A Unified Deep-Learning-Based Framework for Cochlear Implant Electrode Array Localization

Yubo Fan[1]([envelope]), Jianing Wang[2], Yiyuan Zhao[3], Rui Li[4], Han Liu[1], Robert F. Labadie[5], Jack H. Noble[4], and Benoit M. Dawant[4]

[1] Department of Computer Science, Vanderbilt University, Nashville, TN 37235, USA
yubo.fan@vanderbilt.edu
[2] Digital Technology and Innovation, Siemens Healthineers, Princeton, NJ 08540, USA
[3] Digital and Automation, Siemens Healthineers, Malvern, PA 19355, USA
[4] Department of Electrical and Computer Engineering, Vanderbilt University, Nashville, TN 37235, USA
[5] Department of Otolaryngology - Head and Neck Surgery, Medical University of South Carolina, Charleston, SC 29425, USA

Abstract. Cochlear implants (CIs) are neuroprosthetics that can provide a sense of sound to people with severe-to-profound hearing loss. A CI contains an electrode array (EA) that is threaded into the cochlea during surgery. Recent studies have shown that hearing outcomes are correlated with EA placement. An image-guided cochlear implant programming technique is based on this correlation and utilizes the EA location with respect to the intracochlear anatomy to help audiologists adjust the CI settings to improve hearing. Automated methods to localize EA in postoperative CT images are of great interest for large-scale studies and for translation into the clinical workflow. In this work, we propose a unified deep-learning-based framework for automated EA localization. It consists of a multi-task network and a series of postprocessing algorithms to localize various types of EAs. The evaluation on a dataset with 27 cadaveric samples shows that its localization error is slightly smaller than the state-of-the-art method. Another evaluation on a large-scale clinical dataset containing 561 cases across two institutions demonstrates a significant improvement in robustness compared to the state-of-the-art method. This suggests that this technique could be integrated into the clinical workflow and provide audiologists with information that facilitates the programming of the implant leading to improved patient care.

Keywords: Cochlear Implant · Electrode Localization · Object Detection

1 Introduction

A cochlear implant (CI) has an electrode array (EA) that is surgically inserted into the cochlea to stimulate the auditory nerves and treat patients with severe-to-profound sensorineural hearing loss. Although CIs have achieved great success, hearing outcomes among recipients vary significantly [1, 2]. Recent studies have shown that factors related to electrode positioning impact on audiological outcomes [3, 4]. These studies require knowing the precise electrode locations, which can be obtained by postoperative CT imaging. Another clinical application that requires precise electrode locations is image-guided cochlear implant programming [5]. After surgical implantation, the CI needs to be programmed such that an optimized frequency mapping [24] can be determined for each patient. Knowledge about the spatial relationship between the electrodes and the intracochlear anatomy permits to generate programming solutions which have been shown to significantly improve the hearing outcomes for both adult and pediatric CI recipients [6, 7].

To facilitate the EA localization process, automated methods have been proposed by several groups. In [8] EAs are categorized as closely-spaced and distantly-spaced. EAs with an interelectrode spacing such that the intensity contrast between electrodes cannot be distinguished are considered closely-spaced, while the opposite applies to distantly-spaced EAs. Note that a given EA model could be categorized as closely-spaced or distantly-spaced, depending on the image resolution. For example, some closely-spaced EAs included in [8] could be considered distantly-spaced in some images acquired at very high resolution (0.08mm isotropic) as in [9]. For algorithms localizing closely-spaced EAs, extracting the centerline of an EA is often an important step [8, 10, 11]. It is achieved by either using intensity-based features only [11] or combining both intensity-based and EA shape-based features [8]. For localizing distantly-spaced EAs, hand-crafted feature extractors are utilized to detect individual blobs [9, 12–14]. To link the electrode candidates in the correct order and remove false positive candidates, graph-based path-finding algorithms [12] or Markov random field models [14] have been proposed. While there has been an emergence of methods based on deep learning (DL) recently [15–17], they cannot be viewed as a complete solution to the automatic EA localization problem because the networks are only trained and validated on one type of EA model [17] or cannot order/link the detected electrodes to form a complete array [15, 16].

