L}_{BCE} = -\frac{1}{N}\sum_{j=1}^{A}\sum_{i=1}^{N} y_{ij}\log \hat{y_{ij}} \\ +(1-y_{ij})\log(1-\hat{y_{ij}})\end{aligned} \tag{4}$$

where, y_{ij} is the groundtruth of j^{th} affordance probability for i^{th} point, similarly $\hat{y}_{ij}$ is the corresponding prediction value. Here, A represents the total number of considered affordance classes.

A point cloud consists of affordable and inhibited regions. Affordable regions denote areas where human interaction is feasible, while inhibited regions refer to areas where interaction is hindered. Affordable regions typically constitute a smaller portion of the point cloud compared to inhibited regions. To handle this imbalance between the inhibited region and the affordable regions, we employ Dice Loss as shown in Equation 5.

$$\begin{aligned}\mathcal{L}_{DICE} = \sum_{j=1}^{A} 1 - \frac{\sum_{i=1}^{N} y_{ij}\hat{y}_{ij} + \epsilon}{\sum_{i=1}^{N} y_{ij} + \hat{y}_{ij} + \epsilon} \\ - \frac{\sum_{i=1}^{N}(1-y_{ij})(1-\hat{y}_{ij}) + \epsilon}{\sum_{i=1}^{N} 2 - y_{ij} - \hat{y}_{ij} + \epsilon}\end{aligned} \tag{5}$$

Here, ϵ prevents the loss from diverging towards infinity. The incorporation of Dice Loss helps to mitigate the class imbalance, ensuring effective training and improving the model's ability to accurately detect affordances.

3. Results and Discussions

In this section, we discuss the results (Section 3.3) of proposed methodology and compare them with state-of-the-art methods. We conduct experiments on benchmark datasets (Section 3.1) and show improved performance in comparison with state-of-the-art methods.

3.1. Datasets Description

To train and evaluate the performace of the proposed LGAfford-Net, we use 3D AffordanceNet dataset [6] and compare the results with the state-of-the-art benchmarking proposed in 3D AffordanceNet dataset [6]. To demonstrate the generalizability of the proposed methodology, we encorporate ModelNet40 dataset [43] and compare the object classification with state-of-the-art methods.

- **3DAffordanceNet [6]**: dataset consists of 23K shapes with 23 semantic object categories and consist of 18 affordance labels. Each point cloud in the dataset is meticulously annotated with affordance labels, indicating the likelihood of specific human interactions with objects. The annotations provide probabilistic information, offering insights into the probability of various affordance categories associated with individual points within the point cloud.
- **ModelNet40 [43]**: ModelNet40 dataset consists of CAD models with 40 categories. Each object category in the dataset comprises a varying number of CAD models, with a total of over 12,000 models across all categories. These CAD models are sampled to 1024 points to form a point-cloud.

3.2. Experimental Setup

In this section, we discuss the architectural design and optimization methodology considering loss functions as discussed in Section 2.3.

- **Architectural Details**: The architectural specifications of LGAfford-Net with $A = 18$ are presented in detail as shown in Table 1. This network comprises of 18 Class Specific Classifiers (CSC), each configured according to the parameters outlined in Table 1. Specifically, the Semantic Geometric Correlator (SGC) incorporates Instance Normalization and Leaky ReLU activation with a parameter $p = 0.2$. Class Specific Classifiers utilize Batch Normalization for improved stability during training. Additionally, sigmoid activation functions are employed at the final layer of each Class Specific Classifier to compute the affordance probabilities.
- **Training Setup towards Affordance detection**: The training parameters f_θ, h_θ, and g_θ are initialized using a uniform distribution. LGAfford-Net is trained for 200 epochs using the Adam optimizer with Cosine Annealing. We consider 2048 points as input for each training

