HHS Public Access

Author manuscript

Med Image Comput Comput Assist Interv. Author manuscript; available in PMC 2020 December 03.

Published in final edited form as:

Med Image Comput Comput Assist Interv. 2020 October ; 12261: 802–812. doi:10.1007/978-3-030-59710-8_78.

Cartilage Segmentation in High-Resolution 3D Micro-CT Images via Uncertainty-Guided Self-training with Very Sparse Annotation

Hao Zheng[1], **Susan M. Motch Perrine**[2], **M. Kathleen Pitirri**[2], **Kazuhiko Kawasaki**[2], **Chaoli Wang**[1], **Joan T. Richtsmeier**[2], **Danny Z. Chen**[1]

[1]Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, IN 46556, USA

[2]Department of Anthropology, Pennsylvania State University, University Park, PA 16802, USA

Abstract

Craniofacial syndromes often involve skeletal defects of the head. Studying the development of the *chondrocranium* (the part of the endoskeleton that protects the brain and other sense organs) is crucial to understanding genotype-phenotype relationships and early detection of skeletal malformation. Our goal is to segment craniofacial cartilages in 3D micro-CT images of embryonic mice stained with phosphotungstic acid. However, due to high image resolution, complex object structures, and low contrast, delineating fine-grained structures in these images is very challenging, even manually. Specifically, only experts can differentiate cartilages, and it is unrealistic to manually label whole volumes for deep learning model training. We propose a new framework to progressively segment cartilages in high-resolution 3D micro-CT images using extremely sparse annotation (e.g., annotating only a few selected slices in a volume). Our model consists of a lightweight fully convolutional network (FCN) to accelerate the training speed and generate pseudo labels (PLs) for unlabeled slices. Meanwhile, we take into account the reliability of PLs using a bootstrap ensemble based uncertainty quantification method. Further, our framework gradually learns from the PLs with the guidance of the uncertainty estimation via self-training. Experiments show that our method achieves high segmentation accuracy compared to prior arts and obtains performance gains by iterative self-training.

Keywords

Cartilage segmentation; Uncertainty; Sparse annotation

1 Introduction

Approximately 1% of babies born with congenital anomalies have syndromes including skull abnormalities [13]. Anomalies of the skull invariably require treatments and care, imposing high financial and emotional burdens on patients and their families. Although prenatal development data are not available for study in humans, the deep conservation of

hzheng3@nd.edu.

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

mammalian developmental systems in evolution means that laboratory mice give access to embryonic tissues that can reveal critical molecular and structural components of early skull development [3, 18]. The precise delineation of 3D chondrocranial anatomy is fundamental to understanding dermatocranium development, provides important information to the pathophysiology of numerous craniofacial anomalies, and reveals potential avenues for developing novel therapeutics. An embryonic mouse is tiny (~2 cm^3), and thus we dissect and reconstruct the *chondrocranium* from 3D micro-computed tomography (micro-CT) images of specially stained mice. However, delineating fine-grained cartilaginous structures in these images is very challenging, even manually (e.g., see Fig. 1).

Although deep learning has achieved great success in biomedical image segmentation [11, 12, 19, 20, 22], there are three main challenges when applying existing methods to cartilage segmentation in our high-resolution micro-CT images. (1) The topology variations of craniofacial cartilages are very large in the anterior, intermediate, and posterior of the skull (as shown in Fig. 1(a)). Known methods for segmenting articular cartilages in knees [2, 17] only deal with relatively homogeneous structures. (2) Such methods deal with images of much lower resolutions (e.g., 200×512^2), and simple scaling-up would precipitate huge computation requirements. Micro-CT scanners work at the level of one micron (i.e., $1 \mu m$), and a typical scan of ours is of size 1500×2000^2. In Fig. 1(c), the cropped sub-region is of size 400^2, and the region-of-interest (ROI) is only 5 pixels thick. (3) More importantly, only experts can differentiate cartilages, and it is unrealistic to manually label whole volumes for training fully convolution networks (FCNs) [12]. While some semi-supervised methods [21, 23] were studied very recently, how to acquire and make the most out of very sparse annotation is seldom explored, especially for real-world complex cartilage segmentation tasks.

To address these challenges, we propose a new framework that utilizes FCNs and uncertainty-guided self-training to gradually boost the segmentation accuracy. We start with extremely sparsely annotated 2D slices and train an FCN to predict pseudo labels (PLs) for unseen slices in the training volumes and the associated uncertainty map, which quantifies pixelwise prediction confidence. Guided by the uncertainty, we iteratively train the FCN with PLs and improve the generalization ability of FCN in unseen volumes. Although the above process seems straightforward, we must overcome three difficulties. (1) The FCN should have a sufficiently large receptive field to accommodate such high-resolution images yet needs to be lightweight for efficient training and inference due to the large volumes. (2) Bayesian-based uncertainty quantification requires a linear increase of either space or time during inference. We integrate FCNs into a bootstrap ensemble based uncertainty quantification scheme and devise a K-head FCN to balance efficiency and efficacy. (3) The generated PLs contain noises. We consider the quality of PLs and propose an uncertainty-guided self-training scheme to further refine segmentation results.