In this work, we present a novel DL-based framework that consists of a multi-task network and a set of postprocessing algorithms to localize the cochlear implant electrode array. Our contribution is three-fold: (1) To the best of our knowledge, it is the first unified DL-based framework designed for localizing both distantly- and closely-spaced EAs in CT and cone-beam CT (CBCT) images. (2) We propose four (three detection and one segmentation) tasks for the multi-task network such that this single network can be trained on various kinds of EAs. (3) We extensively evaluate this framework on datasets that significantly exceed the scale of all datasets reported in the literature to date: a heterogeneous clinical _test_ set with CT or CBCT images of 561 implanted ears and a test set with gold standard ground truth for 27 implanted cadaveric ears. Results show that the proposed framework is significantly more robust (generates results that require manual adjustments less often) than the state-of-the-art (SOTA) techniques, while also being slightly more accurate. These findings indicate that the proposed framework could

be reliably used to support large-scale quantitative studies and deployed in the clinical workflow to provide clinicians with critical information at the time and point of care.
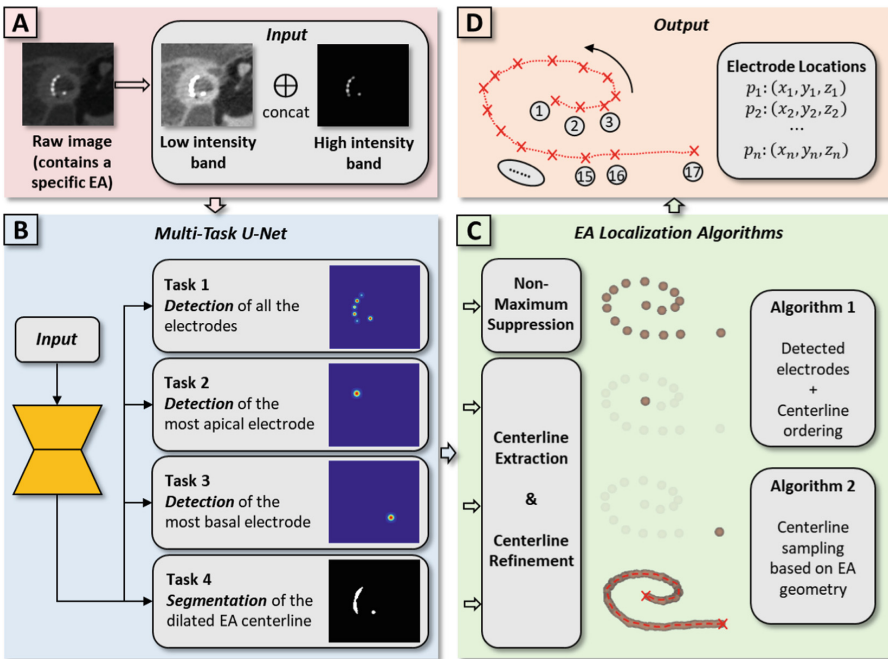

Fig. 1. An overview of the proposed framework. It consists of two major modules: a multi-task U-Net that is generic for various EA types (B) and a set of postprocessing algorithms that utilize the known EA geometry (C). The input to the multi-task U-Net is the two-channel image that contains low and high intensity bands extracted from the raw image (A). The output of the framework is the ordered electrode locations (D).

2   Method

2.1   Data

The data in this study consist of a large-scale clinical dataset from CI recipients (dataset #1) as well as 27 cadaveric samples (dataset #2). Dataset #1 includes 1324 implanted ears from CI recipients treated at two institutions (datasets #1A and #1B). 8 types of distantly- and closely-spaced EA from 3 manufacturers are included in these cases. Dataset #1A has 958 implanted ears of which 97% are scanned with CBCT scanners, and the remaining (3%) are scanned with conventional CT scanners. Dataset #1B includes 366 implanted ears of which most (98%) are scanned with conventional CT scanners and the remaining (2%) are scanned with CBCT scanners. Dataset #2 contains 4 types of distantly-spaced EAs, and these specimens are scanned with conventional CTs.

The training and validation sets are all from dataset #1A and constitute 60% and 20% of that dataset, respectively. The remaining 20% of dataset #1A along with dataset #1B (561 implanted ears in total) are used to test the robustness of the proposed framework. Dataset #2 is used to test its accuracy because the gold standard ground truth can be obtained using their paired micro-CT [16, 18]. Details on the datasets and EA specifications can be found in the supplementary material.

