Academic Journal of Science and Technology
ISSN: 2771-3032 | Vol. 20, No. 3, 2026

EARSeg3D: An Automated Segmentation Method for Ear Canal Impressions

Ye Cheng[1, *], Sheng Chen[2]

[1] School of Shanghai for Science and Technology University, Shanghai 200000, China
[2] School of Shanghai for Science and Technology University, Shanghai 200000, China
[*] **Corresponding author**

Abstract: Hearing impairment is a significant global public health issue. Hearing aids, as critical devices for improving the auditory experience of individuals with hearing loss, require personalized customization that decisively affects wearing comfort and acoustic performance. Traditional ear canal impression fabrication relies on manual procedures, which are time-consuming, cause patient discomfort, and introduce subjective errors. Although deep learning has achieved remarkable progress in medical image analysis, its application in the core steps of hearing aid customization—namely, automated semantic understanding and three-dimensional structural analysis of ear canal impressions—remains limited. To address this research gap, this paper proposes EARSeg3D, an end-to-end semantic segmentation method for ear canal impression point clouds based on PointNet++. Targeting the complex geometric characteristics of ear canal morphology, the method introduces a Local Context Fusion Module (LCFM) built upon PointNet++'s hierarchical feature learning framework. By encoding relative spatial positional relationships and feature gradients among points within local neighborhoods, LCFM enhances the network's ability to perceive the intricate geometric topology of the ear canal. Following a six-class semantic labeling standard, this study partitions ear canal impression point clouds into six semantic regions: ear canal tip, middle ear canal, ear canal base, redundant material, concha cavity, and concha cymba. Ablation study results on a self-constructed dataset show that with the LCFM module, the model achieves a mean Intersection over Union (mIoU) of 92.31% and an overall accuracy (OA) of 83.36%, which are 1.18 and 1.07 percentage points higher than the baseline model, respectively. Visualization results further demonstrate that the method effectively mitigates jagged segmentation boundaries, achieving smoother and more precise segmentation in curved regions of the ear canal and at the junctions between different anatomical structures.

Keywords: Hearing Aids, Deep Learning, Semantic Segmentation, Ear Canal Impression.

1. Introduction

Hearing impairment is a major global public health issue, affecting the daily communication and quality of life of hundreds of millions of people[1]. As critical devices for improving the auditory experience of individuals with hearing loss, hearing aids have evolved from early manual sound collectors (e.g., ear trumpets) to electric hearing aids, digital hearing aids, and intelligent hearing aids. Modern hearing aids not only amplify sound but also integrate digital signal processing, noise suppression, and speech recognition to adapt to complex auditory environments and deliver personalized hearing solutions[2].

In the customized design of hearing aids, individual variations in ear canal morphology significantly influence wearing comfort, acoustic performance, and real-world usability[3]. Traditional ear canal impression fabrication requires clinical personnel to manually create physical impressions using materials such as silicone. This process is time-consuming, uncomfortable for patients, and prone to subjective errors, as well as the complexity of subsequent digitization. With the advancement of three-dimensional scanning technology, ear canal impressions can now be digitally acquired, allowing precise recording of ear canal morphology in the form of 3D point clouds. Digital 3D impressions provide direct data input for computer-aided design (CAD) and computer-aided manufacturing (CAM), greatly improving the efficiency and accuracy of personalized hearing aid shell design. Some studies have explored using external 3D scanners to obtain ear canal models directly, replacing the digital workflow based on traditional physical impressions. The feasibility and efficiency advantages of this digital scanning approach in clinical fitting are gradually being validated[4].

Meanwhile, deep learning, an important branch of artificial intelligence, has demonstrated strong capabilities in medical image analysis in recent years. Convolutional neural networks (CNNs) have achieved remarkable success in classifying, segmenting, and assisting diagnosis of 2D medical images such as CT, MRI, and ultrasound data, enabling automated and high-precision lesion detection and tissue segmentation in complex structures[5, 6]. With the increasing availability of rich 3D data and computational resources, researchers have begun extending deep learning methods to 3D data analysis, with particular emphasis on point cloud processing. Due to the unordered and irregular nature of point clouds, traditional convolution operations based on pixels or grids cannot be directly applied. This challenge has motivated the development of point-cloud-specific neural network architectures such as PointNet[7] and its improved version, PointNet++[8]. By leveraging symmetric functions, local neighborhood aggregation, and hierarchical feature learning, these methods effectively handle the unordered structure of point clouds, enabling direct extraction of semantic information from raw 3D point sets.