Experiments show that our proposed framework achieves an average Dice of 78.98% in segmentation compared to prior arts and obtains performance gains by iterative self-training (from 78.98% to 83.16%).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

2 Method

As shown in Fig. 2, our proposed framework contains a new FCN, which can generate PLs and uncertainty estimation at the same time, and an iterative uncertainty-guided self-training strategy to boost the segmentation results.

2.1 K-Head FCN

Initial Labeling and PL Generation.—We consider two sets of 3D data, $\mathscr{A} = \{\mathscr{A}_i\}_{i=1}^{L}$ and $\mathscr{B} = \{\mathscr{B}_i\}_{i=1}^{U}$, for training and testing respectively, where each $\mathscr{A}_i$ (or $\mathscr{B}_i$) is a 3D volume and L (or U) is the number of volumes in $\mathscr{A}$ (or $\mathscr{B}$). Each 3D volume can be viewed as a series of 2D slices, i.e., $\mathscr{A}_i = \{\mathbf{A}_i^j\}_{j=1}^{i_Q}$, where i_Q is the number of slices in $\mathscr{A}_i$. To begin with, experts chose representative slices in each $\mathscr{A}_i$ from the anterior, intermediate, and posterior of the skull and annotated them at the pixel level. Due to the high resolution of our micro-CT images, the annotation ratio is rather sparse (e.g., 25 out of 1600 slices). Thus, each $\mathscr{A}_i$ can be divided into two subsets $\mathscr{A}l_i = \{\mathbf{l}_i^j\}_{j=1}^{i_P}$ and $\mathscr{A}u_i = \{\mathbf{u}_i^j\}_{j=1}^{i_R}$, where each slice $\mathbf{l}_i^j$ has its associate label $\mathbf{m}_i^j$, and $i_Q > i_R \gg i_P$. Conventionally, using such sparse annotation, a trained FCN lacks generalization ability to the unseen volumes B. Hence, a key challenge is how to make the most out of the labeled slices. We will show that an FCN can delineate ROIs in unseen slices of the training volumes (i.e., $\mathscr{A}u_i$) with very sparsely labeled slices. For this, we propose to utilize these true labels (TLs) and generate PLs to expand the training data.

Uncertainty Quantification.—Since FCN here is not trained by standard protocol, its predictions may be unreliable and noisy. Thus, we need to consider the reliability of the PLs (which may otherwise lead to meaningless guidance). Bayesian methods [7] provided a straightforward way to measure uncertainty quantitatively by utilizing Monte Carlo sampling in forward propagation to generate multiple predictions. Prohibitively, the computational cost grows linearly (either time or space). Since our data are large volumes, such cost is unbearable. To avoid this issue, we need to design a method that is both time- and space-efficient. Below we illustrate how to design a new FCN for this purpose.

There are two main types of uncertainty in Bayesian modelling [8, 16]: *epistemic uncertainty* captures uncertainty in the model (i.e., the model parameters are poorly determined due to the lack of data/knowledge); *aleatoric uncertainty* captures genuine stochasticity in the data (e.g., inherent noises). Without loss of generality, let $f_\theta(x)$ be the output of a neural network, where θ is the parameters and x is the input. For segmentation tasks, following the practice in [8], we define pixelwise likelihood by squashing the model output through a softmax function S: $p\left(y \mid f_\theta(x), \sigma^2\right) = S\left(\frac{1}{\sigma^2} f_\theta(x)\right)$. The magnitude of $\boldsymbol{\sigma}$ determines how 'uniform' (flat) the discrete distribution is. The log likelihood for the output

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

is: $\log\left(p\left(y = c \mid f_\theta(x), \sigma^2\right)\right) = \frac{1}{\sigma^2} f_\theta^c(x) - \log \sum_{c'} \exp\left(\frac{1}{\sigma^2} f_\theta^{c'}(x)\right) = \frac{1}{\sigma^2} \log \frac{\exp\left(f_\theta^c(x)\right)}{\sum_{c'} \exp\left(f_\theta^{c'}(x)\right)}$, where $f_\theta^c(x)$ $- \log \frac{\sum_{c'} \exp\left(\frac{1}{\sigma^2} f_\theta^{c'}(x)\right)}{\left(\sum_{c'} \exp\left(f_\theta^{c'}(x)\right)\right)^{\frac{1}{\sigma^2}}} \approx \frac{1}{\sigma^2} \log S\left(f_\theta(x)\right)^c - \frac{1}{2} \log \sigma^2$ is the c-th class of output $f_\theta(x)$, and we use the explicit simplifying assumption $\left(\sum_{c'} \exp\left(f_\theta^{c'}(x)\right)\right)^{\frac{1}{\sigma^2}} \approx \frac{1}{\sigma} \sum_{c'} \exp\left(\frac{1}{\sigma^2} f_\theta^{c'}(x)\right)$. The objective is to minimize the loss given by the negative log likelihood:

