

Learning diffusion on global graph: A PDE-directed approach for feature detection on geometric shapes



Nannan Li^a, Shengfa Wang^{b,*}, Risheng Liu^b, Ziqiao Guan^c, Zhixun Su^d,
Zhongxuan Luo^{b,e}, Hong Qin^c

^a Information Science and Technology, Dalian Maritime University, China

^b DUT-RU International School of Information and Software Engineering, Dalian University of Technology, China

^c Computer Science, Stony Brook University, United States of America

^d School of Mathematical Science, Dalian University of Technology, China

^e Institute of Artificial Intelligence, Guilin University of Electronic Technology, China

ARTICLE INFO

Article history:

Available online 14 June 2019

Keywords:

Partial differential equations (PDEs)

Global graph

Submodularity

Small-sample learning

Feature detection

ABSTRACT

Feature and saliency analyses are crucial for various graphics applications. The key idea is to automatically compute and recommend the salient or outstanding regions of concerned models. However, there is no universally-applicable criterion for the detection results stemming from the personalized viewpoints for interest features on each specific model. This paper proposes a human-oriented feature detection framework, learning diffusion on global graph (LDGG), to understand personalized interests in a simple and low-cost way. A user-friendly interaction method is introduced to incorporate specific human interests as detection criteria in a small training set. Given a test model, we model the interest feature detection process as partial differential equations (PDEs)-directed diffusion on the global graph composed of nodes extracted from all training and test models. To infer the real interest points of users, submodular optimization is employed to select the source seeds adaptively for the diffusion system. By introducing diffusion guidance based on interest information, the PDEs become learnable. Extensive experiments and comprehensive comparisons have exhibited many attractive advantages of the proposed framework, such as capable of small-sample learning, easy-to-implement, extendable, self-correction, discriminative power, etc.

© 2019 Elsevier B.V. All rights reserved.

1. Introduction

Feature analysis and saliency detection have always been the central problems for mainstream graphics applications such as shape segmentation, retrieval, recognition, etc. (Bronstein et al., 2011; Kin-Chung Au et al., 2012; Li et al., 2016). The main tasks for these basic analyses are to compute and recommend the salient or discriminative geometric elements. However, since the regions of interest on the models differ among different people and tasks, a universal detection criterion does not exist. Furthermore, as newly emerged applications (Wang et al., 2015; Sipiran and Bustos, 2012) require high-level or semantic information abstracted from the models, feature detection should not be confined to detecting exactly similar regions (regions that overlap well after matching), but also considers functional

* Corresponding author.

E-mail address: sfwang@dlut.edu.cn (S. Wang).

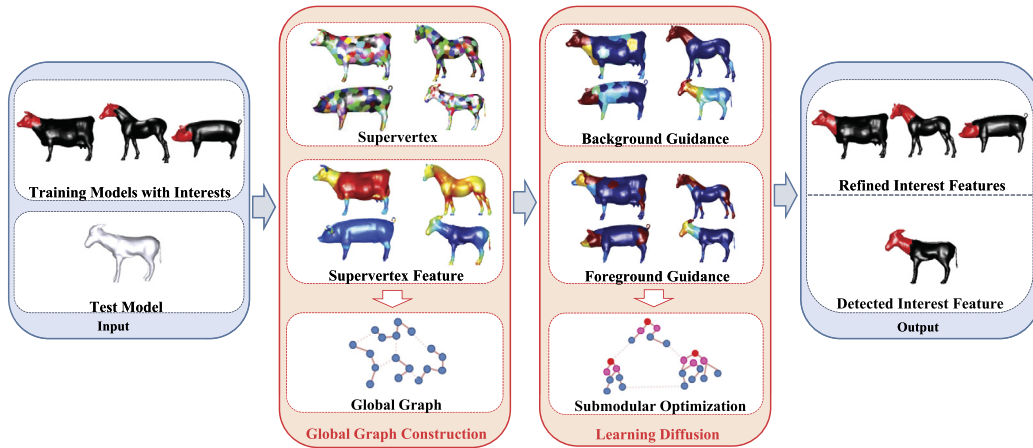


Fig. 1. The functional pipeline of our LDGG framework.

or semantic similarities. Therefore, in this paper, we propose a human-oriented interest feature detection framework to understand specific human interests via small-sample learning.

Existed works on feature analysis mainly focus on depicting and detecting the point features, line features, and regional features. By analyzing these shape elements based on their geometric attributes (normal, curvature, etc.), the features can effectively highlight the salient regions in a geometry aware sense. However, early works on feature detection mainly aim to find the exact similar elements across different models and ignore the functional and semantic information contained. Mesh saliency detection (Wang et al., 2015; Song et al., 2014) aims to find the most eye-catching regions on meshes and has been gaining momentum these years. Local saliency tries to capture the local geometric changes by using attributes like curvatures, Zernike map, spectrum, and so on. In comparison, global saliency describes geometric elements through the distribution of features regarding the entire model, e.g., global rarity and low rank analysis. Though saliency analysis claims to find the salient regions from the perspectives of humans, the analysis methods are usually lack of considerations for the actual demand of users and specific applications, nor are there proper evaluation criteria.

Current works considering high-level and semantic features of geometric elements mainly include labeling and segmentation (Kalogerakis et al., 2010; Huang et al., 2010). The central idea of these two connected tasks is to figure out the separate parts of the models concerned in terms of geometry functionality and semantic information. There are numerous methods with some success, including those based on unsupervised learning (Sidi et al., 2011), semi-supervised learning (Huang et al., 2010; Wang et al., 2012; Lv et al., 2012), supervised learning (Kalogerakis et al., 2010; Hu et al., 2012), and deep learning (Guo et al., 2015; Yi et al., 2016; Boscaini et al., 2016a). Unsupervised methods compute the shape features of concerned models and analyze them together to utilize the joint information. Supervised and deep learning methods usually can achieve somewhat high-precision results, however, they require intensive training with large volume of data. Semi-supervised methods try to combine geometric information with manual labelings via iterative user-interactive refinements. Due to the purposes of the labeling and segmentation tasks, the parts considered are all structure-aware but confined to some fixed scales with regular boundaries. As for interest feature detection, the scales and shapes of the feature regions concerned should not be restricted because interests vary among different people and applications.

In this paper, we propose a LDGG framework (as shown in Fig. 1) to automatically learn to understand human interests from a very small training set. By introducing a simple interest feature specification process, we can incorporate human interests directly into the training models. We model the interest feature detection problem as a learned anisotropic diffusion guided by PDEs on a global graph, which is constructed with nodes (i.e., supervertices, which will be introduced later) extracted from all training and test models. To infer the real interest points of users, we utilize the submodular optimization to select the source seeds for diffusion. Then we introduce the guidance based on geometric and interest information to make the PDE-directed diffusion learnable. Our framework can understand the real interests of human, and figure out the potential interest regions on randomly given test model, and simultaneously refine the mis-specified interest features on the training models. The primary contributions of this paper can be summarized as follows:

- We propose a simple but effective personalized feature detection framework to recommend potential interest regions on arbitrary models.
- We formulate the interest understanding task as PDEs-directed learned diffusion on a global graph using small training sets (with simple, user-interactive interest specifications).
- We utilize the submodular optimization to find the real interest points of users and exploit guidance map to guide the interest diffusion, making the PDEs adaptive and learnable.

- We extend our framework to the graphics task of mesh labeling and show its attractive properties under various challenging situations.

2. Related work

Exact regional feature detection. Till now, several categories of techniques have been employed with respect to exact regional feature detection or partial detection. Skeletal-graph-based approaches (e.g., Biasotti and Marini, 2006) co-analyze geometry and structure in the skeletal descriptor based on the Reeb graph theory, however, the independent sub-parts cannot always be recognized automatically. Bag-of-words (BoWs) technique is adopted in Bronstein et al. (2011) to represent a subject as a collection of feature signatures quantized in “geometric words”, and it depends much on the scale and quality of the geometric vocabulary. Multi-criterion optimization approaches (Lavoué, 2011; Lavoué, 2012) match subparts by balancing between significance and similarity. These methods require the prior knowledge of shape correspondence, and if not given, it can only be solved via time-consuming alternation between correspondence and part area. Non-point-wise correspondence is first proposed in Pokrass et al. (2013) by optimizing over the whole integration domain with descriptors based on integral operators. It can precisely match the fragments with entire shapes when the matching regions are exactly similar. Furthermore, most of the mentioned approaches rely largely on exact or meaningful shape decomposition results as in Lavoué (2012) or adopt the post-processing as done in Gal and Cohen-Or (2006) for more meaningful results, both of which are computationally expensive.