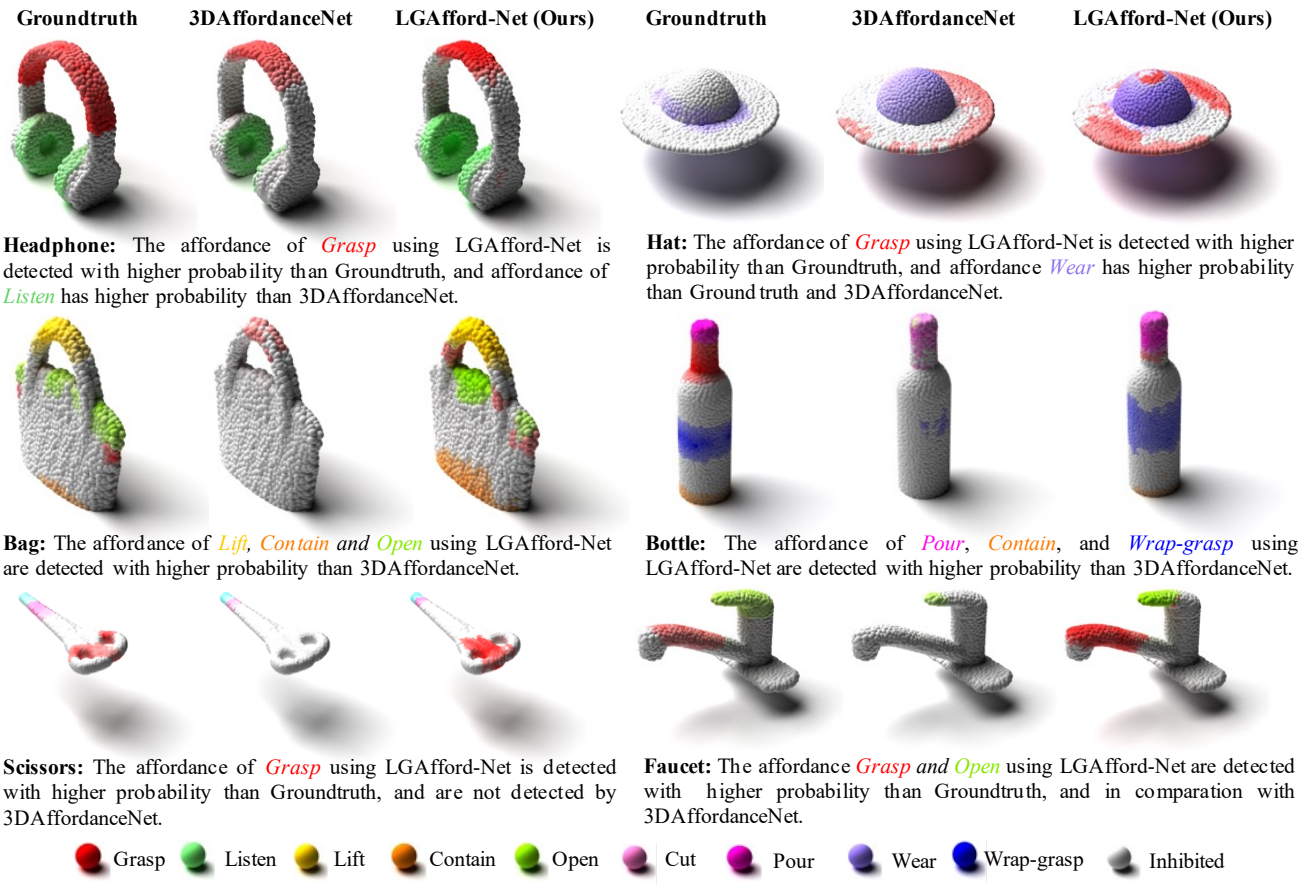

Figure 5. Qualitative analysis of proposed **LGAfford-Net** for Affordance detection in comparison with 3DAffordanceNet [6].

Table 1. Architectural Details and specification of LGAfford-Net which includes SGC and CSC as shown in Figure 2.