$$\mathscr{L}_{UC}\left(\theta, \sigma^2\right) = -\frac{1}{N} \sum_i^N \sum_m^M \mathbb{1}_{m = c} \log\left(p\left(y_i = c \mid f_\theta(x_i), \sigma^2\right)\right), \tag{1}$$

where N is the number of training samples and $\mathbb{1}_{m = c}$ is the one-hot vector of class c. In practice, we make the network predict the log variance $s := \log \sigma^2$ for numerical stability. Now, the aleatoric uncertainty is estimated by e^{-s}, and we can quantify the epistemic uncertainty by the predictive variance by $\frac{1}{K} \sum_k^K \hat{y}_k^2 - \left(\frac{1}{K} \sum_k^K \hat{y}_k\right)^2$, where $\hat{y}_k = f_\theta(x)$ is the k-th sample from the output distribution.

K-Head FCN.—To sample K samples from the output distribution, we adopt the bootstrap method into the FCN design. A naïve way would be to maintain a set of K networks $\{f_{\theta_k}\}_{k = 1}^K$ independently on K different bootstrapped subsets (i.e., $\{D_k\}_{k = 1}^K$) of the whole dataset D and treat each network f_{θ_k} as independent samples from the weight distribution. However, it is computationally expensive, especially when each neural net is large and deep. Hence, we propose a single network that consists of a shared backbone architecture with K lightweight bootstrapped heads branching on/off independently. The shared network learns a joint feature representation across all the data, while each head is trained only on its bootstrapped sub-sample of the data. The training and inference of this type of bootstrap can be conducted in a single forward/backward pass, thus saving both time and space. Besides, in contrast to previous methods where σ^2 is assumed to be constant for all inputs, we estimate it directly as an output of the network [7, 16]. Thus, our proposed network consists of a total of $K + 1$ branches—K heads corresponding to the segmentation prediction map and an extra head corresponding to σ^2. In all the experiments, K is set as 5, and the input image size is 512 × 512.

Figure 3 shows the detailed structure of our new K-head FCN. There are 7 residual blocks (RBs) and max-pooling operations in the encoding-path to deliver larger reception fields, each RB containing 2 cascaded residual units as in ResNet [6]. To save parameters, we maintain the number of channels in each residual unit and a similar number of feature channels at the last 4 scales. Rich contextual and semantic information is extracted in shallower and deeper scales in the encoding-path and is up-sampled to maintain the same size for the input and output and then concatenated to generate the final prediction. The output layer splits near the end of the model for two reasons: (1) ease the training difficulty and improve the convergence speed; (2) incur minimal computation resource increases (both time and space) in training and inference. To train the network, we randomly choose one head in each iteration and compute the cross-entropy loss $\mathscr{L}_{CE}$. It is combined with the uncertainty loss $\mathscr{L}_{UC}$ to update the parameters in the chosen head branch and the shared backbone only (i.e., freezing the other $K-1$ head branches). Specifically, $\mathscr{L} = \mathscr{L}_{CE} + 0.04$ $\mathscr{L}_{UC}$.

2.2 Iterative Uncertainty-Guided Self-Training

Since both $\mathscr{A}l_i$ and $\mathscr{A}u_i$ come from the same volume $\mathscr{A}_i$ and are based on the assumption that the manifolds of the seen/unseen slices (of $\mathscr{A}_i$) are smooth in high dimensions [15], our generated PLs bridge the annotation gap. However, the K predictions, $\{\widehat{\mathbf{m}}_i^{j,k}\}_{k=1}^{K}$, obtained from the output distribution for each $\mathbf{u}_i^j \in \mathscr{A}u_i$ could be unreliable and noisy. Thus, we propose an uncertainty-guided scheme to reweight PLs and rule out unreliable (highly uncertain) pixels in subsequent training. Specifically, we calculate the voxel-level cross-entropy loss weighted by the epistemic uncertainty $\boldsymbol{\sigma}_i^j$ for

$\mathbf{u}_i^j : \mathscr{L}_{CE}(\overline{\mathbf{m}}_i^j, \widetilde{\mathbf{m}}_i^j) = \frac{\sum_v e^{-\sigma_v} \mathscr{L}_{ce}(\overline{m}_v, \widetilde{m}_v)}{\sum_v e^{-\sigma_v}}$, where $\overline{\mathbf{m}}_i^j$ is the prediction at the current iteration and $\widetilde{\mathbf{m}}_i^j = \sum_{k=1}^{K} \widehat{\mathbf{m}}_i^{j,k}$; $\overline{m}_v$ and $\widetilde{m}_v$ are the values of the v-th pixel (for simplicity, we omit i and j); σ_V is the sum of normalized epistemic and aleatoric uncertainties at the v-th pixel; $\mathscr{L}_{ce}$ is the cross-entropy error at each pixel. Note that we do not choose a hard threshold to convert the average probability map $\widetilde{\mathbf{m}}_i^j$ to a binary mask, as inspired by the "label smoothing" technique [14] which may help prevent the network from becoming over-confident and improve generalization ability.