Segmentation and semantic labeling. Segmentation and labeling, considering meaningful and semantic parts, have shown their significance and attracted much attention in recent years. Till now, many approaches have been proposed for mesh labeling (Kalogerakis et al., 2010; Huang et al., 2010). In these works, the faces of a triangular mesh are usually characterized by heuristically designed geometric features. For instance, shape diameter function is utilized in Shapira et al. (2010) to partition a 3D object, where a context-aware distance measure is defined to figure out part similarity among objects. Subsequently, Sidi et al. (2011) proposed an unsupervised co-segmentation algorithm, which firstly provided a pre-defined feature set for per-object segmentation, and then embedded the segments into a common space by using diffusion maps. These approaches oftentimes achieve promising performance, however, due to the limited feature space, they often lack the ability to generate accurate and consistent labels for models with large style variations. Later works propose to address the problems by introducing high-level knowledge. A data-driven approach is proposed in Kalogerakis et al. (2010) to label the mesh by optimizing the conditional random field (CRF). Lv et al. (2012) presented a semi-supervised approach to reduce the influence of the size of the training set. Xie et al. (2014) proposed a fast method by employing a set of features and the extreme learning machine (ELM) to train a neural network classifier. Liu et al. (2015) proposed to label the test model via low-rank analysis of the feature space consisted of all the training and test models. Guo et al. (2015) aimed to learn mesh representations from a large pool of geometric features with the deep convolutional neural networks (CNNs). By using labeled meshes for training, these methods successfully learn various types of segmentations for corresponding tasks. However, the training set should be large enough and be labeled consistently, or the results may be undesirable.

PDEs in image and graphics processing. Conventional methods to design PDEs can be roughly divided into two categories: direct and variational methods. Direct methods (e.g., curve evolutions Sapiro, 2006) directly document PDEs based on the mathematical understandings on the physical problems (e.g., heat flow) or the geometric characteristics (e.g., curvature) of the problems. In contrast, variational methods (Scherzer et al., 2008) and total variation (TV) functionals (Rudin et al., 1992) define an energy to describe the desired properties of certain problem and derive PDEs by the Euler-Lagrange equation or its associated flows. In recent years, PDEs have been widely used in image processing such as restoration, inpainting, and multi-scale representations. In the primal works, PDEs-directed diffusions are regularly performed on image pixels and the differential operators are defined on Cartesian grids of the local image domains. Not surprisingly, these diffusions can only reflect interactions on the image locally. The graph-based PDEs are proposed in Liu et al. (2016) to capture the visual saliency. As for graphics processing, PDEs are mainly used to govern the construction of modeled surfaces, namely, PDE surfaces (Du and Qin, 2003). It emerges as a powerful modeling technique and gains popularity for surface modeling and design, and it allows geometric objects to be defined and governed by a group of differential equations. Besides, diffusion geometry (Sun et al., 2009; Wang et al., 2013; Boscaini et al., 2016b) has widely exploited PDEs in describing the local-to-global geometric and topological variations of mesh models, empowering applications such as feature abstraction, shape retrieval, and segmentation. However, PDE-directed diffusions in these works are only performed on the vertices of one single model, and have not yet been fully exploited in graph settings, where multiple shapes may be considered as a whole.

3. Problem statement and overview

The key contribution of this paper is a new approach to detect interest feature by learning diffusion on a global graph. The goal is to detect and recommend the potential interest regions on a new given model after user interests are specified on a small training set.

Since the existed feature and saliency detection works lack the consideration of high-level human requirements in specific applications as well as the evaluation criteria for analysis results, we propose to understand and detect the personalized

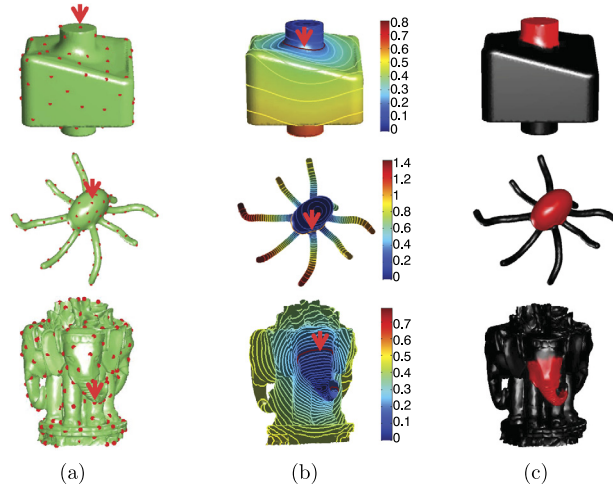


Fig. 2. Interests specified on different examples. Red arrows in (a) and (b) denote the specified points of interest and the specified interest scopes, respectively. (c) shows various specified interest regions. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

interest features. To directly incorporate human interests, we design a simple but flexible specification process, providing criteria for the detection results.

When the interests are specified on test models, our task aims to transfer the interest information from training samples to the test model. We model this problem as a diffusion process (guided by PDEs), namely, to diffuse the interest score from the specified regions on the training models to the test model. First, we establish the base for diffusion, namely, the global graph. We construct the graph utilizing supervertices abstracted from all the training and test models, and set up the connections and weights based on geometric and interest information.

Then, the source seeds should be selected for the diffusion, and the graph nodes within the specified interest features provide the backup set. However, there may exist redundancy and inaccuracy in the set, so we have to infer the real interest points of human and select the ideal seeds. On one hand, user interactions may be not precise, and on the other hand, the scales of regions of interest specified may vary drastically from a small local patch to a much larger one (combining multiple functional parts of an object). With a small set of training samples, we assume that recurring regions within specifications can best reflect the real human interests. Therefore, we introduce submodular optimization for seed selection, which can afford a reasonable distribution of the seeds and is capable of finding the maximal global interest score with adaptive PDE boundaries.

With the seed selection scheme prepared, we further propose to make the interest diffusion process aware of human interests by guiding the diffusion directions. To realize this, two diffusion guidance, foreground and background guidance, are proposed based on both geometric information and interest information. By incorporating the diffusion guidance, we construct adaptive and learnable PDEs to make the interest diffuse more wisely to detect the potential human interests on the test model and refine specifications on training models.

4. Interest diffusion on global graph

In order to perform the interest diffusion on a global graph, we shall first introduce the interest specification process and the construction of global graph.

4.1. Interest specifications on training models

Firstly, we perform a modified normalized cuts (Ncuts) (Golovinskiy and Funkhouser, 2008) with normal and shape diameter function (SDF) (Shapira et al., 2010) being combined to generate our so-called “supervertex”, which can partition the models with respect to the local geometry and represent the original models in a piecewise fashion (as shown in Fig. 1). The number of supervertices generated on each mesh (in our experiments) is roughly 200 and it may be set according to task requirements. Hereafter, we use lowercase bold letters (e.g., $\mathbf{q}$) to represent vector points and capital calligraphic ones (e.g., $\mathcal{S}$) to denote sets of points. Here we denote the center of a supervertex as $\mathbf{p}$ and the supervertex set (of all training and test models concerned) as $\mathcal{V}$.

Interest specification can be performed either on the supervertices or on the original vertices, and various methods may be used here. Here we introduce one specification method on the original mesh as shown in Fig. 2. We let the user specify their interest on the mesh vertices. On the supervertices, we set those containing over a certain portion of specified vertices as regions of interests. The interest region specified is defined with two terms: the central point of interest and the interest

scope. The central point can be picked directly from the vertices as shown in Fig. 2(a), or the extreme points of a function defined on the mesh. The interest scope is set with respect to the metric of the bi-harmonic distance field (Lipman et al., 2010), and the distance between the i -th and the j -th vertex, $\mathbf{p}_i$ and $\mathbf{p}_j$, is:

$$D_{bh}(i, j)^2 = \sum_{k=1}^m \frac{(\chi_k(i) - \chi_k(j))^2}{\gamma_k^2}, \quad (1)$$

where $\{\gamma_k\}$ and $\{\chi_k(\cdot)\}$ are, respectively, the first m non-zero eigenvalues and their corresponding eigenfunctions of the Laplacian-Beltrami operator.

Here Eq. (1) is used to define the distance field of the specified interest point. By constructing a set of contours (Fig. 2(b)) across the entire model, scales are thus provided for users to select and set the scope of the interest region. Some examples with different selected interest scopes are shown in Fig. 2(b), and Fig. 2(c) illustrates the interest regions specified in red with the black backgrounds. It can be seen that the scales of the regions of interest are flexible and can be adjusted to characterize the interest objects very well.

4.2. Global graph construction

The global graph in our approach is constructed by considering both the geometric information and the human interest information. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ represent an undirected graph, where $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ is a finite set of edges, the weights of which are decided by the graph affinities that are set as follows.

Feature-driven graph affinity. We first incorporate the feature information and the following feature descriptors are employed to depict each supervertex on the training models and the test model: SDF, Gauss Curvature (GC) (Gal and Cohen-Or, 2006), Conformal Factor (CF) (Ben-Chen and Gotsman, 2008), Shape Context (SC) (Belongie et al., 2002), Average Geodesic Distance (AGD) (Hilaga et al., 2001), and Wave Kernel Signature (WKS) (Aubry et al., 2008). The six descriptors are first computed on faces or vertices, and then on supervertex level, we compute a histogram for each descriptor to encode the feature distribution within each supervertex, each with 100 bins in our implementation. Finally, we concatenate the histograms to construct the feature vector for each supervertex, denoted as $\{\mathbf{h}_p, \mathbf{p} \in \mathcal{V}\}$.