2.2   Multi-task U-Net

The proposed framework is designed to localize various types of EAs in both CT and CBCT. This is challenging because the number of electrodes to be detected and the interelectrode spacing is different among EA models, and the postoperative images can have different intensity characteristics depending on the type of scanner, i.e., CT or CBCT, that is used for their acquisition. To normalize the input image in a way that enhances both the nearby anatomy, which contains contextual information, and the high intensity (usually the electrodes) component, inspired by [16], we separate the raw image into two channels. One is the original image clipped at the 99.9% of its intensity histogram. The other is the original image in the [99.9%-100%] interval of its intensity histogram. Each channel is linearly normalized to [0,1] based on its own min-max values. We refer to these as the low and high intensity bands and an example can be seen in Fig. 1A.

To avoid having to train a network for each EA type, we define four tasks that a single multi-task network can learn simultaneously and whose output can be used to localize and order the electrodes in all arrays. Specifically, as shown in Fig. 1B, the four tasks include (1) detection of all the electrodes on the EA by heatmap regression, (2) detection of the most apical endpoint (tip) of the EA by heatmap regression, (3) detection of the most basal endpoint (the farthest electrode from the tip) of the EA by heatmap regression, and (4) segmentation of the EA centerline that starts and ends with the two endpoints. Although the network is a simple U-Net-like encoder-decoder architecture, the innovation is more focused on the multi-task strategy such that the network can serve as a robust feature extractor that leverages a large heterogeneous dataset.

We train the multi-task network using the electrode positions obtained with the methods described in [8, 12] corrected manually when a large localization error is visually observed. For tasks (#1,2, and 3) aiming at localizing electrodes, the ground truth is a one-channel heatmap for each task with a Gaussian kernel (variance of 2 voxels) at each electrode location. Following [19], we use a penalty-reduced voxel-wise logistic regression with the focal loss [20] as the training objective.

Depending on the image quality and the interelectrode spacing, EAs can appear as a whole bright (i.e., with high Hounsfield Unit (HU) values) tubular structure or distinguishable bright blobs representing individual electrodes. We define the centerline of the EA as a line connecting each electrode in sequential order (from most apical to most basal or the opposite). The motivation for segmenting the EA centerline is twofold. First, after we extract all the electrodes from the predicted heatmap (Task #1), we need to order them. We will show later in our postprocessing algorithms that these detected electrodes can be linked in the correct order using the segmented centerline and the detected two endpoints (Algorithm 1 in Fig. 1C). The second reason is that it

can serve as an alternative EA localization method by sampling the centerline using the known interelectrode spacing specific to a particular EA model (Algorithm 2 in Fig. 1C). It is essential when the electrodes are not discernible due to the low image resolution and/or the small interelectrode spacing. Different from [8], we resort to a DL approach rather than human-crafted feature extractors. To make the training of the centerline segmentation easier, we dilate the ground truth centerline to a tubular-structured mask with a radius of 3 voxels as the learning target. In addition to the Dice loss, we adopt a clDice loss which has shown its superiority in improving the accuracy and preserving the topology of the underlying one-voxel wide centerline [21].

2.3 Postprocessing Algorithms for EA Localization

As said above, although the output of the multi-task network contains essential information to identify the EA, it does not provide the desired final output, i.e., the position (coordinates) of the ordered electrodes in the image. To do so, we have designed a series of postprocessing algorithms that can effectively extract the ordered electrode locations from the four output maps. As shown in Fig. 1C, there are two main algorithms (Algorithm 1 and Algorithm 2) that are suitable for distantly- and closely-spaced EAs, respectively.

Algorithm 1 takes the heatmap of all the electrodes (Task #1) as input and utilizes a non-maximum suppression (NMS) algorithm, which is a common postprocessing step for object detection [22], to obtain the desired number of electrodes. Then, the centerline of the EA is extracted by skeletonizing its segmented mask. Its two endpoints are further refined by merging the detection results of the most apical and basal points. Finally, all the detected electrodes are linked with the guidance of the centerline along the direction from the most apical to the most basal. Algorithm 1 works well if there is an apparent contrast between the electrodes in the heatmap, which is the case for most distantly-spaced EAs.