With the expansion of the training set (TLs ∪ PLs), our FCN can distill more knowledge about the data (e.g., topological structure, intensity variances), thus becoming more robust and generalizing better to unseen data $\mathscr{B}$. However, due to the extreme sparsity of annotation at the very beginning, not all the generated PLs are evenly used (i.e., highly uncertain and assigned with low weights). Hence, we propose to conduct this process iteratively.

Overall, with our iterative uncertainty-guided self-training scheme, we can further refine the PLs and FCN at the same time. In practice, it needs 2 or 3 rounds, but we do not have to train from scratch, incurring not too much cost.

3 Experiments and Results

Data Acquisition.

Mice were produced, sacrificed, and processed in compliance with animal welfare guidelines approved by the Pennsylvania State University (PSU). Embryos were stained with phosphotungstic acid (PTA), as described in [10]. Data were acquired by the PSU Center for Quantitative Imaging using the General Electric v|tom|x L300 nano/micro-CT system with a 180-kV nanofocus tube and were then reconstructed into micro-CT volumes with a resulting

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

average voxel size of 5μm and volume size of 1500 × 2000^2. Seven volumes are divided into the training set $\mathscr{A} = \{\mathscr{A}_i\}_{i=1}^{4}$ and test set $\mathscr{B} = \{\mathscr{B}_i\}_{i=1}^{3}$. Only a very small subset of slices in each $\mathscr{A}_i$ is labeled for training (denoted as $\mathscr{A}l_i$) and the rest unseen slices $\mathscr{A}u_i$ and $\mathscr{B}$ are used for the test. Four scientists with extensive experience in the study of embryonic bones/cartilages were involved in image annotations. They first annotated slices in the 2D plane and then refined the whole annotation by considering 3D information of the neighboring slices.

Evaluation.

In the 3D image regions not considered by the experts, we select 11 3D subregions (7 from $\mathscr{B}$ and 4 from $\mathscr{A}u_i$), each of an average size 30×300^2 and containing at least one piece of cartilages. These subregions are chosen for their representativeness, i.e., they cover all the typical types of cartilages (e.g., nasal capsule, Meckel’s cartilage, lateral wall, braincase floor, etc). Each subregion is manually labeled by experts as ground truth. The segmentation accuracy is measured by Dice-Sørensen Coefficient (DSC).

Implementation Details.

All our networks are implemented with TensorFlow [1], initialized by the strategy in [5], and trained with the Adam optimizer [9] (with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 1\text{e-}10$). We adopt the “poly” learning rate policy $L_r \times \left(1 - \frac{iter}{\#\ iter}\right)^{0.9}$, where the initial rate $L_r = 5\text{e-}4$ and the max iteration number is set as 60k. To leverage the limited training data and reduce overfitting, we augment the training data with standard operations (e.g., random crop, flip, rotation in 90°, 180°, and 270°). Due to large intensity variance among different images, all images are normalized to have zero mean and unit variance.

Main Results.

The results are summarized in Table 1. To our best knowledge, there is no directly related work on cartilage segmentation from embryonic tissues. We compare our new framework with the following methods. (1) A previous work which utilizes U-Net [19] to automatically segment knee cartilages [2]. We also try another robust FCN model DCN [4]. For a fair comparison, we scale up U-Net [19] and DCN [4] to accommodate images of size 512^2 as input and match with the number of parameters of our K-head FCN (denoted as U-Net* and DCN*). (2) A semi-supervised method that generates PLs and conducts self-training (i.e., 1-head FCN-R3).

First, compared with known FCN-based methods, our K-head FCN yields better performance for cartilages in different positions. We attribute this to its deeper structures and multi-scale extracted feature fusion design, which leads to larger receptive fields and richer spatial and semantic features. Hence, our backbone model can capture significant topology variances in skull cartilages (e.g., relatively small but thick nasal parts, and large but thin shell-like cranial base and vault). Second, to show that our K-head FCN is comparable with Monte Carlo sampling based Bayesian methods, we implement 1-head FCN and conduct sampling K times to obtain PLs. Repeating the training process 3 times (denoted as ‘-R3’),

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

we observe that using PLs, K-head FCN-R3 achieves similar performance as 1-head FCN-R3. However, in each forward pass, we obtain K predictions at once, thus saving ~ K× the time/space costs. Qualitative results are shown in Fig. 4. Third, we further show that under the guidance of uncertainty, our new method (K-head FCN-R3-U) attains performance gain (from 82.45% to 83.16%). We attribute this to that unreliable PLs are ruled out, and the model optimizes under cleaner supervisions.

Discussions.