In the high-dimensional feature space, we follow Liu et al. (2017) to build the affinity matrix using the shape interaction matrix which is indeed the row space projection of the feature matrix. It has been verified in Liu et al. (2012) that such matrix can significantly enhance the multiple subspace discrimination for the feature set. By assembling $\{\mathbf{h}_p, \mathbf{p} \in \mathcal{V}\}$ into a matrix $\mathbf{X}$, we analyze the linear relevance among the data via the row space projection of $\mathbf{X}$. Then the graph affinity $\mathbf{A}$ of $\mathcal{G}$ is defined as:

$$\mathbf{A} = \mathbf{Q}\mathbf{Q}^T, \quad (2)$$

where $\mathbf{Q}$ is obtained by the SVD decomposition of $\mathbf{X}$, i.e., $\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{Q}$.

Interest-driven and consistency-driven graph connection. Interest information is incorporated into graph construction by adjusting the connection of the global graph $\mathbf{A}$ as:

$$\mathbf{A}(i, j) = \begin{cases} \mathbf{A}(i, j) & \text{if } (i, j) \in \mathcal{E}, \\ 0 & \text{if } (i, j) \notin \mathcal{E}. \end{cases} \quad (3)$$

It is set under the following rules: First, if $\mathbf{p}_i$ and $\mathbf{p}_j$ both belong to the test model and are adjacent to each other, then $(i, j) \in \mathcal{E}$; Second, if $\mathbf{p}_i$ belongs to training models (test model), $\mathbf{p}_j$ belongs to test model (training models) $\mathcal{T}$ and $\mathbf{A}(i, j)$ is among the top m values of $\{\mathbf{A}(i, t) | \mathbf{p}_t \in \mathcal{T}\}$, then $(i, j) \in \mathcal{E}$; Third, if $\mathbf{p}_i$ belongs to interest region $\mathcal{U}$ and $\mathbf{A}(i, j)$ is among the top m values of $\{\mathbf{A}(i, t) | \mathbf{p}_t \in \mathcal{U}\}$, then $(i, j) \in \mathcal{E}$, m is empirically set to 5. Through such settings, the diffusion routes are designed to reflect the interest information obtained from the user's input.

4.3. Diffusion on global graph

We formulate the interest feature detection as the task of performing interest diffusion on the global graph $\mathcal{G}$. To be more intuitive and numeric, we define a real-value interest score function $f(\mathbf{p}) : \mathcal{V} \rightarrow \mathbb{R}$ (i.e., temperature in heat diffusion theory) to measure the interest level of $\mathbf{p}$. Then we denote interest seeds as $\mathcal{S}$ (i.e., heat source), and its corresponding scores as $f(\mathbf{p}) = s_p, \mathbf{p} \in \mathcal{S}$.

PDEs-guided diffusion. Mathematically, we can obtain the interest diffusion result as the stable status of an evolutionary PDEs with Dirichlet boundary condition, namely, a linear elliptic system with Dirichlet boundary (LESD):

$$F(f, \nabla f) = 0, f(\mathbf{e}) = 0, f(\mathbf{p}) = s_p, \mathbf{p} \in \mathcal{S}, \quad (4)$$

where $\nabla f = [f(\mathbf{p}) - f(\mathbf{q}_1), \dots, f(\mathbf{p}) - f(\mathbf{q}_{|\mathcal{N}_p|})]$, $\mathcal{N}_p = \{\mathbf{q}_1, \dots, \mathbf{q}_{|\mathcal{N}_p|-1}, \mathbf{e}\}$ is the neighborhood set of $\mathbf{p}$, $\mathbf{e}$ is an environment point with 0 score (outside $\mathcal{V}$). In our experimental settings, all the nodes are connected to an environment node $\mathbf{e}$, which

is used to set a lower bound to control the diffusion value of the system. F is a function of f and ∇f , and it can be any smooth function, here, we introduce an anisotropic diffusion term $\text{div}(\mathbf{K}\nabla f)$ to control the diffusivity:

$$F(f, \nabla f) = \text{div}(\mathbf{K}\nabla f), \quad (5)$$

where $\mathbf{K}$ is an inhomogeneous metric tensor defined (at $\mathbf{p}$) as:

$$\mathbf{K}_{\mathbf{p}} = \text{diag}(k(\mathbf{p}, \mathbf{q}_1), \dots, k(\mathbf{p}, \mathbf{q}_{|\mathcal{N}_{\mathbf{p}}|-1}), z_{\mathbf{e}}), \quad (6)$$

where $k(\mathbf{p}, \mathbf{q}) = \mathbf{A}(\mathbf{p}, \mathbf{q})$ as defined in Eq. (3), and $z_{\mathbf{e}}$ is a small constant (usually set to 0.1) for measuring the dissipation conductance at $\mathbf{p}$.

Task formulation. For the purpose of propagating the interest scores of seeds in $\mathcal{S}$ to all the other nodes in $\mathcal{V}$ with a stable status, we define an objective function to formulate and enforce the purpose of our task, that is, to maximize the sum of scores with regard to the whole domain of $\mathcal{V}$ when the diffusion is stable:

$$\begin{aligned} \max_{\mathcal{S} \in \mathbb{M}^n} T(\mathcal{S}) &= \sum_{\mathbf{p} \in \mathcal{V}} f(\mathbf{p}; \mathcal{S}), \\ \text{s.t. } F(f, \nabla f) &= 0, f(\mathbf{e}) = 0, f(\mathbf{p}) = s_{\mathbf{p}}, \mathbf{p} \in \mathcal{S}, \end{aligned} \quad (7)$$

where $\mathbb{M}^n = \{\mathcal{S} | \mathcal{S} \subset \mathcal{V}, |\mathcal{S}| \leq n\}$ is a uniform matroid to enforce the cardinality of $\mathcal{S}$ to be no more than a given number n , which in our tests is set as the number of supervertices in $\mathcal{S}$. And in our experiments, we set $s_{\mathbf{p}} = 1$.

Now, we have established a general diffusion system for our task. However, using all the specified supervertices (denoted as backup seed set $\tilde{\mathcal{S}}$) as seeds is not ideal. This is because, the geometric features are not effective on every node, and user specifications may contain scale ambiguities. Thus we will introduce a seed selection technique as follows.

5. LDGG framework for interest feature detection

For our interest feature detection task, the LESD system should simultaneously identify the most representative source seed set $\mathcal{S}$, and propagate the interest score wisely to extract the relevance between seeds and other graph nodes.

5.1. Submodular optimization for adaptive PDEs

It should be noted that, there exists significant redundancy in the backup seed set $\tilde{\mathcal{S}}$. Specifically, in our task, the recurred regions in specifications on all test models should be the real interests. So we introduce the submodular optimization to conduct the seed selection. The theorem proved in Kim et al. (2011), Liu et al. (2016) enables us to explore the good properties of the interest score function and the objective function in our problem.

The monotonicity and submodularity of T can be easily proved, and then by combining these properties with the uniform matroid constraint in Eq. (7), we can employ a greedy algorithm to solve Eq. (7) and a $(1 - 1/e)$ -approximation can be got (Nemhauser and Wolsey, 1978). The submodular function is characterized as a diminishing return property stating that, the marginal gain of adding an element to a smaller subset of $\mathcal{S}$ is higher than that of adding it to a larger subset of $\mathcal{S}$. That is to say, the submodular optimization tends to select the source seeds in relatively larger connected subgraph (thus is more representative). Undoubtedly, the property of submodular function can help select the recurred supervertices within the interest regions as seeds. And the optimization process is similar as that in Kim et al. (2011). One specific seed selection process for our example is shown in Fig. 3, the seeds are gradually selected in the labeled order.

5.2. Diffusion guidance for learnable PDEs

To make the interest diffusion aware of both geometric and interest information, we introduce a guidance to our PDEs-directed diffusion system.