Type	Parameters	#	Keys
SGC	$(K_L, K_S, C_L, C_S, C_{out})$	(40,40,20,16,16)	SGC 1
SGC	$(K_L, K_S, C_L, C_S, C_{out})$	(40,40,82,64,64)	SGC 2
SGC	$(K_L, K_S, C_L, C_S, C_{out})$	(40,40,322,512,512)	SGC 3
Conv1d	(C_{in}, C_{out})	(512,1024)	SharedMLP
CSC Linear 1	(C_{in}, C_{out})	(1024,128)	SharedMLP
CSC Linear 2	(C_{in}, C_{out})	(128, 1)	SharedMLP

iteration and set the learning rate to $\mathbf{10^{-3}}$. During training, LGAfford-Net optimizes its parameters using the loss function as discussed in Section 2.3. These experiments are conducted on Tesla V100 with 32 GB VRAM, and Pytorch Framework. The above configuration is aligned with the settings as described in [6].

- **Training Setup for Classification**: Freezed weights of series of Semantic Geometric Correlators (SGC), previously trained for affordance detection is extended for another downstream task, classification using ModelNet40 dataset [43]. We train a classifier with three layers using a learning rate of $\mathbf{10^{-3}}$, Adam optimizer, and Cross Entropy loss function for 200 epochs.

3.3. Results

In this section, we present the results of LGAfford-Net for affordance detection on the 3DAffordanceNet dataset [6]. We discuss the qualitative and quantitative results of LGAfford-Net in comparison with 3DAffordanceNet [6]. Subsequently, we provide a comprehensive analysis of results, focusing on the effectiveness of our approach.

The evaluation and comparison of our proposed methodology for Affordance Detection are conducted on the 3DAffordanceNet Dataset as shown in Table 2. We employ multiple performance metrics to assess the effectiveness of our approach, including mean Intersection over Union (mIoU), mean Average Precision (mAP), mean Area Under the Curve (mAUC), and Mean Squared Error (MSE). For mIoU, mAP, and mAUC, higher values indicate better performance, reflecting a greater accuracy and precision in affordance detection. Conversely, for MSE, lower values indicate better performance, showing less deviation between predicted and ground truth affordances.

In Figure 5, we show comparative result analysis of both LGAfford-Net and 3DAffordanceNet. Analyzing the detec-

Table 2. Evaluation and comparison of proposed methodology for Affordance Detection using mIoU, mAP, mAUC, and MSE on the 3DAffordanceNet Dataset as in [6] considering 18 Affordance. Here '↑' represent higher is better, and '↓' represent lower is better. Blue represents increase in mIoU and Red represents decrease in mIoU of LGAfford-Net over 3DAffordanceNet [6].

	3DAffordanceNet [6]				LGAfford-Net (Ours)				Change
Affordances	**mIoU ↑**	**mAP ↑**	**mAUC ↑**	**MSE ↓**	**mIoU ↑**	**mAP ↑**	**mAUC ↑**	**MSE ↓**	**in mIoU**
Grasp	13.9	43.9	82.5	0.0030	18.3	38.8	79.9	0.0030	4.3
Contain	21.6	57.6	89.9	0.0070	22.2	51.4	86.6	0.0060	0.6
Lift	40.2	85.2	98.7	0.0001	38.9	77.8	94.7	0.0001	1.3
Open	25.4	51.8	91.6	0.0030	25.7	49.5	90.2	0.0028	0.2
Lay	1.0	12.3	50.1	0.0006	23.6	50.9	89.5	0.0005	22.6
Sit	34.9	80.9	96.1	0.0060	41.9	83.6	96.6	0.0053	7.0
Support	18.8	54.0	90.2	0.0130	19.9	54.9	90.8	0.0121	1.1
Wrap-grasp	5.6	20.7	74.6	0.0070	4.2	16.2	68.3	0.0028	1.4
Pour	17.7	47.7	89.2	0.0050	17.7	40.2	86.1	0.0025	0.0
Move	9.9	35.5	78.9	0.0250	10.3	35.3	80.4	0.0208	0.4
Display	32.1	65.5	92.1	0.0020	37.9	64.4	92.3	0.0020	5.8
Push	5.5	20.5	85.0	0.0020	8.4	24.0	84.0	0.0005	2.9
Pull	11.8	40.5	89.7	0.0006	36.5	62.9	90.0	0.0002	24.7
Listen	11.9	36.0	86.1	0.0020	15.8	34.4	80.7	0.0007	3.9
Wear	5.9	18.3	61.0	0.0020	6.9	19.1	67.0	0.0009	1.0
Press	14.8	34.2	91.8	0.0007	20.4	43.9	94.5	0.0007	6.6
Cut	14.5	40.2	91.7	0.0002	9.8	29.4	91.0	0.0003	4.7
Stab	35.4	91.4	98.7	0.0001	32.5	84.3	99.3	0.0001	2.9
Average	17.8	46.4	85.5	0.0800	21.7	47.8	86.8	0.0630	3.9