However, for closely-spaced EAs, it is nearly impossible to differentiate the individual electrodes. Algorithm 2 is designed to localize EAs in such situations. After the extraction and refinement of the EA centerline, the centerline is smoothed with a cubic spline, and the final electrode positions are obtained by resampling along it using known interelectrode spacing for the EA [8]. Note that for distantly-spaced EAs, Algorithm 1 can occasionally lead to abnormal localization results, such as an incorrect number of detected electrodes or the spacing between the detected electrodes being inconsistent with the known interelectrode spacing for the EA. This is most often caused by poor image quality that affects the creation of the subsequent heatmap (Task #1). We have designed simple rules to detect these anomalies. When one is detected, the framework switches to Algorithm 2 for a more reliable sampling-based localization for distantly-spaced EAs.

3 Experiments and Results

3.1 Implementation Details

The multi-task U-Net is trained with PyTorch 1.12 on an NVIDIA RTX 2080 Ti GPU. We use MONAI [23] for data augmentation which contains an additive Gaussian noise and random affine transformations. The images are preprocessed by being rigidly registered

to the left ear (mirroring is performed if one case is a right ear) of a template volume (atlas), resampled to a 0.1mm isotropic voxel size, and cropped into a region of interest (ROI) with a dimension of $320 \times 320 \times 320$. Due to GPU memory size limit, we use a patch-based strategy (with a dimension of $256 \times 256 \times 192$) for training. The batch size is set to 1 and we use AdamW optimizer with learning rate of 5e-4. At inference we use a sliding-window approach to merge the results. The network is trained for 250 epochs, and we select the epoch with the lowest validation loss as our final model. Note that a minority (3%) of the images in dataset #1 are reconstructed with a limited HU range, i.e., [-1024,3071]HU. In these images bone and electrodes have similar intensity values. The remaining (97%) images have maximum intensities far larger than 3071HU. In these images electrodes have larger intensity values than bone. To address this issue, we train another DL model with the same training strategy and dataset, but the intensity of all the training images is saturated at 3071HU, i.e., every pixel with an intensity above 3071HU is assigned a value of 3071HU. All the results obtained for limited HU range images presented herein are obtained with this dedicated model. The postprocessing algorithms are implemented in Python 3.9 and NumPy. We use skimage for 3D skeletonization and the csaps package for calculating the cubic spline. The inference time for the proposed framework (from loading the image to outputting the electrode locations) ranges between 5 to 20 s.

3.2   Evaluation and Results

The techniques described in [8] (for closely-spaced EAs) and [12] (for distantly-spaced EAs) are designed and validated with EAs and imaging protocols that are similar to those used in this study. They are also in routine use with over 2000 ears processed at the two institutions in dataset #1. They are considered to be the SOTA methods for comparison.
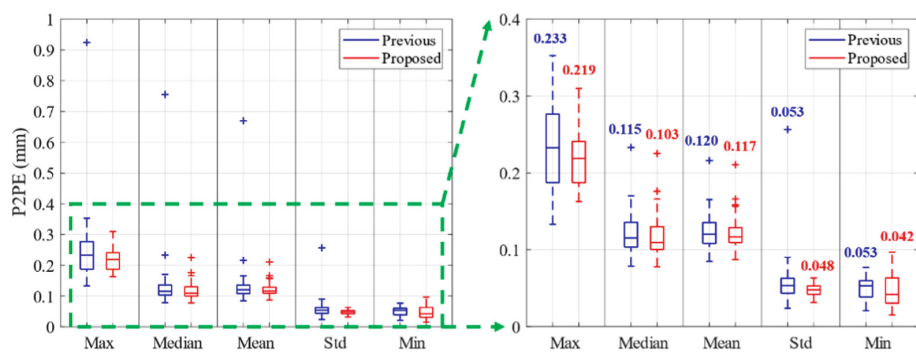

Fig. 2. Box plot of the P2PE in dataset #2 (27 implanted cadaveric ears). The median value for each metric is shown above each box in the right figure.

