A Deep Learning Driven Active Framework for Segmentation of Large 3D Shape Collections

D. George[a], X. Xie[a], Y.-K. Lai[b], G.K.L. Tam[a]

[a]*Department of Computer Science, Swansea University, United Kingdom*
[b] *School of Computer Science and Informatics, Cardiff University, United Kingdom*

Abstract

High-level shape understanding and technique evaluation on large repositories of 3D shapes often benefit from additional information known about the shapes. One example of such information is the semantic segmentation of a shape into functional or meaningful parts. Generating accurate segmentations with meaningful segment boundaries is, however, a costly process, typically requiring large amounts of user time to achieve high-quality results. In this paper we propose an active learning framework for large dataset segmentation, which iteratively provides the user with new predictions by training new models based on already segmented shapes. Our proposed pipeline consists of three components. First, we propose a fast and accurate feature-based deep learning model to provide dataset-wide segmentation predictions. Second, we develop an information theory measure to estimate the prediction quality and for ordering subsequent fast and meaningful shape selection. Our experiments show that such suggestive ordering helps to reduce users' time and effort, produce high-quality predictions, and construct a model that generalizes well. Lastly, we provide interactive segmentation refinement tools, helping the user quickly correct any prediction errors. We show that our framework is more accurate and in general more efficient than the state-of-the-art for large dataset segmentation, while also providing consistent segment boundaries.

Keywords: shape segmentation, active learning, shape collections, user interaction

1. Introduction

Segmented datasets have already been shown incredibly useful for many applications, including shape matching [1], retrieval [2] and modeling [3]. Semantic labels are also useful for shape understanding and abstraction [4], and shape parsing and partial shape recovery [5]. Shape segmentation techniques often benefit the most from such fully labeled datasets. Supervised techniques require ground truth labels to train segmentation classifiers [6], and both supervised and unsupervised techniques need ground truth labels to evaluate their methods [7]. While existing works have shown good efforts and results [8, 9, 10], clear ground truth inconsistencies still exist [11]. This means both existing and new techniques could perform better with higher quality ground truth segmentations.

Generating high-quality segmentations for shape datasets is a time-consuming and interaction-heavy task. Smaller datasets, with only small numbers of inconsistencies or errors may be manageable through manual effort [12, 7]. Massive datasets would take a great amount of user effort however [13]. Further, these massive datasets typically consist of non-manifold (multiple components, holes, zero thickness, etc.) and low-resolution shapes. These shapes are very difficult to process in segmentation pipelines. Recent works employ point cloud projection [14, 9], or further KD-connected point cloud projection [15]. While these are viable techniques, there may be information loss when using point clouds, e.g., connectivity and topology of the shape. Without these, certain reliable features are much harder to compute or are inaccurate when computed (e.g., Shape Diameter Function (SDF) [16], Geodesic Distance). Although connectivity can be re-established (e.g., through K Nearest Neighbors, assuming the resolution of the point cloud is high enough), thin regions of the shape could be wrongly connected, leading to undesirable connections. More recently, there are increasing interests to use mesh-based representations to develop robust CNN techniques [17, 18]. For this reason, in our proposed pipeline, we largely focus on input meshes. We further show that by re-meshing these non-manifold 3D models into manifold meshes, our pipeline can handle very large datasets very well.

Previous works that generate ground truth segmentations for large datasets typically focus on active learning approaches, where a user has some control over the system and influences the decisions in some way. [19] first used an unsupervised co-segmentation algorithm, where the user interactively selects pairs of parts between shapes to connect or disconnect. [14] used a supervised algorithm to label a single part at a time. Users are asked to paint two 2D views of a 3D shape. A learning model is trained based on the painted regions and similar shapes (according to global shape descriptors) are evaluated on that model. However, these techniques can only provide a coarse segmentation and output segmentations may have errors. Further, [14] requires one part to be labeled at a time, so datasets with high numbers of parts will take longer and more iterations to label. Here, we developed an active framework which allows full shape segmentation of a shape dataset, to ensure good segmentation quality and it scales well with the number of parts in the dataset.