tion results in conjunction with the groundtruth from Figure 5, Figure 1, Figure 6, and Table 2, several observations were made:

- LGAfford-Net demonstrated superior capture of semantic local features on objects such as Bottle, Scissors, Hat, and Bag as shown in Figure 5, compared to 3DAffordanceNet.
- Examination of Table 2 revealed that affordance categories such as *Lay* and *Pull* exhibited superior performance in terms of mIoU compared to 3DAffordanceNet.
- Figure 1 shows instance where LGAfford-Net successfully detected *Grasp* affordances that were absent in the ground truth, suggesting its ability to generalize affordance detection.
- Conversely, 3DAffordanceNet exhibited erroneous results compared to the ground truth, as evident in Figure 1 and further visualized in Figure 5 for objects like Scissors, Faucets, and Bottle.
- Comparing the results for Headphones, both methods yielded similar outcomes with differences in confidences.
- On the Hat object, 3DAffordanceNet demonstrated a greater degree of generalization compared to LGAfford-Net in distinguishing between the flat regions.
- Affordances such as *Wrap*, *Cut*, *Stab*, and *Lift* exhibited superior performance in terms of mIoU for 3DAffordanceNet, indicating its proficiency in understanding physics-related affordances.
- We demonstrate the generalizability of LGAfford-Net in Figure 6 on the Door object, where the *Pull* affordance is detected despite being unavailable in the Ground Truth.
- In Figure 6, we demonstrate the superiority of LGAfford-Net on the Laptop and Table objects, showcasing improved detection of affordance regions such as *Press*, *Move*, and *Support*. Notably, LGAfford-Net successfully detects the *Press* affordance in the keyboard region, which is unavailable in the Ground Truth, unlike 3DAffordanceNet.

Overall, our comprehensive analysis suggests that LGAfford-Net exhibits greater robustness compared to 3DAffordanceNet in detecting affordances within 3D point clouds even when groundtruth labels are unavailable. LG Afford-Net outperforms ground truth data in affordance detection due to its ability to generalize patterns beyond explicit labels and leverage semantic understanding and local context, which may not be fully captured in the ground truth annotations. Additionally, the model's complexity and capacity to capture intricate affordance patterns contribute to its superior performance in certain instances.

Our proposed methodology demonstrate promising results across all metrics as shown in Table 2. Specifically, we achieved high values for mIoU, mAP, and mAUC, signifying accurate localization and classification of affordances within the point cloud data. Additionally, the MSE was minimized, indicating minimal error between predicted and ground truth affordances, further validating the robustness of our approach.