Prior-guided PDEs. To realize this goal, we incorporate a fidelity term $\mathcal{P}_{\mathcal{V} \setminus \mathcal{S}}(f - f_u)$ to the governing function F , where f_u is a guidance map to direct the diffusion, $\mathcal{V} \setminus \mathcal{S}$ denotes the complement of $\mathcal{S}$ when $\mathcal{S} \subset \mathcal{V}$, and $\mathcal{P}_{\mathcal{V} \setminus \mathcal{S}}$ is the projection on $\mathcal{V} \setminus \mathcal{S}$. We will show in the following that, f_u can be learned from both interest and geometric information. And the governing equation finally becomes:

$$F_u(f, \nabla f) = \text{div}(\mathbf{K}\nabla f) + \lambda \mathcal{P}_{\mathcal{V} \setminus \mathcal{S}}(f - f_u), \quad (8)$$

where $\lambda \geq 0$ is a parameter to control the trade-off between the diffusion rate and the fidelity to the guidance, and it will be discussed in Section 6. Then we can discretize the PDE formulation as:

$$f(\mathbf{p}) = \begin{cases} \frac{1}{d_{\mathbf{p}} + \lambda} (\sum_{\mathbf{q} \in \mathcal{N}_{\mathbf{p}}} \mathbf{K}_{\mathbf{p}}(\mathbf{q}) f(\mathbf{q}) + \lambda f_u), & \mathbf{p} \in \mathcal{V} \setminus \mathcal{S}, \\ s(\mathbf{p}), & \mathbf{p} \in \mathcal{S}, \end{cases} \quad (9)$$

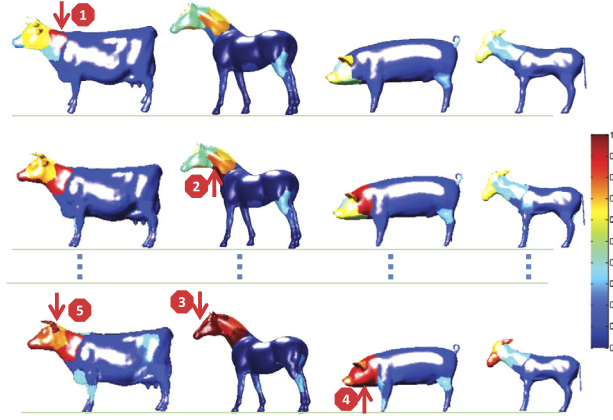


Fig. 3. Illustration of seed selection utilizing submodular optimization. The first and second rows show the first two selected seeds, and the last row shows several other seeds achieved in the labeled order.

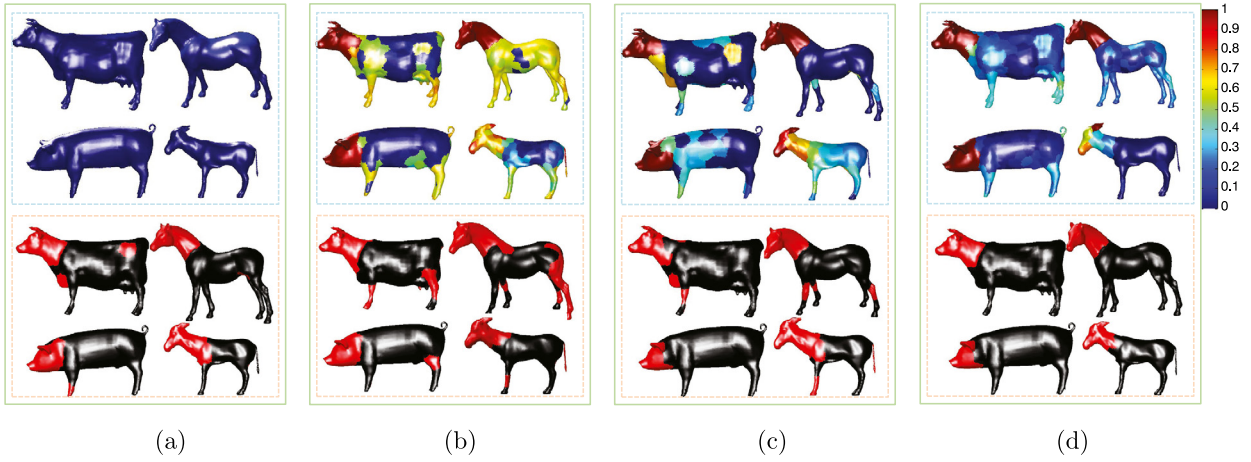


Fig. 4. Interest feature detection results (in the bottom row) with different guidance maps (in the top row). (a)-(d) show results without guidance, with foreground guidance, with background guidance, and with final combined guidance, respectively.

where $K_p(q)$ represents the diagonal element of K_p corresponding to q and $d_p = \sum_{q \in \mathcal{N}_p} K_p(q)$.

Diffusion guidance construction. Our diffusion guidance is constructed based on geometric and interest information by considering the interest and non-interest regions, namely, the regions most similar and most dissimilar to the specified interest regions, which constitute the final foreground and background guidance.

First, we achieve the foreground guidance by performing interest diffusion based on Eq. (9) with all nodes within $\tilde{S}$ as source seeds. We denote the foreground guidance as $f_{\mathcal{F}}$, Fig. 4(b) shows one case. Then, for the background guidance, we empirically select the top 70% dissimilar supervertices (according to the distance measurement in the feature space) for each interest supervertex, denoted as $\mathcal{B}_i$. Then the most non-interest regions $\mathcal{B}$ are obtained by taking the intersection of all $\mathcal{B}_i$, that is $\mathcal{B} = \cap \mathcal{B}_i$. We also adopt Eq. (9) to perform the non-interest score diffusion with the boundaries as $\mathcal{B}$, and the values of nodes in the final status (denoted as $\tilde{f}_{\mathcal{B}}$) represent probabilities of belonging to the non-interest region. Therefore, we define the background guidance as $f_{\mathcal{B}} = 1 - \tilde{f}_{\mathcal{B}}$ as shown in Fig. 4(c). In this view, we obtain the probability of a node belonging to the interest region, namely guidance map, as $f_u(p) = f_{\mathcal{F}}(p) + f_{\mathcal{B}}(p)$, which is shown in Fig. 4(d).

5.3. LDGG framework

Till now, we have introduced all the techniques required for interest feature detection, therefore we will integrate them as a whole and set up the combinatorial framework of LDGG.

In order to provide an adaptive way to identify the number of heat sources and further suppress the redundancy in $\tilde{S}$, we introduce a penalty term to the objective function as:

$$U(S) = \sum_{p \in S} u(p), \quad (10)$$

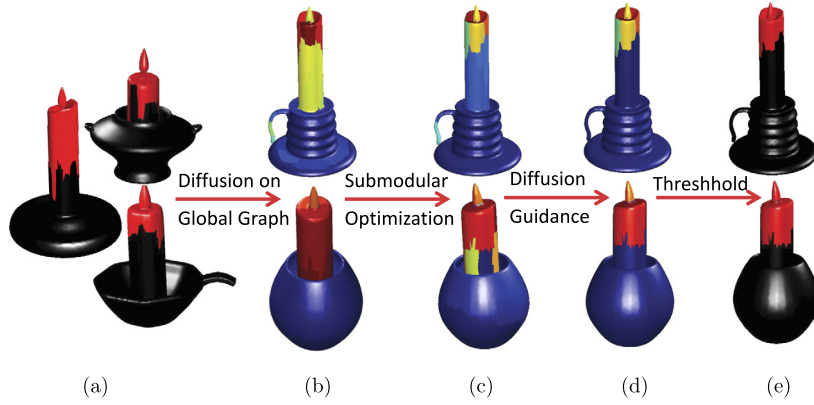


Fig. 5. Step-by-step illustration of our pipeline with (a) training models, (b) diffusion on global graph, (c) submodular optimization, (d) guidance map, and (e) detected interest regions.

Table 1

Running time for the cooperations (graph construction, diffusion, detection) on three training models and one test model (with the scale similar to the listed models) with average 20 seeds, and $\sharp$ SV denotes the number of supervertices.

Model	$\sharp$ SV	Timing (sec)			
		Graph	Diffusion	Detection	Total
Dinosaur	150	0.65	1.44	0.14	2.51
Dog	200	0.87	1.84	0.26	3.25
Armadillo	250	1.20	2.35	0.34	4.26
Santa	350	1.91	3.22	0.75	6.47
Dragon	500	2.65	4.52	1.23	8.56

where $u(\mathbf{p}) = \exp(-f_u(\mathbf{p})^2)$, and larger $u(\mathbf{p})$ implies that $\mathbf{p}$ has a higher probability of belonging to $\tilde{S} \setminus S$ and should be suppressed. Finally, we formulate the LDGG framework as the following combinatorial optimization:

$$\begin{aligned} \max_{S \in \mathbb{M}^n} O(S) &= T(S) - U(S), \\ \text{s.t. } F_u(f, \nabla f) &= 0, f(\mathbf{e}) = 0, f(\mathbf{p}) = s_{\mathbf{p}}, \mathbf{p} \in S. \end{aligned} \quad (11)$$

It should be noted that, the objective function O is also a submodular function, the optimization process stated above is also applicable for the LDGG framework by changing the objective function from T to O , and governing function from F to F_u . As the interest scores can be considered as the relevance between nodes and the seeds, the maximum criterion in Eq. (11) naturally tends to choose seeds in relatively larger connected subgraph, therefore, the nodes in interest regions with high local contrast (i.e., less paths or small connection weights to other nodes) will be removed from $\tilde{S}$. One may concern that non-interest nodes will also have a large O because they may connect to a large number of other non-interest nodes. Fortunately, our guidance map f_u can enforce small interest scores in non-interest regions, thus they cannot be included in S .