(1) *Iteration Numbers*. We measure DSC scores on both unseen slices in the training volumes $(\{\mathscr{A}u_i\}_{i=1}^{L})$ and unseen slices in the test volumes $(\{\mathscr{B}_i\}_{i=1}^{U})$ during the training of "K-head FCN-R3-U" (see Table 1 bottom-left). We notice significant performance gain after expanding the training set (i.e., TLs → TLs ∪ PLs, as Iter-1 → Iter-2). Meanwhile, because the uncertainty of only a small amount of pixels changes during the whole process, the performance gain is not substantial from Iter-2 to Iter-3. **(2)** *Annotation Ratios*. As shown in Table 1 bottom-right, the final segmentation results can be improved using more annotation, but the improvement rate decreases when labeling more slices. **(3)** *Uncertainty Estimation*. We visualize the samples along with estimated segmentation results and the corresponding epistemic and aleatoric uncertainties from the test data in Fig. 5. It is shown that the model is less confident (i.e., with a higher uncertainty) on the boundaries and hard mimic regions where the epistemic and aleatoric uncertainties are prominent.

4 Conclusions

We presented a new framework for cartilage segmentation in high-resolution 3D micro-CT images with very sparse annotation. Our K-head FCN produces segmentation predictions and uncertainty estimation simultaneously, and the iterative uncertainty-guided self-training strategy gradually refines the segmentation results. Comprehensive experiments showed the efficacy of our new method.

Acknowledgement.

This research was supported in part by the US National Science Foundation through grants CNS-1629914, CCF-1617735, IIS-1455886, DUE-1833129, and IIS-1955395, and the National Institute of Dental and Craniofacial Research through grant R01 DE027677.

References

1. Abadi M, et al.: TensorFlow: a system for large-scale machine learning. In: OSDI, vol. 16, pp. 265–283 (2016)
2. Ambellan F, Tack A, Ehlke M, Zachow S: Automated segmentation of knee bone and cartilage combining statistical shape knowledge and convolutional neural networks: data from the osteoarthritis initiative. Med. Image Anal 52, 109–118 (2019) [PubMed: 30529224]
3. Brinkley JF, et al.: The facebase consortium: a comprehensive resource for craniofacial researchers. Development 143(14), 2677–2688 (2016) [PubMed: 27287806]
4. Chen H, Qi XJ, Cheng JZ, Heng PA: Deep contextual networks for neuronal structure segmentation. In: Thirtieth AAAI Conference on Artificial Intelligence, pp. 1167–1173 (2016)
5. He K, Zhang X, Ren S, Sun J: Delving deep into rectifiers: surpassing human level performance on imagenet classification. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1026–1034 (2015)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

6. He K, Zhang X, Ren S, Sun J: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
7. Kendall A, Gal Y: What uncertainties do we need in Bayesian deep learning for computer vision? In: Advances in Neural Information Processing Systems, pp. 5574–5584 (2017)
8. Kendall A, Gal Y, Cipolla R: Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7482–7491 (2018)
9. Kingma DP, Ba J: Adam: a method for stochastic optimization. In: Third International Conference on Learning Representations (2015)
10. Lesciotto KM, et al.: Phosphotungstic acid-enhanced microCT: optimized protocols for embryonic and early postnatal mice. Dev. Dyn 249, 573–585 (2020). 10.1002/dvdy.136 [PubMed: 31736206]
11. Liang P, Chen J, Zheng H, Yang L, Zhang Y, Chen DZ: Cascade decoder: a universal decoding method for biomedical image segmentation. In: IEEE 16th International Symposium on Biomedical Imaging (ISBI), pp. 339–342 (2019)
12. Long J, Shelhamer E, Darrell T: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431–3440 (2015)
13. Mossey PA, Catilla EE, et al.: Global registry and database on craniofacial anomalies: report of a WHO registry meeting on craniofacial anomalies (2003)
14. Müller R, Kornblith S, Hinton GE: When does label smoothing help? In: Advances in Neural Information Processing Systems, pp. 4696–4705 (2019)
15. Niyogi P: Manifold regularization and semi-supervised learning: some theoretical analyses. J. Mach. Learn. Res 14(1), 1229–1250 (2013)
16. Oh M.h., Olsen PA, Ramamurthy KN: Crowd counting with decomposed uncertainty. In: Thirty-Fourth AAAI Conference on Artificial Intelligence, pp. 11799–11806 (2020)
17. Prasoon A, Petersen K, Igel C, Lauze F, Dam E, Nielsen M: Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network In: Mori K, Sakuma I, Sato Y, Barillot C, Navab N (eds.) MICCAI 2013. LNCS, vol. 8150, pp. 246–253. Springer, Heidelberg (2013). 10.1007/978-3-642-40763-531
18. Richtsmeier JT, Baxter LL, Reeves RH: Parallels of craniofacial maldevelopment in Down syndrome and Ts65Dn mice. Dev. Dyn 217(2), 137–145 (2000) [PubMed: 10706138]
19. Ronneberger O, Fischer P, Brox T: U-Net: convolutional networks for biomedical image segmentation In: Navab N, Hornegger J, Wells WM, Frangi AF (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). 10.1007/978-3-319-24574-428
20. Wang Y, et al.: Deep attentional features for prostate segmentation in ultrasound In: Frangi AF, Schnabel JA, Davatzikos C, Alberola-López C, Fichtinger G (eds.) MICCAI 2018. LNCS, vol. 11073, pp. 523–530. Springer, Cham (2018). 10.1007/978-3-030-00937-360
21. Yu L, Wang S, Li X, Fu C-W, Heng P-A: Uncertainty-aware self-ensembling model for semi-supervised 3D left atrium segmentation In: Shen D, Liu T, Peters TM, Staib LH, Essert C, Zhou S, Yap P-T, Khan A (eds.) MICCAI 2019. LNCS, vol. 11765, pp. 605–613. Springer, Cham (2019). 10.1007/978-3-030-32245-867
22. Zheng H, et al.: HFA-Net: 3D cardiovascular image segmentation with asymmetrical pooling and content-aware fusion In: Shen D, et al. (eds.) MICCAI 2019. LNCS, vol. 11765, pp. 759–767. Springer, Cham (2019). 10.1007/978-3-030-32245-884
23. Zheng H, Zhang Y, Yang L, Wang C, Chen DZ: An annotation sparsification strategy for 3D medical image segmentation via representative selection and self-training. In: Thirty-Fourth AAAI Conference on Artificial Intelligence, pp. 6925–6932 (2020)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