Discrepancies may arise between the labels present in

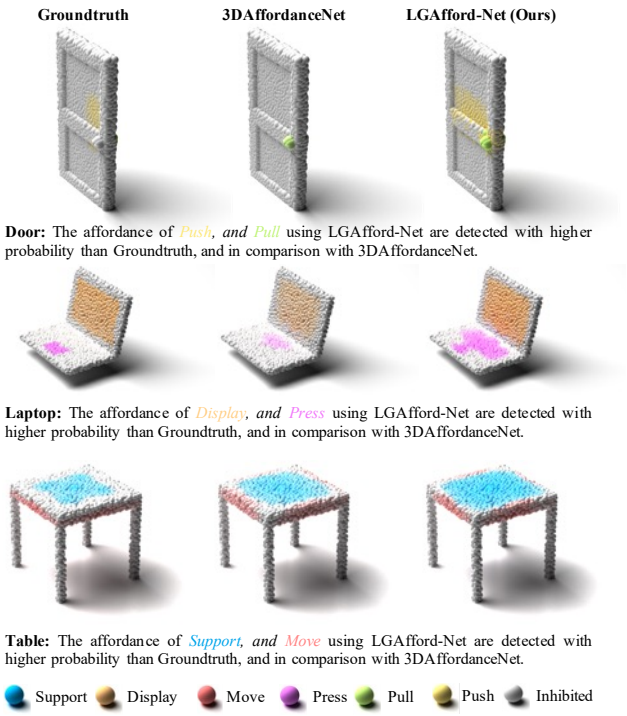

Figure 6. Qualitative analysis of proposed **LGAfford-Net** for Affordance detection in comparison 3DAffordanceNet [6]. We show better affordance detection of LGAfford-Net than groundtruth in certain region with high probability.

the ground truth dataset for affordance and the labels predicted by our approach as shown in Figure 6. Interestingly, we often observe that our model assigns relevant labels to certain instances where the ground truth dataset may lack specificity as shown in Figure 1, Figure 5, and Figure 6. This phenomenon could potentially explain why certain metrics may not reflect promising results, despite our approach providing meaningful and contextually relevant outputs.

Table 3. The classification accuracy of proposed methodology in comparison with state-of-the-art method on ModelNet40 with 1024 point density.

Methods	Accuracy
PointNet [26](2017)	89.2
PointNet++ [27](2017)	90.7
MRTNet [7](2018)	91.2
Spec-GCN [41](2018)	91.5
PCNN [4](2018)	92.3
DGCNN [42](2019)	92.2
RSCNN [22](2019)	92.9
Point Transformer [44](2020)	93.7
KCNet [13](2021)	91.0
PCT [9](2021)	93.2
LGAfford-Net + Classifier(Ours)	91.1

Such discrepancies highlight the complexity inherent in affordance detection tasks and underscore the challenges associated with accurately annotating real-world data [6]. While the ground truth dataset serves as a valuable reference point, it may not always capture the full spectrum of affordance scenarios present in diverse environments. Rather than solely relying on quantitative assessments, it is essential to complement the results with qualitative analyses to gain a comprehensive understanding of our approach's performance. By considering both quantitative metrics and qualitative observations, we can better evaluate the effectiveness and robustness of our methodology for affordance detection.

We assess the effectiveness of LGAfford-Net, previously trained for affordance detection, in an extended application for object classification. We compare the classification accuracy achieved by our method with state-of-the-art end-to-end trained point cloud classification methods, as outlined in Table 3. It is important to note that the end-to-end trained models referenced in the comparison are trained directly on the ModelNet40 dataset [43], whereas LGAfford-Net does not have information about the dataset. This comparison allows us to measure the adaptability and efficacy of features learned by LGAfford-Net for the task of point cloud classification, demonstrating its potential utility in scenarios where dataset-specific training data may be limited or unavailable.

Overall, the results highlight the efficacy of our proposed methodology for affordance detection, showcasing its superiority over existing methods and its capability to accurately perceive and understand affordances within 3D environments.