We first evaluate the accuracy of the proposed framework on dataset #2 (27 implanted cadaveric ears) for which the localization ground truth is known. Since the EAs in dataset #2 are all distantly-spaced, [12] is used as the previous SOTA method for comparison. We define the point-to-point error (P2PE) as the Euclidean distance between the predicted
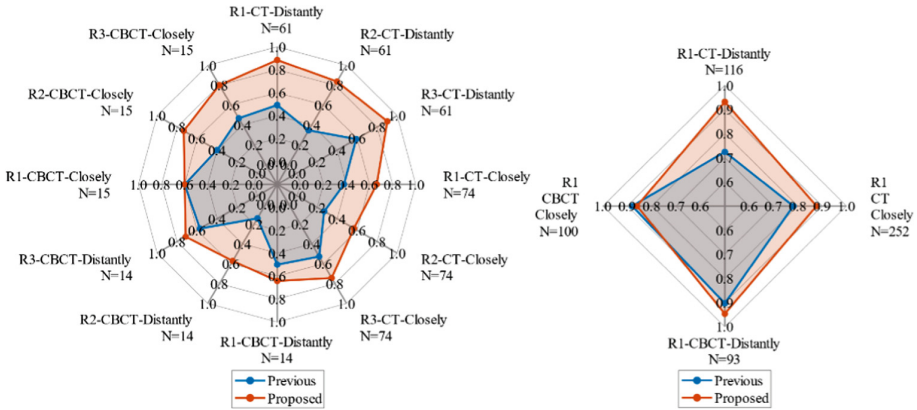
Fig. 3. Acceptance rate (higher means better) of the SOTA and proposed localization results on the large-error subset (left) and the whole clinical test set (right). R1, R2, and R3 are different raters.

electrode location and the ground truth location, and we calculate five P2PE-based metrics (maximum, median, mean, standard deviation (std), and the minimum P2PE) for all the electrodes in each case. The quantitative results are shown in Fig. 2. We can see from the left box plot that the results of the previous method [12] contain an outlier with relatively large P2PE, which is attributed to the low quality of the image. The median values in the right box plot show that the proposed framework has slightly smaller P2PE across the five metrics, but the differences are not statistically significant.

As introduced in Sect. 2.1, the 561 clinical cases in dataset #1 used for testing are highly heterogeneous. They contain CBCT and CT images from dataset #1A and #1B. Since manual annotations are not available for these images, we ask three experts to evaluate and compare the results obtained with the previous SOTA methods ([8] and [12]) and those obtained with the proposed method. Specifically, as shown in Fig. 4, we generate MIP (Maximum Intensity Projection) images and mark the locations at which the electrodes have been localized. This permits a rapid visual assessment of the localization quality. Next, for each case, we present the visualization of the results generated by the previous SOTA methods and by the proposed method side-by-side but in a random order, such that the experts are blind to the methods used to localize the contacts. They are then asked to rate the localization results as acceptable, i.e., no need to adjust the localization result or "Needs Adjustment (NA)", i.e., at least one electrode location needs to be adjusted by more than half a contact size. When both results are acceptable, raters are also asked to decide which one is preferred (i.e., has a more accurate localization result) or there is no preference.

Before performing the expert evaluation, we calculate the maximum P2PE between the results from the previous SOTA and the proposed methods. For cases with maximum P2PE larger than 0.3 mm (we refer to them as the large-error subset), there is a high probability that at least one of the methods generates NA results. For cases with maximum P2PE smaller than 0.3 mm (we refer to them as the small-error subset), we presume that the probability of NA results is relatively small. To limit the demand on the experts'
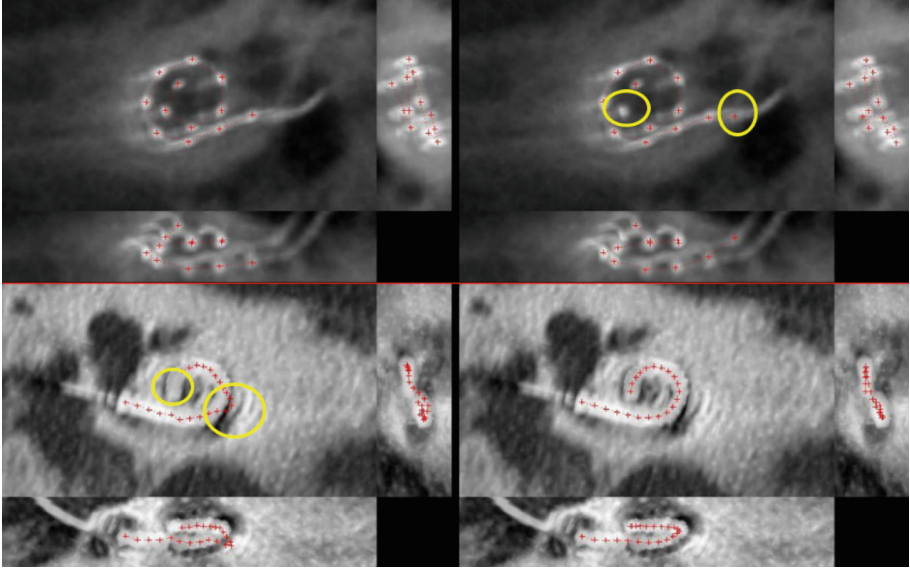
Fig. 4. Two representative cases from the large-error subset are shown with the display used to visually evaluate localization quality. Yellow circles highlight the regions that contain the NA electrode localization. Top row: a distantly-spaced EA. Bottom row: a closely-spaced EA. The Left and right images in each row represent results obtained with one of the methods; the method that is used for each is not disclosed to the evaluator.