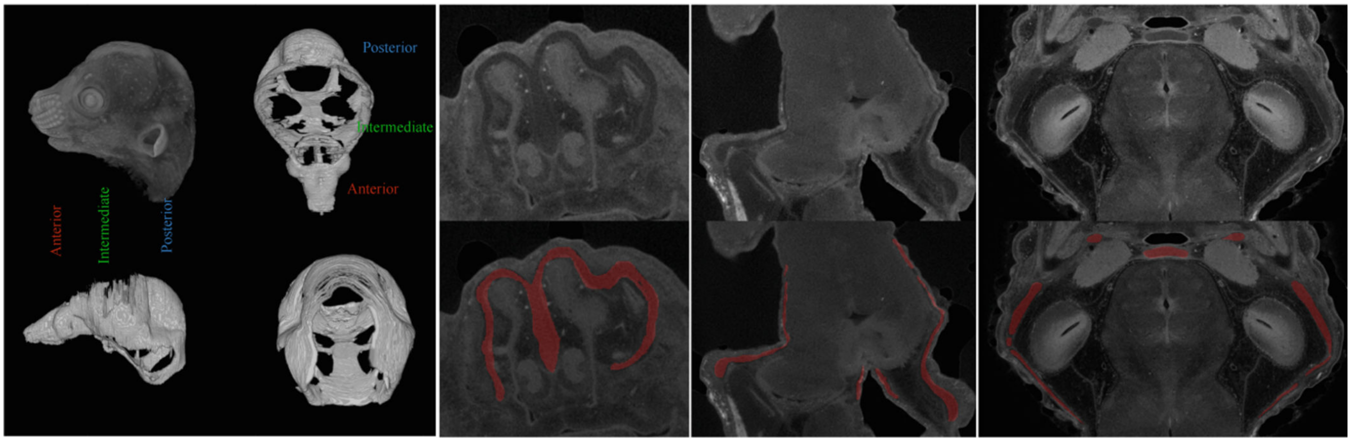


Fig. 1.

Examples of micro-CT images of stained mice. (a) A raw 3D image and its manual annotation. The shape variations are large: the front nasal cartilage is relatively small (i.e., 300^2); the cranial vault is very big (i.e., 900 × 500) but extremely thin like a half-ellipsoid surface. (b) A 2D slice from the nasal cartilage (top) and its associated label (bottom); the image contrast is low and there are many hard mimics in surrounding areas. (c) Two 2D slices from the cranial vault (top) and their associated labels (bottom); the cartilage is very thin. Best viewed in color.

Author Manuscript

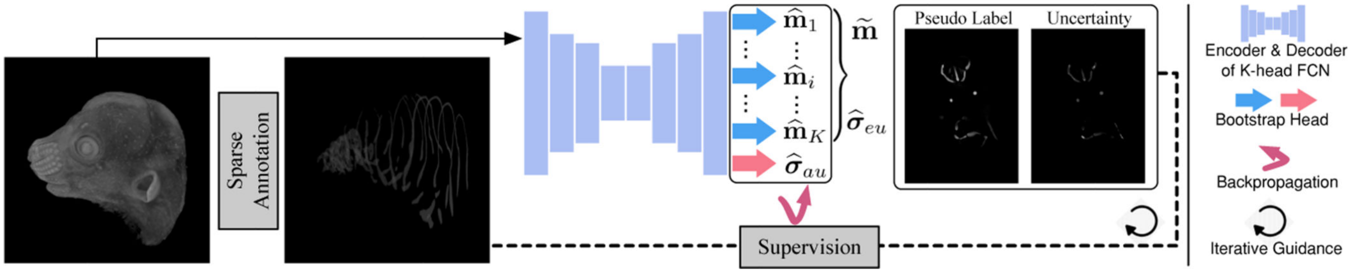


Fig. 2.