Despite numerous breakthroughs in medical imaging, deep learning remains relatively limited in its application to the hearing aid manufacturing process[9]. Existing research has largely focused on optimizing acoustic performance, such as noise suppression and speech enhancement, to improve

device performance in complex auditory environments[10]. However, for the core step of hearing aid customization—automated semantic understanding and precise 3D structural analysis of ear canal impressions—mature automated solutions are still lacking. Traditional point cloud post-processing methods often rely on hand-crafted features or rule-based approaches, which struggle to capture the intricate geometric details of diverse ear canal morphologies and exhibit poor robustness to noise and irregular sampling[11].

Given this status quo, this paper proposes EARSeg3D, an end-to-end ear canal impression point cloud segmentation method based on PointNet++. Through hierarchical feature learning, the model directly extracts local and global geometric features from raw 3D point clouds to achieve multi-category semantic segmentation of ear canal impressions. This method aims to fill the gap in deep learning applications within the hearing aid production process, improve the efficiency of personalized design, and provide feasible technical support for the digital manufacturing of hearing aids

2. Anatomical Features of the Ear Canal and the Semantic Segmentation Standard

(1) Dataset and Preprocessing

From an anatomical perspective, the auricle has an extremely complex structure that exhibits significant individual variation. Its main components include the helix, antihelix, concha, tragus, and lobule. The external auditory canal is of particular interest. It is a canal approximately 2.5–3.5 cm in length and 0.7–1.0 cm in diameter. Rather than being straight, it consists of two segments—a bony segment and a cartilaginous segment—oriented in different directions[12].

The external auditory canal itself presents an "S"-shaped curvature, which is primarily determined by the anatomy of the auricle. When viewed from the auricular entrance (the concha), the anterior wall of the external auditory canal inclines slightly downward and forward (though "anterior" already implies forward), while the posterior wall is relatively vertical. In the deep portion of the external auditory canal, there are two major physiological curvatures:

First Bend: Approximately 0.5–1.0 cm inward from the canal entrance, the canal begins to curve inward and downward.

Second Bend: Approximately 1.5–2.0 cm inward from the canal entrance, the canal curves again, primarily inward and upward. This bend is critically important for the fitting position of many deep-canal hearing aids (e.g., CIC, IIC), as it determines the deepest point of the hearing aid shell.

In the auricular region, the structures closely related to the entrance of the external auditory canal include:

Cavum Conchae: This is the largest concave area at the entrance of the external auditory canal. For in-the-ear (ITE) hearing aids, the morphology of the cavum conchae directly determines the size and available space for the hearing aid shell.

Cymba Conchae: A relatively shallow area located above the cavum conchae and below the crus of the helix. Part of the ITE hearing aid shell may also extend into this region.

(2) Semantic Segmentation Standard for Ear Canal Impressions

To enable automated generation of customized hearing aid shells, accurate semantic segmentation of the anatomical structures of the ear canal is essential. This study adopts the six-category semantic labeling standard for ear canal impressions (ECIs) proposed by Zouhar et al[13]., which partitions ear canal impressions into regions of clear clinical and engineering relevance. This standardized definition addresses the need, during hearing aid customization, to flexibly extract the corresponding morphological components from ECIs based on the requirements of different hearing aid types (e.g., ITE, ITC, CIC, IIC; see Figure 1). The ECI point cloud is divided into the following six semantic categories:

Figure 1. Hearing Aid Type Classification

Category 0: Ear Canal Tip – Typically located distal to the second bend of the ear canal, this is the target region that IIC (invisible-in-canal) hearing aids need to reach.

Category 1: Middle Ear Canal – Located between the first bend and the second bend, this is a critical region for ITE (in-the-ear), ITC (in-the-canal), and CIC (completely-in-canal) hearing aid shells.

Category 2: Ear Canal Base – Located between the ear canal entrance and the first bend.

Category 3: Redundant Material – Refers to non-canal anatomical structures, such as the handling base retained during scanning and non-essential cartilage areas around the ear. This material has no direct relevance to hearing aid manufacturing and needs to be removed.

Category 4: Cavum Conchae – The bowl-shaped depression deep in the auricle, which is the main embedding area for ITE and ITC hearing aids.