4. Conclusions

In this paper, we have introduced LGAfford-Net, a novel architecture specifically tailored for affordance detection in 3D point clouds. Unlike prior approaches that overlooked local context and semantic cues, the proposed LGAfford-Net takes into consideration of semantic and local geometries for affordance detection. Through extensive experimentation on the 3D-AffordanceNet dataset, along with comparative analyses against state-of-the-art methods, we have demonstrated the improved performance of our proposed architecture. Comprehensive analysis on benchmark datasets shows that LGAfford-Net exhibits greater robustness compared to 3DAffordanceNet in detecting affordances within 3D point clouds even when groundtruth labels are unavailable. We observe 3.9↑ increase in average mIoU across 18 affordance classes, and maximum of 24.7↑ increase in mIoU (for *"Pull"* affordance). We showcase the generalizability of the extracted features by LGAfford-Net through classification of 3D objects using the ModelNet40 dataset.

References

[1] Tejas Anvekar and Dena Bazazian. GPr-Net: Geometric Prototypical Network for Point Cloud Few-Shot Learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4178–4187, 2023. 2, 4

[2] Tejas Anvekar, Ramesh Ashok Tabib, Dikshit Hegde, and Uma Mudenagudi. Metric-KNN is All You Need. In *SIGGRAPH Asia 2022 Posters*, pages 1–2. 2022. 3

[3] Tejas Anvekar, Ramesh Ashok Tabib, Dikshit Hegde, and Uma Mudengudi. VG-VAE: a venatus geometry point-cloud variational auto-encoder. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2978–2985, 2022. 2, 4

[4] Matan Atzmon, Haggai Maron, and Yaron Lipman. Point Convolutional Neural Networks by Extension Operators. *CoRR*, abs/1803.10091, 2018. 2, 8

[5] Ching-Yao Chuang, Jiaman Li, Antonio Torralba, and Sanja Fidler. Learning to act properly: Predicting and explaining affordances from images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 975–983, 2018. 2

[6] Shengheng Deng, Xun Xu, Chaozheng Wu, Ke Chen, and Kui Jia. 3D AffordanceNet: A benchmark for visual object affordance understanding. In *proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1778–1787, 2021. 1, 2, 5, 6, 7, 8

[7] Matheus Gadelha, Rui Wang, and Subhransu Maji. Multiresolution Tree Networks for 3D Point Cloud Processing. In *ECCV*, 2018. 8

[8] James J Gibson. *The ecological approach to visual perception: classic edition*. Psychology press, 2014. 1

[9] Meng-Hao Guo, Jun-Xiong Cai, Zheng-Ning Liu, Tai-Jiang Mu, Ralph R. Martin, and Shi-Min Hu. PCT: Point Cloud Transformer. *Computational Visual Media*, 7(2):187–199, 2021. 2, 8

[10] Mohammed Hassanin, Salman Khan, and Murat Tahtali. Visual affordance and function understanding: A survey. *ACM Computing Surveys (CSUR)*, 54(3):1–35, 2021. 2

[11] Deepti Hegde, Dikshit Hegde, Ramesh Ashok Tabib, and Uma Mudenagudi. Relocalization of camera in a 3d map on memory restricted devices. In *Computer Vision, Pattern Recognition, Image Processing, and Graphics: 7th National Conference, NCVPRIPG 2019, Hubballi, India, December 22–24, 2019, Revised Selected Papers 7*, pages 548–557. Springer, 2020. 2

[12] Girish Hegde, Tushar Pharale, Soumya Jahagirdar, Vaishakh Nargund, Ramesh Ashok Tabib, Uma Mudenagudi, Basavaraja Vandrotti, and Ankit Dhiman. DeepDNet: Deep dense network for depth completion task. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2190–2199, 2021. 2

[13] Jinyung Hong and Theodore P. Pavlic. KCNet: An Insect-Inspired Single-Hidden-Layer Neural Network with Randomized Binary Weights for Prediction and Classification Tasks. *CoRR*, abs/2108.07554, 2021. 8