An overview of our proposed framework.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

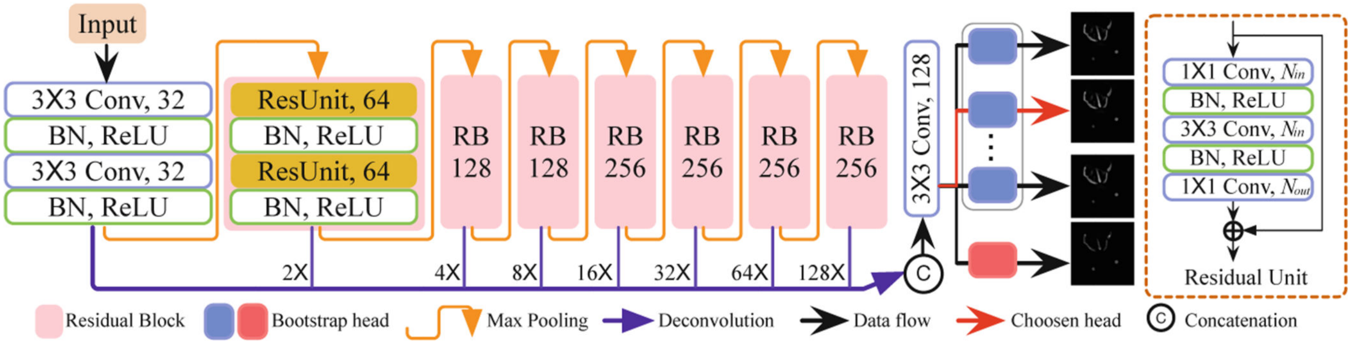


Fig. 3.

The network architecture of our proposed method, K-head FCN. The output layer branches out to *K* bootstrap heads and an extra log-variance output.

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

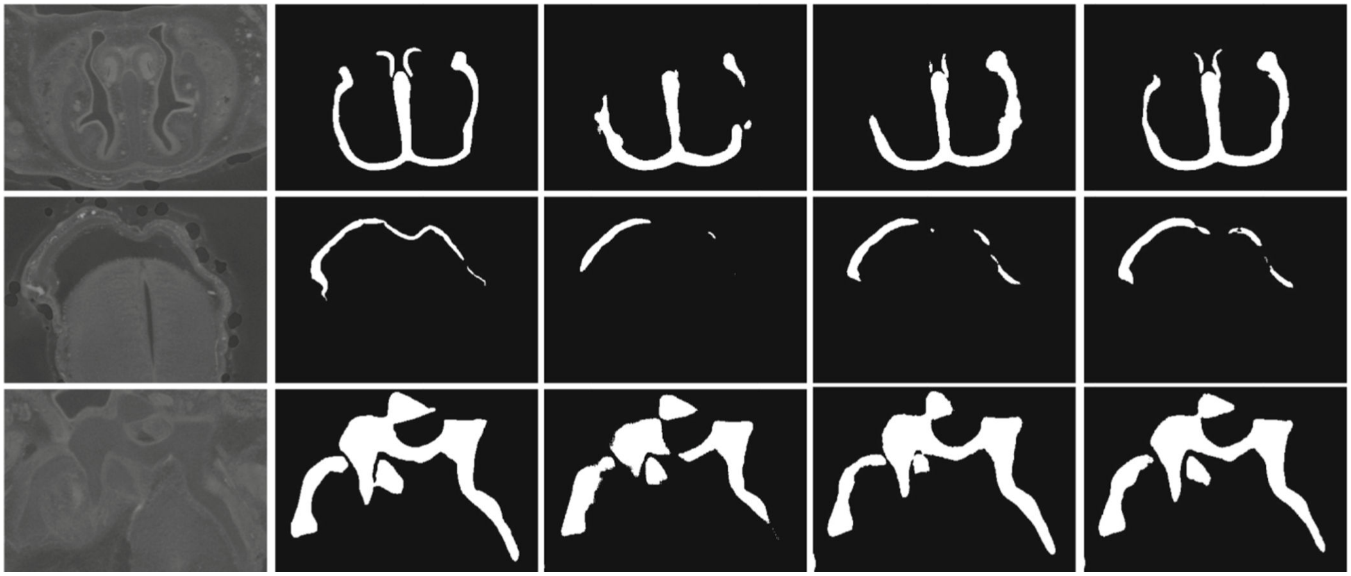


Fig. 4.