Category 5: Cymba Conchae – The shallow fossa area in the lower part of the auricle, between the cavum conchae and the helix. For ITE hearing aids, complete preservation of the cymba conchae region is an important component of the shell.

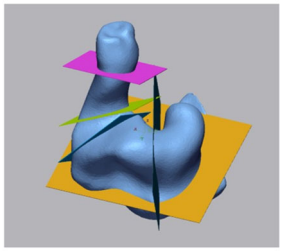

Figure 2. Segmentation Standard for Ear Canal Impressions

Figure 2 visually illustrates the anatomical specification of

the above six semantic categories on an ECI point cloud. This classification system achieves functional decoupling of irregular ear canal morphologies and can flexibly accommodate the manufacturing requirements of mainstream hearing aid types—CIC, ITC, IIC, and ITE—by combining different semantic regions. For example, ITE hearing aids require preservation of both the cymba conchae and the cavum conchae; CIC hearing aids need only be truncated at the first bend; and IIC hearing aids require preservation of the region distal to the second bend.

3. Method

The primary task of ear canal impression (ECI) segmentation is to remove redundant regions irrelevant to hearing aid manufacturing (e.g., scanning bases) and to achieve accurate semantic segmentation of valid anatomical components (e.g., cavum conchae, cymba conchae, ear canal, etc.). To accomplish this, this paper employs PointNet++, a powerful feature extractor, as the backbone network. The segmentation network first takes mesh vertices and vertex normals as input, then performs sampling, grouping, and feature extraction through Set Abstraction (SA) layers. It encodes the internal structural relationships among neighboring points and the interactions across global feature channels, and introduces a Local Context Fusion Module (LCFM) to enhance the network's perception of local geometric structures. This module is embedded into the hierarchical architecture of PointNet++, improving the model's segmentation accuracy for complex ECI morphologies.

PointNet++ is a deep learning model for point set processing that enhances the capture of local and global point cloud features through a hierarchical design. Its core idea is to build upon the original PointNet by introducing a multi-scale, hierarchical feature extraction framework, which enables deep semantic understanding via the following key modules:

Set Abstraction (SA) Layer: Performs sampling and grouping operations on the point cloud layer by layer, partitions the original point cloud into local neighborhoods, and applies PointNet for local feature learning. This layer incorporates local geometric information, allowing the model to understand fine-grained shapes.

Feature Propagation (FP) Layer: In segmentation tasks, high-level abstract features must be propagated back to each point to provide rich semantic descriptions. The FP layer propagates features via interpolation and skip connections.

Multi-scale Fusion Mechanism: PointNet++ extracts and fuses features from neighborhoods of different scales, improving robustness and detail representation ability under conditions of non-uniform point cloud density.

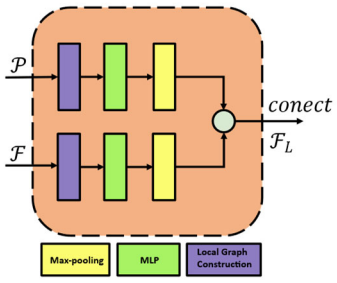

Figure 3. LCFM Architecture

Unlike the original PointNet, which performs only global feature learning, PointNet++ captures both local details and the overall structure of point clouds, making it more effective for complex 3D morphology segmentation. Despite its strong performance in local feature extraction, PointNet++ has limitations when processing ear canal point clouds. Specifically, the original SA layer relies primarily on Euclidean distance for grouping, which is insufficiently sensitive to the complex geometric topology of the ear canal and can lead to incorrect clustering in that region. The LCFM module introduced in this paper deeply encodes the internal structural relationships among neighboring points to compensate for the shortcomings of Euclidean distance and enhance the network's perceptual accuracy for local complex geometric topology.