To clearly show the functionality of each technique in our framework, we show a step-by-step illustration of the pipeline in Fig. 5. It can be seen that, the global graph sets the foundation for diffusion, the submodular optimization helps select the real interest regions, the guidance map directs the diffusion to more reasonable regions, and the overall framework is capable of determining the interest regions robustly.

6. Experiments and discussions

This section will demonstrate the performance of our approach via experiments in various aspects. All the experiments are conducted on a 3.5 GHz Intel(R) Core(TM) i7 computer with 16 GB memory. We employ data from the Princeton Segmentation Benchmark (Chen et al., 2009) (19 categories, 20 meshes per category) and 4 additional categories including Lamp (20 meshes), Candelabra (28 meshes), Iron (18 meshes), and Guitar (44 meshes) from Sidi et al. (2011). Without any complex training process, the numerical optimization process of our framework is very fast. The timing details are shown in Table 1.

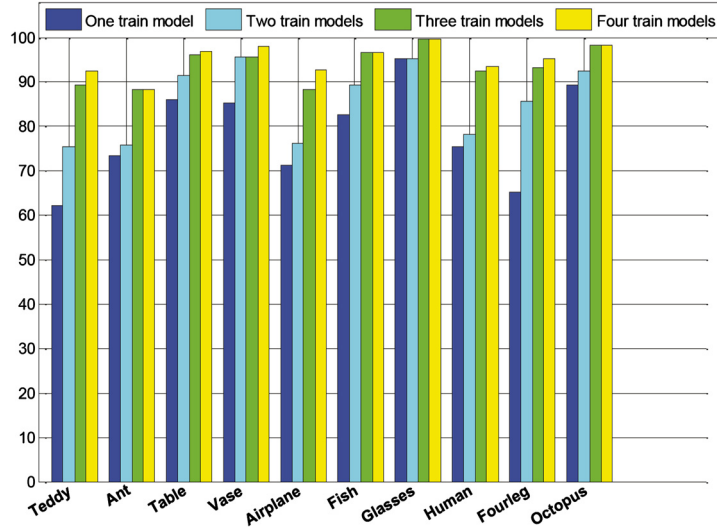


Fig. 6. Detection scores with different numbers of training models.

6.1. Parameter settings

The settings of parameters in our framework are almost stated in the corresponding sections with empirical values or optimal selections. Here we specifically make clear the settings of two most important parameters: the number of training models and the parameter λ in Eq. (8) for the trade-off between the diffusion rate and the fidelity to the guidance.

In order to demonstrate the influence of the number of training models, we randomly select one, two, three and four models from each category in the benchmark and calculate the average detection score as defined in Kalogerakis et al. (2010) and Lv et al. (2012):

$$Acc = \frac{\sum_i a_i \frac{I(y_i, y_i^*) + 1}{2}}{\sum_i a_i}, \quad (12)$$

where a_i is the area of the i -th face, y_i is the ground truth label for the i -th face, y_i^* is the estimated label for the i -th face, and $I(y, y^*)$ is an indicator function that is 1 when $y = y^*$ and -1 otherwise. Since our interest detection task is different from that of labeling, the labels in our work are just two (interest and non-interest), and we set up the ground truth labels manually for our problem specifically. The bars in Fig. 6 show the detection scores achieved for 10 categories with three training models randomly selected and the interest regions are specified with various scopes specifically. We can infer from the results that an average of three training models can achieve ideal detection results. To visualize the inside details, we show two cases in Fig. 7 with green and red boxes denoting interest detection with initial and additional training models respectively. It's obvious that interest regions on the corresponding models can be detected with higher precision when more training models are used.

Then the setting of parameter λ can be inferred from Fig. 8. It can be seen that an average of 0.6 can achieve relatively high recognition scores with respect to each category of models.

6.2. Experiments on interest feature detection

The interest feature detection results of our method for meshes from different categories are shown in Fig. 9. We specifically set the interest regions to be varying in scopes, and the detection results are all visually correct.

Robustness. In practice, the inconsistent and incorrect specifications from users are inevitable due to objective and subjective factors. Previous methods based on learning may over-fit the training data. The introduction of submodular optimization and guidance map can effectively make our method robust to both “under-specifications” (where some regions of interest are left out) and “over-specifications” (where some regions specified by user are not what they intend to select) (as illustrated in the first two rows of Fig. 10). Here, a large part of areas are mis-specified manually (highlighted with red dash circles), such as the over specification of table leg, the different-scale specifications on human legs, and the under specification of chair seat. Our algorithm can correctly detect the interest regions and simultaneously achieve the self-refinement as highlighted by the red circles. We also show in Fig. 10 (in the third row) that our algorithm can suppress the interferences of data imperfections to some level. We specifically add 0.3 mean-edge-length noises to the first human model, drill holes of different sizes on the second human model, and perform a 30% downsampling to the third model. It can be seen that the detection results are still correct. Besides, through the experiments on human legs (the scopes of the three specifications

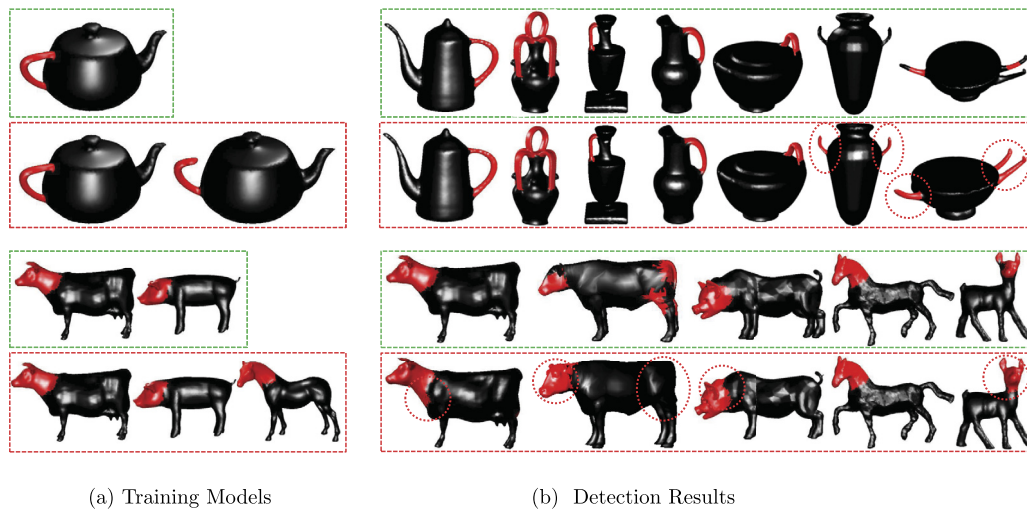


Fig. 7. The influence of the number of training models. Green and red boxes denote interest feature detection with initial and additional training models respectively, and red circles denote the modified detection results.

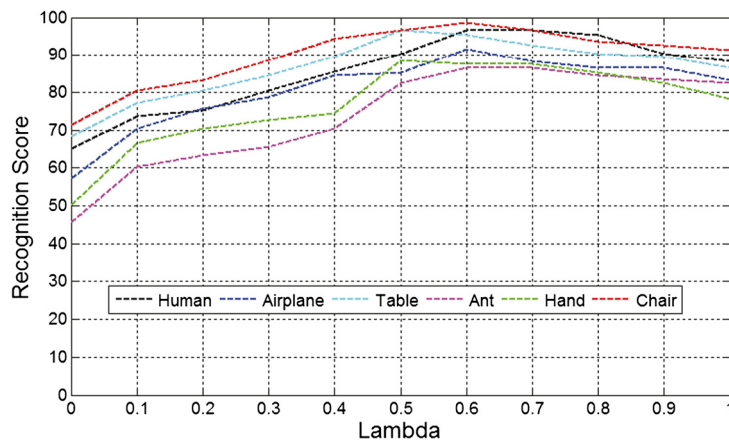


Fig. 8. Detection scores with different settings of parameter λ .

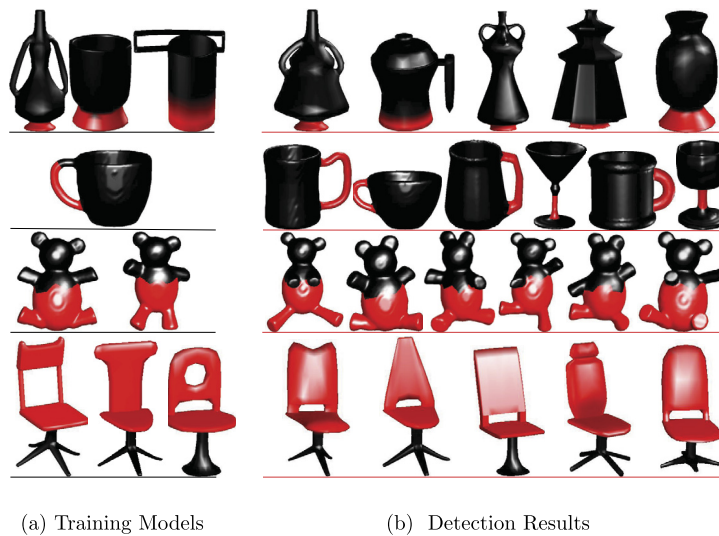


Fig. 9. Experiments on interest understanding on models from different categories.