Qualitative examples: (a) Raw subregions; (b) ground truth; (c) U-Net* (TL); (d) K-head FCN (TL); (e) K-head FCN-R3-U (TLUPL). (XX) = (trained using XX).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

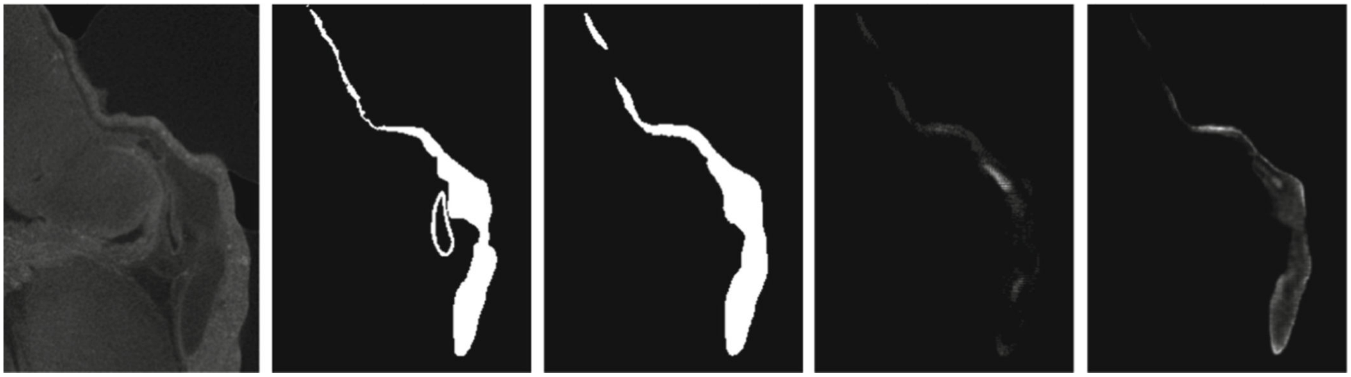


Fig. 5.

Visualization of uncertainty. From left to right: a raw image region, ground truth, prediction result, estimated epistemic uncertainty, and estimated aleatoric uncertainty. Brighter white color means higher uncertainty.

Author Manuscript

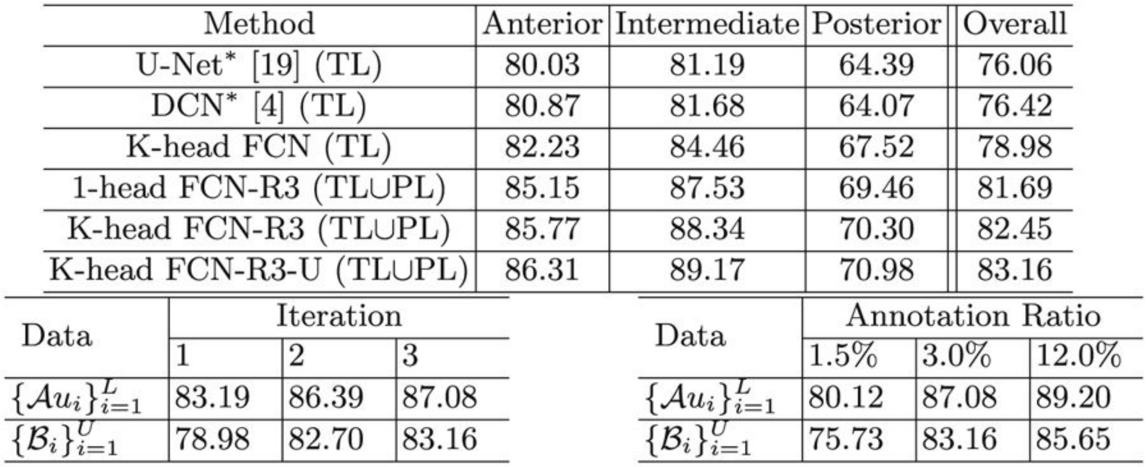
Author Manuscript

Author Manuscript

Author Manuscript

Table 1.

Segmentation results. Top: DSC (%) comparison of cartilages in the anterior, intermediate, and posterior skull, w/annotation ratio of 3.0%. TL: true labels; PL: pseudo labels. Bottom-left: “K-head FCN-R3-U (TLUPL)” w/ annotation ratio of 3.0%. Bottom-right: “K-head FCN-R3-U (TLUPL)” w/different annotation ratios.



Method	Anterior	Intermediate	Posterior	Overall
U-Net* [19] (TL)	80.03	81.19	64.39	76.06
DCN* [4] (TL)	80.87	81.68	64.07	76.42
K-head FCN (TL)	82.23	84.46	67.52	78.98
1-head FCN-R3 (TL∪PL)	85.15	87.53	69.46	81.69
K-head FCN-R3 (TL∪PL)	85.77	88.34	70.30	82.45
K-head FCN-R3-U (TL∪PL)	86.31	89.17	70.98	83.16

Data	Iteration		
	1	2	3
$\{\mathcal{A}u_i\}_{i=1}^{L}$	83.19	86.39	87.08
$\{\mathcal{B}_i\}_{i=1}^{U}$	78.98	82.70	83.16

Data	Annotation Ratio		
	1.5%	3.0%	12.0%
$\{\mathcal{A}u_i\}_{i=1}^{L}$	80.12	87.08	89.20
$\{\mathcal{B}_i\}_{i=1}^{U}$	75.73	83.16	85.65