Given that the ear canal is smooth and lacks sharp geometric variations, this study embeds the LCFM between consecutive SA layers, as illustrated in Figure 3. The goal is to overcome PointNet++'s limitation of relying solely on absolute spatial distance for feature extraction, and instead explicitly encode the relative topological relationships and feature gradients among points within local neighborhoods. The specific implementation process is as follows: Given a point cloud feature from an SA layer, denoted as $X \in \mathbb{R}^{B\times 3\times N}$ and $F \in \mathbb{R}^{B\times C\times N}$, which represent the spatial coordinates and input feature dimensions of the point cloud, respectively, where B is the batch size and N is the number of points. First, to capture the local geometric context for each point P_i, the K-Nearest Neighbors (KNN) algorithm is used to find its K nearest neighbors in Euclidean space, forming a local neighborhood. By computing the relative coordinate differences ΔX and the relative feature differences ΔY between each neighbor point P_j and the center point P_{i} the network is provided with explicit local spatial structure information. Subsequently, two shared Multi-Layer Perceptrons (MLPs) are used to encode these two types of differences separately:

$$G_{xyz} = \mathrm{Re}\,LU\left(BN\left(Conv2d\left(\Delta X\right)\right)\right)$$

Where $G_{xyz} \in \mathbb{R}^{B\times C/2\times N\times K}$ encodes the geometric structure of the local space, and $G_{feat} \in \mathbb{R}^{B\times C/2\times N\times K}$ encodes the variation of local features. After concatenating G_{xyz} and G_{feat} ,a max pooling operation is applied to aggregate information from the K neighboring points, yielding an enhanced local feature $F_{local} \in \mathbb{R}^{B\times C\times N}$:

$$F_{local} = \max_{k=i,\ldots,K}\left(Concat\left(G_{xyz}, G_{feat}\right)\right)$$

This fusion operation enables the network to simultaneously perceive the geometric features of the neighborhood and their precise spatial distribution, which helps the network distinguish boundaries of complex structures such as curvatures and folds within the ear canal.

4. Experiment

(1) To strengthen the combination of traditional hand embroidery skills and modern industrialization

The dataset used in this paper is the public dataset provided by Zouhar et al. The sample data were obtained from patients with sensorineural hearing loss. All samples are stored in STL format, with each model containing 20,000 to 60,000 points. The number of points varies depending on individual differences in ear canal morphology. Due to occlusion by the fixed base of the ear mold and the limited viewing angle of structured light imaging, redundant silicone material is present in the base connection area.

In the semantic segmentation experiment, the training batch size was set to 16, the learning rate to 0.0001, and the number of training epochs to 200. The proposed method was implemented on a single NVIDIA RTX 3090 GPU using the Windows operating system, with Python 3.9 and PyTorch 1.15.

To facilitate comparison with other methods, this study adopted the two most commonly used metrics for evaluating point cloud semantic segmentation performance—mean Intersection over Union (mIoU) and overall accuracy (OA).

(2) Quantitative and Visualization Analysis

To evaluate the effect of the proposed LCFM module on ECI point cloud semantic segmentation, this section presents ablation experiments comparing the baseline model (PointNet++) with the model incorporating the LCFM module. Under identical training and testing settings, the experimental results are shown in Table 1. After introducing LCFM, the mIoU increases from 91.13% to 92.31%, and the OA increases from 82.29% to 83.36%.

Table 1. Ablation Study Results

Model	mIoU%	OA%
Baseline	91.13	82.29
Baseline+LCFM	92.31	83.36

In the task of ECI point cloud segmentation without manual intervention, achieving high-quality segmentation is critical for subsequent hearing aid shell manufacturing. Figure 4 presents visualization results for selected samples after incorporating the LCFM module, along with boundary heatmaps generated by the model. It is evident that the baseline PointNet++ produces noticeable jagged artifacts at segmentation boundaries. These jagged boundaries not only degrade visual quality but also adversely affect subsequent shell manufacturing, reducing the efficiency and accuracy of the entire system. With the LCFM module, the segmentation results become more stable in terms of both overall structural consistency and local boundary representation, aligning more closely with the ground truth labels. Particularly in curved regions of the ear canal and at the junctions of different anatomical structures, the predictions are closer to the ground truth, and the segmentation boundaries are relatively smooth. Such high-quality segmentation is essential for ensuring precise fitting of the hearing aid shell to the human ear, and can effectively enhance the efficiency and accuracy of the entire manufacturing process.