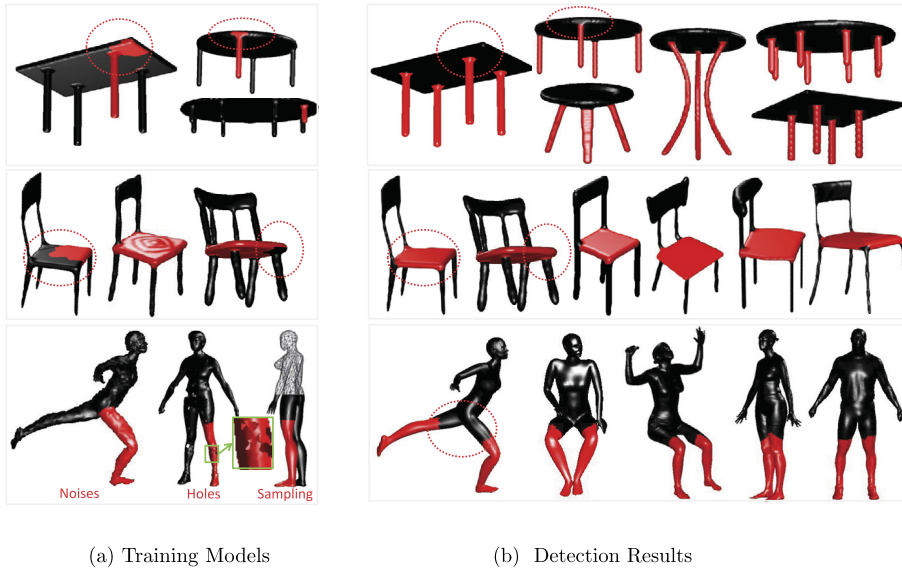


Fig. 10. Experiments on robustness. Our framework can tolerate the mis-specifications and error-specifications. The red circles highlight the mis-specifications and self-refined results.

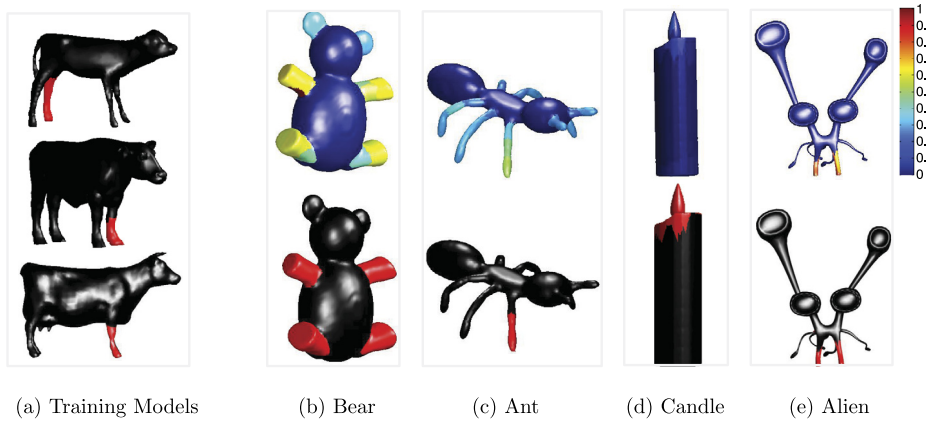


Fig. 11. Experiments on models that training and test categories are different. The upper models in (b)–(e) show the interest scores achieved via diffusion, and the under models show the recommended interest regions by setting a threshold.

are not all the same), we can see that our algorithm can exactly catch the scope that is really preferred by users thanks to the submodularity optimization process.

Cross-category interest understanding. We perform tests to show that our framework can recommend the interest regions not only on new models of the same object category as the training set, but also on models of different categories. Fig. 11 shows results in which users specify on the four-leg animals and test on four models from totally different categories. The upper rows of (b)–(e) show the interest diffusion results, the values may be seen as a generalized kind of saliency with respect to the interests of users. Then the threshold used to recommend the potential interest regions can be set according to different applications. Here, we just use the mean value of the diffusion results as threshold and obtain the results as shown in the second rows of (b)–(e). It can be seen that our framework can generate desired detection results even if the given examples exhibit significant shape variabilities.

Joint understanding with multiple categories. To demonstrate the expandability of our LDGG framework, we augment the global graph by expanding the categories of specified models. It is more challenging, since the features coming from different categories may interact with each other. Fig. 12 shows examples in which the training model sets consist of more than one category and the test models are chosen from the corresponding categories. From the detection results we can see that, our framework can be well extended to multi-category cases thanks to the seed selection and diffusion learning techniques.

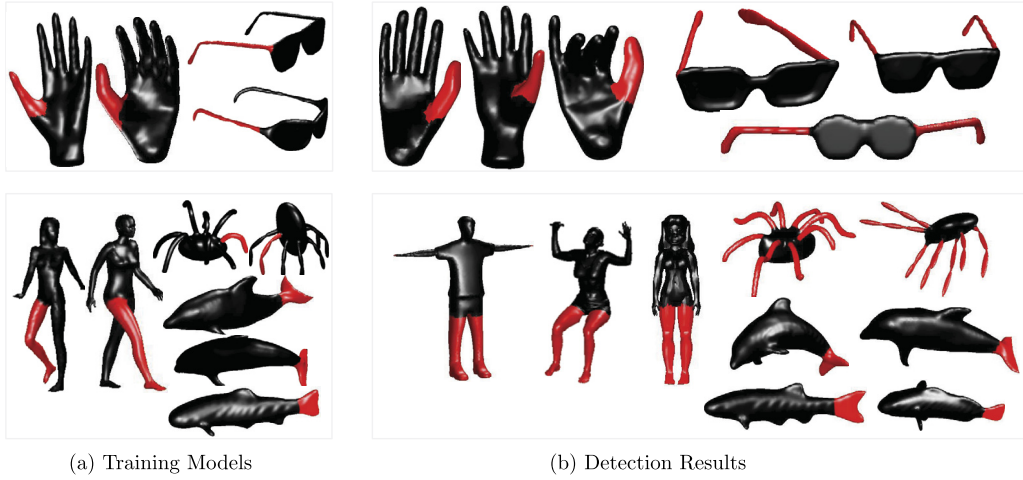


Fig. 12. Experiments on training models consisting of multiple categories.

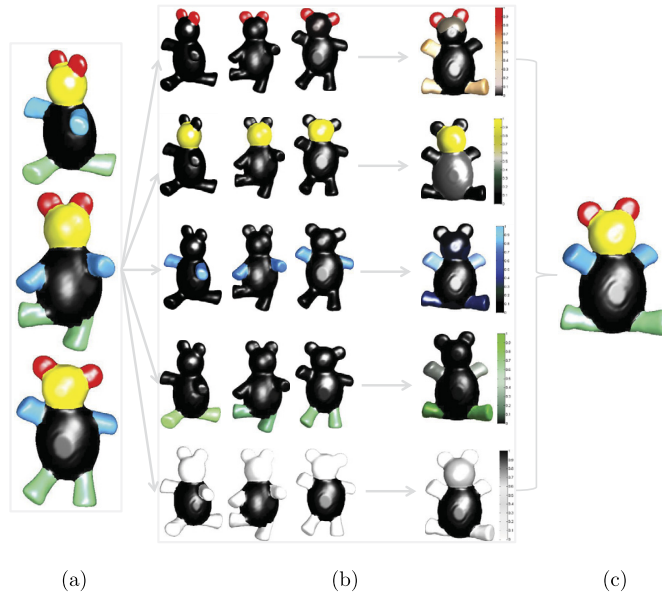


Fig. 13. Application of LDGG framework to multi-labeling. (a)–(c) show labeling samples, learned diffusion per label and the labeling result, respectively.

Algorithm 1 Applying LDGG to mesh labeling.

Require: Given $\mathcal{V}$, specified interest region $\tilde{\mathcal{S}}$ and number of labels on test models n_l .

Ensure: Labels l_i of each supervertex on test model.

1: Construct matrix $\mathbf{C}$ of n rows and n_l columns.

2: **for** $i = 1 \rightarrow n_l$ **do**

3: Perform interest diffusion for the i -th label based on the LDGG framework, and endow the stable diffused value to $\mathbf{C}(:, i)$.