[14] Mahek Jain, Guruprasad Kamat, Rochan Bachari, Vinayak A Belludi, Dikshit Hegde, and Ujwala Patil. AfforDrive: Detection of Drivable Area for Autonomous Vehicles. In *International Conference on Pattern Recognition and Machine Intelligence*, pages 532–539. Springer, 2023. 2

[15] Siddharth Katageri, Shashidhar V Kudari, Akshaykumar Gunari, Ramesh Ashok Tabib, and Uma Mudenagudi. ABD-Net: Attention based decomposition network for 3d point cloud decomposition. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2049–2057, 2021. 2

[16] Siddharth Katageri, Sameer Kulmi, Ramesh Ashok Tabib, and Uma Mudenagudi. PointDCCNet: 3d object categorization network using point cloud decomposition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2200–2208, 2021. 2

[17] Hema S Koppula and Ashutosh Saxena. Anticipating human activities using object affordances for reactive robotic response. *IEEE transactions on pattern analysis and machine intelligence*, 38(1):14–29, 2015. 2

[18] Hema Swetha Koppula, Rudhir Gupta, and Ashutosh Saxena. Learning human activities and object affordances from rgb-d videos. *The International journal of robotics research*, 32(8):951–970, 2013. 2

[19] Oliver Kramer and Oliver Kramer. K-nearest neighbors. *Dimensionality reduction with unsupervised nearest neighbors*, pages 13–23, 2013. 3

[20] Akash Kumbar, Tejas Anvekar, Ramesh Ashok Tabib, and Uma Mudenagudi. ASUR3D: Arbitrary Scale Upsampling and Refinement of 3D Point Clouds using Local Occupancy Fields. In *2023 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, pages 1644–1653, 2023. 2

[21] Akash Kumbar, Tejas Anvekar, Tulasi Amitha Vikrama, Ramesh Ashok Tabib, and Uma Mudenagudi. TP-NoDe: Topology-aware Progressive Noising and Denoising of Point Clouds towards Upsampling. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2272–2282, 2023. 2

[22] Yongcheng Liu, Bin Fan, Shiming Xiang, and Chunhong Pan. Relation-shape convolutional neural network for point cloud analysis. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8895–8904, 2019. 2, 8

[23] Austin Myers, Ching L Teo, Cornelia Fermüller, and Yiannis Aloimonos. Affordance detection of tool parts from geometric features. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1374–1381. IEEE, 2015. 2

[24] Shanthika Naik, Uma Mudenagudi, Ramesh Tabib, and Adarsh Jamadandi. Featurenet: Upsampling of point cloud and it's associated features. In *SIGGRAPH Asia 2020 Posters*, pages 1–2. 2020. 2

[25] Anh Nguyen, Dimitrios Kanoulas, Darwin G Caldwell, and Nikos G Tsagarakis. Object-based affordances detection with convolutional neural networks and dense conditional random fields. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 5908–5915. IEEE, 2017. 2

[26] Charles R. Qi, Hao Su, Kaichun Mo, and Leonidas J. Guibas. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. 2, 8

[27] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space. In *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2017. 2, 8

[28] Siyuan Qi, Siyuan Huang, Ping Wei, and Song-Chun Zhu. Predicting human activities using stochastic grammar. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1164–1172, 2017. 2

[29] Shi Qiu, Saeed Anwar, and Nick Barnes. Geometric Back-projection Network for Point Cloud Classification. *IEEE Transactions on Multimedia*, 2021. 2, 4

[30] Antonio Rodríguez-Sánchez, Simon Haller-Seeber, David Peer, Chris Engelhardt, Jakob Mittelberger, and Matteo Saveriano. Affordance detection with Dynamic-Tree Capsule Networks. In *2022 IEEE-RAS 21st International Conference on Humanoid Robots (Humanoids)*, pages 873–879. IEEE, 2022. 2