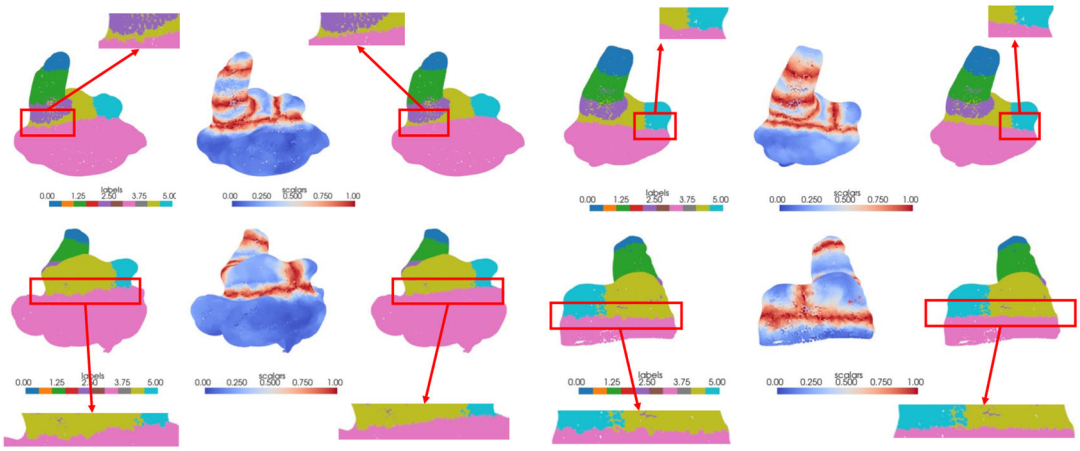

Figure 4: Visualization of Segmentation Results

5. Conclusion

The PointNet++-based semantic segmentation method for ear canal impression point clouds proposed in this paper achieves automatic six-category classification of complex 3D structures, thereby providing stable 3D foundational data support for personalized hearing aid design. The model demonstrates strong performance across multiple evaluation metrics, confirming the technical feasibility of point cloud deep learning methods in medical device manufacturing. Future work combining cross-device adaptation and multimodal fusion is expected to further enhance the clinical and industrial applicability of this technology.

References

[1] HAILE L M, ORJI A U, REAVIS K M, et al. Hearing loss prevalence, years lived with disability, and hearing aid use in the United States from 1990 to 2019: findings from the Global Burden of Disease Study [J]. Ear and hearing, 2024, 45(1): 257-67.

[2] ORGANIZATION W H. World report on hearing [M]. World Health Organization, 2021.

[3] LAUNER S, ZAKIS J A, MOORE B C. Hearing aid signal processing [J]. Hearing aids, 2016: 93-130.

[4] PANAYI N C, EFSTATHIOU S, CHRISTOPOULOU I, et al. Digital orthodontics: Present and future [J]. AJO-DO Clinical Companion, 2024, 4(1): 14-25.

[5] CHUA K W D, YEO H K H, TAN C K L, et al. A novel ear impression-taking method using structured light imaging and machine learning: A pilot proof of concept study with patients' feedback on prototype [J]. Journal of Clinical Medicine, 2024, 13(5): 1214.

[6] DORAN S. 3D Printing Based Silicone Earmold Mold Fabrication for Pediatric Hearing Aids [D]; WORCESTER POLYTECHNIC INSTITUTE, 2025.

[7] QI C R, SU H, MO K, et al. Pointnet: Deep learning on point sets for 3d classification and segmentation; proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, F, 2017 [C].

[8] QI C R, YI L, SU H, et al. Pointnet++: Deep hierarchical feature learning on point sets in a metric space [J]. Advances in neural information processing systems, 2017, 30.

[9] TOGNOLA G, PARAZZINI M, SVELTO C, et al. Design of hearing aid shells by three dimensional laser scanning and mesh reconstruction [J]. Journal of biomedical optics, 2004, 9(4): 835-43.

[10] CORTEZ R, DINULESCU N, SKAFTE K, et al. Changing with the times: applying digital technology to hearing aid shell manufacturing [J]. Hearing Review, 2004, 11(3): 30-41.

[11] CHUNG K. Challenges and recent developments in hearing aids: Part II. Feedback and occlusion effect reduction strategies, laser shell manufacturing processes, and other signal processing technologies [J]. Trends in Amplification, 2004, 8(4): 125-64.

[12] ZOUHAR A, SLABAUGH G G, UNAL G, et al. Generalized rigid alignment of 3d ear impression models [Z]. Google Patents. 2011

[13] ZOUHAR A, ROTHER C, FUCHS S. Semantic 3-D Labeling of Ear Implants using a Global Parametric Transition Prior [J].