4: **end for**

5: $l_i = \operatorname{argmax}_j \mathbf{C}(i, j)$.

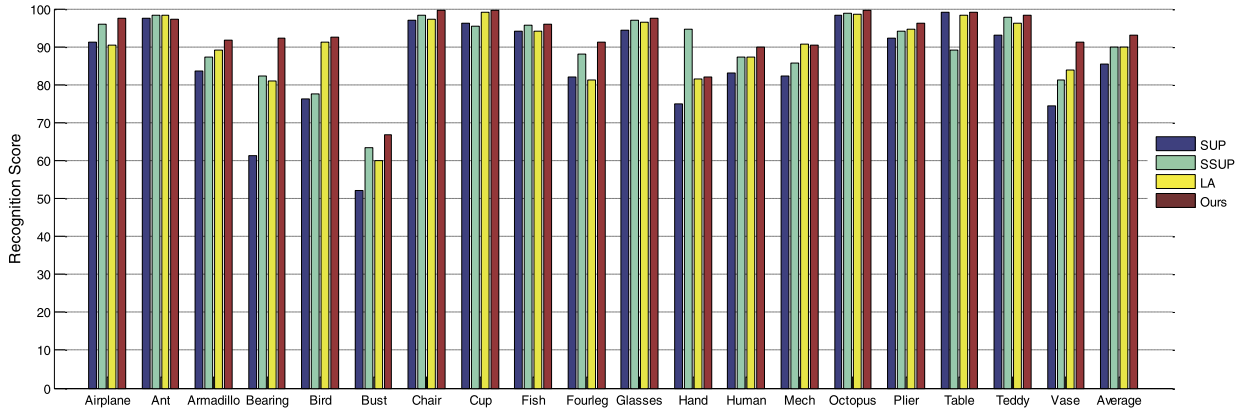
6.3. Comparisons

LDGG framework for mesh labeling. To make our framework extendable to multi-specifications and comparable with the most relative problem of mesh labeling, we apply our framework to labeling as follows. To implement labeling, we just see a label as a specification (as shown in Fig. 13(b)), and repeat the specification and detection process for the number of label times, and then decide the labels of supervertices on the test model according to the results of the diffusions. The above process is detailed in Algorithm 1.

Table 2

Recognition scores of various methods.

Dataset	SUP	SSUP	LA	Ours
Airplane	91.2	95.9	90.5	97.6
Ant	97.4	98.3	98.4	97.2
Armadillo	83.7	87.3	89.2	91.8
Bearing	61.3	82.4	81.1	92.3
Bird	76.3	77.5	91.3	92.5
Bust	52.2	63.4	60.0	66.8
Chair	97.1	98.3	97.2	99.6
Cup	96.3	95.3	99.1	99.6
Fish	94.1	95.7	94.1	96.0
Fourleg	82.0	88.1	81.2	91.2
Glasses	94.4	97.1	96.4	97.6
Hand	74.9	94.7	81.4	82.0
Human	83.2	87.3	87.2	89.8
Mech	82.4	85.6	90.6	90.5
Octopus	98.3	98.7	98.5	99.6
Plier	92.2	94.2	94.5	96.2
Table	99.0	89.1	98.3	99.0
Teddy	93.1	97.9	96.3	98.3
Vase	74.3	81.2	83.8	91.3
Average	85.4	89.9	90.0	93.1

**Fig. 14.** Comparisons of various methods.

Comparisons with labeling methods. We first compare our method with the low-rank analysis method (LA) (Liu et al., 2015), the supervised method (SUP) (Kalogerakis et al., 2010), the semi-supervised method (SSUP) (Lv et al., 2012) on the benchmark. The numeric comparisons are shown in Table 2 and are visualized in Fig. 14. The example models here are 3 randomly chosen meshes. Since the method proposed by Lv et al. (2012) is semi-supervised, 17 unlabeled meshes are chosen as examples as mentioned in Lv et al. (2012). As Table 2 shows, when the size of the training dataset is small, our method achieves better results than the SUP, SSUP and LA methods. Furthermore, the SUP and SSUP methods both require offline training processes, which is not needed by our method.

Furthermore, we make visual comparisons two learning methods, Torralba et al. (2007) and Kalogerakis et al. (2010), and a deep learning method (Guo et al., 2015) as in Fig. 15. Torralba et al. (2007) employs JointBoost, which is a linear classifier, and it can not work that well in face with the large and complex feature space. Kalogerakis et al. (2010) achieves perfect results since it uses hundreds of geometric and contextual label features for training and learns different types of segmentations for different tasks. Guo et al. (2015) can well represent the meshes by introducing multilayer convolutions and nonlinear mappings. It can be seen that, with the smallest training set, our method can achieve almost the same or even better results compared with the other training-based methods with respect to the head part. We also compare with Kalogerakis et al. (2010) and Guo et al. (2015) in cross-category detection as shown in Fig. 16. Without using any tedious training process, we just use three training models as shown in the third row of Fig. 16 and get more natural and meaningful results (e.g. the heads of the planes).

Limitation. One limitation of our approach is that it may have difficulties in dealing with large defects on both training and test models, such as the existence of big holes or missing large organic parts on interest regions, leading to the deficiency of information. So it will be our future work to make the framework applicable to models with various kinds of defects.

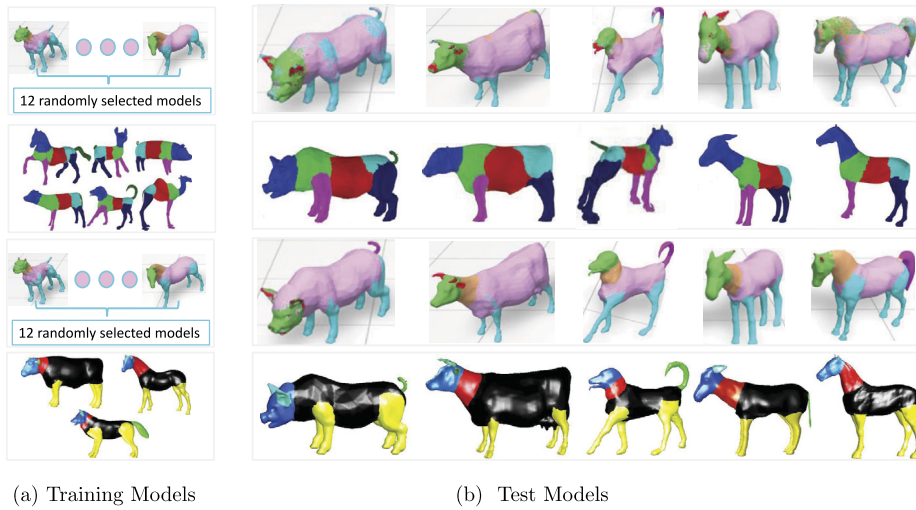


Fig. 15. Comparisons with two learning methods, Torralba et al. (2007) and Kalogerakis et al. (2010) (in the first two rows), and a deep learning method (Guo et al., 2015) (in the third row) (results are cited from the original papers), and the bottom row shows ours.

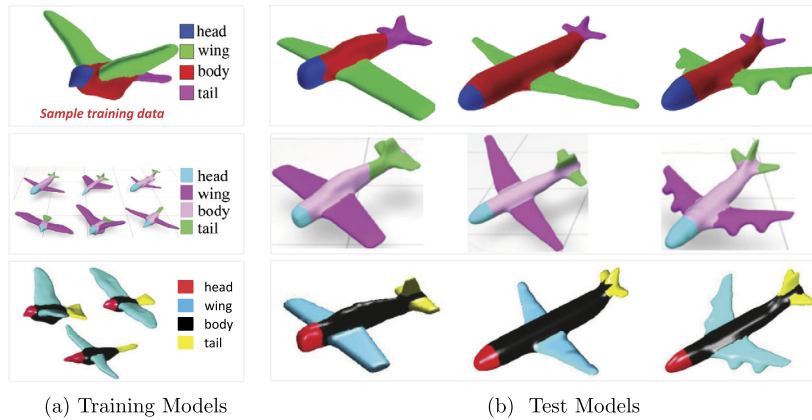


Fig. 16. Comparisons with Kalogerakis et al. (2010) (in the first row) and Guo et al. (2015) (in the second row) in cross-category detection with results cited from the corresponding papers, and the third row shows our results.

7. Conclusion and future work

We have described the first combinatorial framework via global-graph based PDEs in the area of graphics processing. The nature of our framework is a small-sample learning process that can understand human interests from just a few training samples with simple user specifications. We detailed the construction of the global graph, the introduction of submodularity optimization and the design of diffusion guidance, all of which make PDEs adaptive and learnable. Through extensive experiments, our algorithm obtains state-of-the-art interest understanding and multi-labeling results without the need of any tedious manually labeling and training process.

Our future work will try to broaden the applicability of our approach to raw scanned point cloud models with depth information, and to develop an efficient “app” for real-time interest understanding.

Declaration of Competing Interest

No conflict of interest.