time, we ask three experts (R1, R2, and R3) to rate the large error subset (164 cases) and only one expert (R1) to rate the small error subset (397 cases).

The expert evaluation results are shown as radar plots in Fig. 3. The left figure shows the acceptance rate for the large error cases evaluated by raters R1, R2, and R3. Except for the cases rated by R1 that contain closely-spaced EAs in CBCT ("R1-CBCT-Closely"), the localization results generated by the proposed method have a substantially higher acceptance rate than the previous SOTA methods ([8] for closely-spaced and [12] for distantly-spaced). For the cases in which both methods are acceptable, the average preference rates across the three raters are: 20.2% for "SOTA preferred", 30.3% for "Proposed preferred", and 49.5% for "No preference". As can be seen in the right plot of Fig. 3, the overall acceptance rate on the whole clinical test set evaluated by R1 is 80.7% for the previous SOTA methods and 89.7% for the proposed method. It is interesting to note that although the proposed method is trained mostly on CBCT images, it still generalizes well on CT images. Figure 4 shows two representative cases from the large-error subset.

4   Conclusions

In this work, we have proposed a novel DL-based framework for cochlear implant EA localization. To the best of our knowledge, it is the first unified DL-based framework designed for localizing both distantly- and closely-spaced EAs in CT and CBCT images.

Compared to the SOTA methods, the proposed framework is substantially more robust (9% less NA results) when evaluated on a large-scale clinical dataset and achieves slightly more accurate localization results on a dataset containing 27 cadaveric samples with gold standard ground truth. While it may be possible to improve our success rate further, a low percentage of NA results is unavoidable. We are thus developing quality assessment techniques to alert end users when images have poor quality and/or results are unreliable.