[31] Anirban Roy and Sinisa Todorovic. A multi-scale cnn for affordance segmentation in rgb images. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14*, pages 186–201. Springer, 2016. 2

[32] T Santoshkumar, Deepti Hegde, Ramesh Ashok Tabib, and Uma Mudenagudi. Refining SfM reconstructed models of indian heritage sites. In *SIGGRAPH Asia 2020 Posters*, pages 1–2. 2020. 2

[33] Johann Sawatzky, Abhilash Srikantha, and Juergen Gall. Weakly supervised affordance detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2795–2804, 2017. 2

[34] Hyun Oh Song, Mario Fritz, Daniel Goehring, and Trevor Darrell. Learning to detect visual grasp affordance. *IEEE Transactions on Automation Science and Engineering*, 13 (2):798–809, 2015. 2

[35] Ramesh Ashok Tabib, Yashaswini V Jadhav, Swathi Tegginkeri, Kiran Gani, Chaitra Desai, Ujwala Patil, and Uma Mudenagudi. Learning-based hole detection in 3D point cloud towards hole filling. *Procedia Computer Science*, 171:475–482, 2020. 2

[36] Ramesh Ashok Tabib, T Santoshkumar, Dikshit Hegde, Adarsh Jamadandi, and Uma Mudenagudi. Modeling nuisance classifier towards class-incremental learning of crowd-sourced data. In *Proceedings of the Twelfth Indian Conference on Computer Vision, Graphics and Image Processing*, pages 1–7, 2021. 4

[37] Ramesh Ashok Tabib, Dikshit Hegde, Tejas Anvekar, and Uma Mudenagudi. DeFi: Detection and Filling of Holes in Point Clouds Towards Restoration of Digitized Cultural Heritage Models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1603–1612, 2023. 2, 4

[38] Ramesh Ashok Tabib, Nitishkumar Upasi, Tejas Anvekar, Dikshit Hegde, and Uma Mudenagudi. IPD-Net: SO (3) Invariant Primitive Decompositional Network for 3D Point Clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2735–2743, 2023. 2

[39] Vishwanath S Teggihalli, Ramesh Ashok Tabib, Adarsh Jamadandi, and Uma Mudenagudi. A polynomial surface fit algorithm for filling holes in point cloud data. In *Pattern Recognition and Machine Intelligence: 8th International Conference, PReMI 2019, Tezpur, India, December 17-20, 2019, Proceedings, Part I*, pages 515–522. Springer, 2019. 2

[40] Tuan-Hung Vu, Catherine Olsson, Ivan Laptev, Aude Oliva, and Josef Sivic. Predicting actions from static scenes. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, pages 421–436. Springer, 2014. 2

[41] Chu Wang, Babak Samari, and Kaleem Siddiqi. Local Spectral Graph Convolution for Point Set Feature Learning. *arXiv preprint arXiv:1803.05827*, 2018. 8

[42] Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E. Sarma, Michael M. Bronstein, and Justin M. Solomon. Dynamic Graph CNN for Learning on Point Clouds. *ACM Transactions on Graphics (TOG)*, 2019. 2, 3, 8

[43] Zhirong Wu, Shuran Song, Aditya Khosla, Xiaoou Tang, and Jianxiong Xiao. 3D ShapeNets for 2.5D Object Recognition and Next-Best-View Prediction. *CoRR*, abs/1406.5670, 2014. 2, 5, 6, 8

[44] Hengshuang Zhao, Li Jiang, Jiaya Jia, Philip H. S. Torr, and Vladlen Koltun. Point Transformer. *CoRR*, abs/2012.09164, 2020. 8

[45] Bolei Zhou, Hang Zhao, Xavier Puig, Tete Xiao, Sanja Fidler, Adela Barriuso, and Antonio Torralba. Semantic understanding of scenes through the ade20k dataset. *International Journal of Computer Vision*, 127:302–321, 2019. 2