Acknowledgements

This work is partially supported by the National Natural Science Foundation of China grants (61802045, 61772104, 61432003, 61572099, 2018ZX04041001-007, 2018ZX04016001-011), and the National Science Foundation of USA grants IIS-(0949467, 1047715, 1049448, 1715985).

References

- Aubry, M., Schlickewei, U., Cremers, D., 2008. The wave kernel signature: a quantum mechanical approach to shape analysis. In: IEEE International Conference on Computer Vision, pp. 1626–1633.
- Belongie, S., Malik, J., Puzicha, J., 2002. Shape matching and object recognition using shape contexts. *IEEE Trans. Pattern Anal. Mach. Intell.* 24 (4), 509–522.
- Ben-Chen, M., Gotsman, C., 2008. Characterizing shape using conformal factors. In: Proceedings of the 1st Eurographics Conference on 3D Object Retrieval, pp. 1–8.
- Biasotti, S., Marini, S., 2006. Sub-part correspondence using structure and geometry. In: Eurographics Italian Chapter Conference, pp. 23–28.
- Boscaini, D., Masci, J., Rodolà, E., Bronstein, M., 2016a. Learning shape correspondence with anisotropic convolutional neural networks. In: Advances in Neural Information Processing Systems. Curran Associates, Inc., pp. 3189–3197.
- Boscaini, D., Masci, J., Rodolà, E., Bronstein, M.M., Cremers, D., 2016b. Anisotropic diffusion descriptors. *Comput. Graph. Forum* 35 (2), 431–441.
- Bronstein, A.M., Bronstein, M.M., Guibas, L.J., Ovsjanikov, M., 2011. Shape google: geometric words and expressions for invariant shape retrieval. *ACM Trans. Graph.* 30 (1), 1–20.
- Chen, X., Golovinskiy, A., Funkhouser, T., 2009. A benchmark for 3d mesh segmentation. *ACM Trans. Graph.* 28 (3), 73.
- Du, H., Qin, H., 2003. Interactive shape design using volumetric implicit pdes. In: Proceedings of ACM Symposium on Solid Modeling and Applications, pp. 235–246.
- Gal, R., Cohen-Or, D., 2006. Salient geometric features for partial shape matching and similarity. *ACM Trans. Graph.* 25 (1), 130–150.
- Golovinskiy, A., Funkhouser, T., 2008. Randomized cuts for 3d mesh analysis. *ACM Trans. Graph.* 27 (5), 145:1–145:12.
- Guo, K., Zou, D., Chen, X., 2015. 3d mesh labeling via deep convolutional neural networks. *ACM Trans. Graph.* 35 (1), 3:1–3:12.
- Hilaga, M., Shinagawa, Y., Kohmura, T., Kunii, T.L., 2001. Topology matching for fully automatic similarity estimation of 3d shapes. In: Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques, pp. 203–212.
- Hu, R., Fan, L., Liu, L., 2012. Co-segmentation of 3d shapes via subspace clustering. *Comput. Graph. Forum* 31 (5), 1703–1713.
- Huang, Q.-X., Su, H., Guibas, L., 2010. Fine-grained semisupervised labeling of large shape collections. *ACM Trans. Graph.* 32 (6), 190.
- Kalogerakis, E., Hertzmann, A., Singh, K., 2010. Learning 3d mesh segmentation and labeling. *ACM Trans. Graph.* 29 (4), 102:1–102:12.
- Kim, G., Xing, E.P., Fei-Fei, L., Kanade, T., 2011. Distributed cosegmentation via submodular optimization on anisotropic diffusion. In: Proceedings of International Conference on Computer Vision, pp. 169–176.
- Kin-Chung Au, O., Zheng, Y., Chen, M., Xu, P., Tai, C.-L., 2012. Mesh segmentation with concavity-aware fields. *IEEE Trans. Vis. Comput. Graph.* 18 (7), 1125–1134.
- Lavoué, G., 2011. Bag of words and local spectral descriptor for 3D partial shape retrieval. In: Eurographics Conference on 3D Object Retrieval, pp. 41–48.
- Lavoué, G., 2012. Combination of bag-of-words descriptors for robust partial shape retrieval. *Vis. Comput.* 28 (9), 931–942.
- Li, N., Wang, S., Zhong, M., Su, Z., Qin, H., 2016. Generalized local-to-global shape feature detection based on graph wavelets. *IEEE Trans. Vis. Comput. Graph.* 22 (9), 2094–2106.
- Lipman, Y., Rustamov, R.M., Funkhouser, T.A., 2010. Biharmonic distance. *ACM Trans. Graph.* 29 (3), 1–11.
- Liu, G., Liu, Q., Li, P., 2017. Blessing of dimensionality: recovering mixture data via dictionary pursuit. *IEEE Trans. Pattern Anal. Mach. Intell.* 39 (1), 47–60.
- Liu, R., Lin, Z., De la Torre, F., Su, Z., 2012. Fixed-rank representation for unsupervised visual learning. In: 2012 IEEE Computer Vision and Pattern Recognition, pp. 598–605.
- Liu, R., Zhong, G., Cao, J., Lin, Z., Shan, S., Luo, Z., 2016. Learning to diffuse: a new perspective to design pdes for visual analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* 38 (12), 2457–2471.
- Liu, X., Zhang, J., Liu, R., Li, B., Jun, W., Cao, J., 2015. Low-rank 3d mesh segmentation and labeling with structure guiding. *Comput. Graph.* 4 (6), 99–109.
- Lv, J., Chen, X., Huang, J., Bao, H., 2012. Semi-supervised mesh segmentation and labeling. *Comput. Graph. Forum* 31 (7), 2241–2248.
- Nemhauser, G., Wolsey, L., 1978. Best algorithms for approximating the maximum of a submodular set function. *Math. Oper. Res.* 3 (3), 177–188.
- Pokrass, J., Bronstein, A.M., Bronstein, M.M., 2013. Partial shape matching without point-wise correspondence. *Numer. Math.* 6 (1), 223–244.
- Rudin, L., Osher, S., Fatemi, E., 1992. Nonlinear total variation based noise removal algorithms. *Phys. D: Nonlinear Phenom.* 60 (1), 259–268.
- Sapiro, G., 2006. Geometric Partial Differential Equations and Image Analysis. Cambridge University Press.
- Scherzer, O., Grasmair, M., Grossauer, H., Haltmeier, M., Lenzen, F., 2008. Variational Methods in Imaging. Springer.
- Shapira, L., Shalom, S., Shamir, A., Cohen-Or, D., Zhang, H., 2010. Contextual part analogies in 3D objects. *Int. J. Comput. Vis.* 89 (2–3), 309–326.
- Sidi, O., Kaick, V., Kleiman, Y., Zhang, H., Cohen-Or, D., 2011. Unsupervised co-segmentation of a set of shapes via descriptor-space spectral clustering. *ACM Trans. Graph.* 30 (6), 126.
- Sipiran, I., Bustos, B., 2012. Key-component detection on 3d meshes using local features. In: Proceedings of the 5th Eurographics Conference on 3D Object Retrieval, pp. 25–32.
- Song, R., Liu, Y., Martin, R.R., Rosin, P.L., 2014. Mesh saliency via spectral processing. *ACM Trans. Graph.* 33 (1), 1–17.
- Sun, J., Ovsjanikov, M., Guibas, L., 2009. A concise and provably informative multi-scale signature based on heat diffusion. In: Proceedings of the Symposium on Geometry Processing, pp. 1383–1392.
- Torralba, A., Murphy, K.P., Freeman, W.T., 2007. Sharing visual features for multiclass and multiview object detection. *IEEE Trans. Pattern Anal. Mach. Intell.* 29, 854–869.
- Wang, S., Hou, T., Li, S., Su, Z., Qin, H., 2013. Anisotropic elliptic pdes for feature classification. *IEEE Trans. Vis. Comput. Graph.* 19 (10), 1606–1618.
- Wang, S., Li, N., Li, S., Luo, Z., Su, Z., Qin, H., 2015. Multi-scale mesh saliency based on low-rank and sparse analysis in feature space. *Comput. Aided Geom. Des.* 35 (C), 206–214.
- Wang, Y., Asafi, S., van Kaick, O., Zhang, H., Cohen-Or, D., Chen, B., 2012. Active co-analysis of a set of shapes. *ACM Trans. Graph.* 31 (6), 165:1–165:10.
- Xie, Z., Xu, K., Liu, L., Xiong, Y., 2014. 3d shape segmentation and labeling via extreme learning machine. *Comput. Graph. Forum* 33 (5), 85–95.
- Yi, L., Kim, V.G., Ceylan, D., Shen, I.-C., Yan, M., Su, H., Lu, C., Huang, Q., Sheffer, A., Guibas, L., 2016. A scalable active framework for region annotation in 3d shape collections. *ACM Trans. Graph.* 35 (6), 210:1–210:12.