References

1. Sharma, S.D., Cushing, S.L., Papsin, B.C., Gordon, K.A.: Hearing and speech benefits of cochlear implantation in children: a review of the literature. Int. J. Pediatr. Otorhinolaryngol. **133**, 109984 (2020). https://doi.org/10.1016/j.ijporl.2020.109984
2. Boisvert, I., Reis, M., Au, A., Cowan, R., Dowell, R.C.: Cochlear implantation outcomes in adults: a scoping review. PLoS ONE **15**, e0232421 (2020). https://doi.org/10.1371/journal.pone.0232421
3. Holden, L.K., et al.: Factors affecting open-set word recognition in adults with cochlear implants. Ear Hear. **34**, 342 (2013). https://doi.org/10.1097/AUD.0b013e3182741aa7
4. Chakravorti, S., et al.: Further evidence of the relationship between cochlear implant electrode positioning and hearing outcomes. Otol. Neurotol. **40**, 617 (2019). https://doi.org/10.1097/MAO.0000000000002204
5. Noble, J.H., Labadie, R.F., Gifford, R.H., Dawant, B.M.: Image-guidance enables new methods for customizing cochlear implant stimulation strategies. IEEE Trans. Neural Syst. Rehabil. Eng. **21**, 820–829 (2013). https://doi.org/10.1109/TNSRE.2013.2253333
6. Noble, J.H., Gifford, R.H., Hedley-Williams, A.J., Dawant, B.M., Labadie, R.F.: Clinical evaluation of an image-guided cochlear implant programming strategy. Audiol. Neurotol. **19**, 400–411 (2014). https://doi.org/10.1159/000365273
7. Noble, J.H., et al.: Initial results with image-guided cochlear implant programming in children. Otol. Neurotol. **37**, e63 (2016). https://doi.org/10.1097/MAO.0000000000000909
8. Zhao, Y., Dawant, B.M., Labadie, R.F., Noble, J.H.: Automatic localization of closely spaced cochlear implant electrode arrays in clinical CTs. Med. Phys. **45**, 5030–5040 (2018). https://doi.org/10.1002/mp.13185
9. Andersen, S.A.W., et al.: Automated calculation of cochlear implant electrode insertion parameters in clinical cone-beam CT. Otol. Neurotol. **43**, 199–205 (2022). https://doi.org/10.1097/MAO.0000000000003432
10. Zhao, Y., Dawant, B.M., Labadie, R.F., Noble, J.H.: Automatic localization of cochlear implant electrodes in CT. In: Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (eds.) Medical Image Computing and Computer-Assisted Intervention, pp. 331–338. Springer International Publishing, Cham (2014). https://doi.org/10.1007/978-3-319-10404-1_42
11. Bennink, E., Peters, J.P.M., Wendrich, A.W., Vonken, E., van Zanten, G.A., Viergever, M.A.: Automatic localization of cochlear implant electrode contacts in CT. Ear Hear. **38**, e376 (2017). https://doi.org/10.1097/AUD.0000000000000438
12. Zhao, Y., Chakravorti, S., Labadie, R.F., Dawant, B.M., Noble, J.H.: Automatic graph-based method for localization of cochlear implant electrode arrays in clinical CT with sub-voxel accuracy. Med. Image Anal. **52**, 1–12 (2019). https://doi.org/10.1016/j.media.2018.11.005

13. Braithwaite, B., et al.: Cochlear implant electrode localization in post-operative CT using a spherical measure. In: 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), pp. 1329–1333 (2016). https://doi.org/10.1109/ISBI.2016.7493512

14. Hachmann, H., Krüger, B., Rosenhahn, B., Nogueira, W.: Localization of cochlear implant electrodes from cone beam computed tomography using particle belief propagation. In: 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), pp. 593–597 (2021). https://doi.org/10.1109/ISBI48211.2021.9433845

15. Chi, Y., Wang, J., Zhao, Y., Noble, J.H., Dawant, B.M.: A deep-learning-based method for the localization of cochlear implant electrodes in CT images. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), pp. 1141–1145 (2019). https://doi.org/10.1109/ISBI.2019.8759536

16. Wang, J., Zhao, Y., Noble, J.H., Dawant, B.M.: Metal artifact reduction, intra cochlear anat-omy segmentation, and cochlear implant electrodes localization in CT images with a multi-task 3D network. In: Landman, B.A., Išgum, I. (eds.) Medical Imaging 2021: Image Processing. p. 23. SPIE, Online Only, United States (2021). https://doi.org/10.1117/12.2580931

17. Margeta, J., et al.: A web-based automated image processing research platform for cochlear implantation-related studies. J. Clin. Med. **11**, 6640 (2022). https://doi.org/10.3390/jcm11226640

18. Zhao, Y., Labadie, R.F., Dawant, B.M., Noble, J.H.: Validation of automatic cochlear im-plant electrode localization techniques using μCTs. J. Med. Imaging. **5**, 035001 (2018). https://doi.org/10.1117/1.JMI.5.3.035001

19. Zhou, X., Wang, D., Krähenbühl, P.: Objects as Points. ArXiv190407850 Cs. (2019)

20. Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. IEEE Trans. Pattern Anal. Mach. Intell. **42**, 318–327 (2020). https://doi.org/10.1109/TPAMI.2018.2858826

21. Shit, S., et al.: clDice -- a novel topology-preserving loss function for tubular structure seg-mentation. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 16555–16564 (2021). https://doi.org/10.1109/CVPR46437.2021.01629

22. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2014)

23. Cardoso, M.J. et al.: MONAI: An open-source framework for deep learning in healthcare, http://arxiv.org/abs/2211.02701 (2022). https://doi.org/10.48550/arXiv.2211.02701

24. Greenwood, D.D.: A cochlear frequency-position function for several species—29 years later. J. Acoust. Soc. Am. **87**, 2592–2605 (1990). https://doi.org/10.1121/1.399052