One of the challenges when developing an active framework for segmentation is minimizing user interactions while maximizing segmentation quality. To balance the quality and speed, we utilize a deep learning model for segmentation predictions. In general, deep learning models can take a long time to train, and typically require a large amount of training data. To resolve these, we propose to use a small Convolutional Neural Network (CNN), using two 2D histogram features as input. The features have been shown useful in previous work [6, 8] and fit the CNN paradigm as 2D histograms are like images. Our architecture allows for quick model training and we also adopt an ensemble based learning scheme [20] to help generalize with reduced available training data. In our experiments we compare to other feature-based CNN techniques. We show that our model can perform better than existing fast techniques, with results comparable to the state-of-the-art.

Another difficulty of an active learning framework is the exploration and analysis of model predicted results. It often takes a long time for users to choose the next 3D model to segment, and there are no ground truth data to compare the predictions for ranking. We thus use *entropy*, a measure of uncertainty, to define a ranking measure without needing ground truth segmentations. This ranking measure provides a meaningful ordering of the predicted segment labels in an interactive tabular view. This allows users to see which shapes the deep learning model segmented well or struggled with. Our experiments show that by selecting poorly segmented 3D models with respect to the ranking measure, it reduces both time and interactions required to segment the whole dataset.

Lastly, another problem we observed in existing active frameworks (e.g., [14]) is that they do not allow quick boundary refinement. When there are slight errors in the output segmentation, users will likely discard the results, leading to extra manual effort and longer interaction time. With this observation, we propose an interactive segmentation refinement algorithm that takes the current segmentation and information about the shape (e.g., angle and thickness) to refine the segmentation boundaries. This algorithm can quickly provide high-quality segmentations while greatly reducing interactions and time required to refine a shape.

Our proposed framework has been demonstrated to work well on public datasets (including PSB, COSEG), and also on re-meshed datasets from ShapeNet, which contains thousands of shapes.

Contributions. To summarize, the main contribution of this work is to develop the first deep learning driven active framework for segmentation of large 3D shape collections. The focus is to maintain accurate and meaningful segment boundaries, while keeping human effort and time to a minimum. Our active learning framework consists of several key components:

- First, we show and evaluate a novel deep learning pipeline for shape segmentation which is relatively fast and accurate, and is suitable for active learning purpose.
- Second, we use an information-theoretical metric for ordering the prediction of shape segmentation when ground truth data is not available. The metric is designed for our segmentation tasks. Users can still flexibly choose next shape to annotate through our interface. Our extensive experiments show that the ordering can help reduce total segmentation efforts and time.
- Third, we develop a useful technique for interactive segmentation refinement, which takes into account the segmentation boundaries and thickness of shapes. Our experiments show that it can help users to quickly improve segmentation boundaries, reducing effort and time.

We will also release the source codes of our tools for the community, and provide new and more accurate ground truth segmentation for some existing datasets[1].

In the following, Section 2 discusses the existing work for segmentation, feature extraction and entropy in geometry processing. In Section 3, we briefly overview our active learning framework. Section 4 discusses the details of the three novel subsystems. We further discuss our framework interface and flow in Section 5 before outlining our experiments and showing their results in Section 6. Finally, in Section 7 we conclude and discuss possible future work.

2. Related Work

This work relates to several research areas. We summarize the literature with respect to shape features, shape segmentation, active learning in image analysis, active learning in shape analysis, and use of entropy in graphics processing.

Shape Features and Their Uses. Much of the existing work in shape segmentation is driven by features. These can be defined per face, per vertex, per patch (a cluster of faces), or even per shape. These features are designed for different purposes, and many have been successfully applied in mesh segmentation. Per-face features include, SDF [16] which estimates the thickness of a shape at a given face, Conformal Factor (CF) [21] which computes a position invariant representation of the curvature of non-rigid shapes and Spin Images (SI) which capture the surface information around a face using a 2D histogram. Recent work has also adapted image based features to the 3D domain. One notable example is Shape Context (SC) [22] which is a 3D shape descriptor to encode both curvature and geodesic distance distributions in a 2D histogram [6]. However, there are limitations to how useful a feature can be on certain shapes. Examples are that CF is susceptible on shapes with sharp curvature [11], SDF can fail if the shape has holes and geodesic distance will fail if the shape has multiple components. Therefore feature selection for a new technique is very important, as it can greatly impact the accuracy and speed. For these reasons, we opted to use two features for this work, SC and SI. Recent work has shown both features can be very useful in shape segmentation [6, 23, 11]. Further they are both 2D histograms, so can be generated at any scale (number of bins) and CNNs should work well to extract useful information.

[1] https://cs.swan.ac.uk/~csgarykl/ActiveLabeller/ActiveMeshLabeller.html

Figure 1: Proposed Pipeline. After de ning the input data (shapes, features and possible segment labels, Sections 4.1, 4.2), users pick some models (Section Shape Subset Selection) and use the proposed interface and tools (Section 4.4, Patch Labeling, Painting, and Section 4.5 Interactive Boundary Re nement (IBR) annotate ground truth labels. With these ground truths, a fast deep learning model is trained. Graph-cut is applied to re ne the predicted labels (Section 4.6 Traini and Evaluation). The results are then ordered in an interface (Section 4.7, Order and Select Subset) for users to con rm the ground truth or further select a sub for user-driven inspection and IBR re nement (same interface in Sections 4.4, 4.5). The iterative active segmentation repeats until dataset is fully labeled.

Unsupervised and Supervised Shape Segmentation. The goal of a shape segmentation algorithm is to partition a single shape into meaningful parts [24, 16]. These algorithms typically used a feature which drives the partitioning (see Features section), though other work also used di erent strategies like tting of primitive shapes [25]. Recently, unsupervised techniques looked into co-analysis of a set of shapes, using information consistent across the set to improve the nal segmentation [7, 26, 27, 28, 29]. However, these methods struggle with largely varying datasets, especially those with a low number of shapes per set [11]. Further, the segmentation of parts not only relates to the shape geometry, but also the meaning, functionality and designs. All these challenges have led to the recent interests in supervised segmentation techniques.

Supervised segmentation techniques rely on prior knowledge in order to train a model. Typically these methods use large pools of shape features as input and classify them according to segment labels [6]. Subsequent techniques further improve in di erent ways, such as ranking features to nd segment boundaries [30], and training an extreme learning machine [31, 23] to classify the labels. However, similar to unsupervised work, these techniques can struggle when datasets are very diverse. To combat this, work using CNNs was proposed [8]. This work arranges a pool of features as an image, and uses an image-based convolution network to predict face labels. However, the simple arrangement leads to unnecessary interference of relationships between features with no correlation, and [11] reduces such interference using 1D convolutions, leading to better results. Recently, several techniques have shown new and interesting shape segmentation methods such as point cloud segmentation [32, 9], kd-tree point cloud segmentation [15], projecting image segmentations to shapes [33], hierarchical segmentations [34] and graph CNNs [10].

With the recent surge of new segmentation papers, each focusing on larger datasets, there is a need for high-quality ground truth labels. However, currently available ground truths for widely used segmentation datasets have been shown to contain inconsistent and poor labels for certain shapes within the dataset [11]. This can impact the training performance by introducing inconsistent labels for similar samples. It can also impact evaluation, as inconsistencies incorrectly degrade the performance of a model. Due to this, we emphasize providing accurate, high-quality segmentations in this work.

There is a concurrent work [35] that shares similar spirit as ours which produces high quality part annotations. They employed professionals to carry out the annotation. The ne-grained part dataset further inspires more recent interests in hierarchical shape segmentation [36], and grouping and labeling of semantic parts [37]. Compared to [35], our work proposes an active learning framework that allows fast annotation of large datasets with the help of a machine learning model. The framework can be used for other work to complete annotations of large datasets. To our knowledge, it is the rst deep learning driven active learning framework for segmentation of large 3D shape collections that aims at ground truth quality.

Active Image Analysis Active learning image analysis systems have been widely explored to leverage the human user input to explore large datasets. They focus on using user input to aid the classi ers by annotation (painting, strokes) or drawing bounding boxes. This has the advantage that, the user can see what data the classi er is struggling with and incrementally provides new training data to alleviate this problem, making the classi er more generalized and accurate [38, 39, 40, 41]. We utilize this functionality in 3D segmentation by allowing the user to incrementally tune the output labels of our model to make it generalize better, while also incorporating a sorting