



HAL
open science

Segmenting partially annotated medical images

Nicolas Martin, Jean-Pierre Chevallet, Georges Quénot

► **To cite this version:**

Nicolas Martin, Jean-Pierre Chevallet, Georges Quénot. Segmenting partially annotated medical images. CBMI 2022: International Conference on Content-based Multimedia Indexing, Sep 2022, Graz Austria, France. pp.111-115, 10.1145/3549555.3549570 . hal-03811749

HAL Id: hal-03811749

<https://hal.science/hal-03811749>

Submitted on 17 May 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Segmenting partially annotated medical images

NICOLAS MARTIN, Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG, France

JEAN-PIERRE CHEVALLET, Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG, France

GEORGES QUÉNOT, Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG, France

Segmentation of medical images using learning based systems remains a challenge in medical computer vision: training a segmentation model requires medical images exhaustively annotated by experts that are difficult and expensive to obtain. We propose to explore the usage of partially annotated images, i.e., all images are annotated but not all regions of a given class are annotated. In this paper, we propose several approaches and we experiment them on the segmentation of intra-oral images. First, we propose to modify the loss function to consider only the annotated areas, and second to integrate annotation from non-expert, as well as the combination of these methods. The experiments we conducted showed an improvement up to 33% on the segmentation performance. This approach allows to obtain better quality annotation masks than the initial human annotation using only partially annotated areas or non-expert annotations. In the future, these approaches can be extended by combination with active learning methods.

CCS Concepts: • **Computing methodologies** → **Semi-supervised learning settings**; **Image segmentation**.

Additional Key Words and Phrases: Image segmentation, Medical images, Partially annotated data

1 INTRODUCTION

Medical image segmentation is considered as one the most challenging task in computer vision for medical images [8]. Correctly segmenting anatomical structures (e.g., organ) is usually an essential step in the pipeline of the computer-aided detection system[12].

With the emergence of Deep Learning algorithms, specific architectures for segmentation problems have been proposed such as Mask R-CNN [6] or U-Net [16]. Relying on convolutional neural network (CNN), these models aim to classify each pixel of an image to belong to one or several classes [14]. Although these deep learning models generally offer better performance than prior methods (e.g., watershed), they are difficult to use in the medical field: annotation time by expert is costly, also it is a waste of resources to ask expert to annotate obvious element like teeth, or gum. It is better to ask expert to annotate areas of the image where his expertise is mandatory. For these reasons, expert annotations cannot guarantee to be exhaustive.

In this paper, we explore the possibility to segment gum and teeth on dental RGB images with partially annotated data by experts, and annotations from non-experts.

2 RELATED WORK

Segmentation models based on deep learning methods require an annotation for each pixel of the image, i.e., one or more classes to which the pixel belongs. In this paper, we name *masks* these annotations. As they are difficult to obtain, numerous methods have been explored to segment images in the medical field, especially when the quantity of data is limited (see [18] for a review). These methods can be categorized in two classes: methods dealing with scarce annotations (i.e., limited number of annotated data) and methods dealing with sparse annotations (i.e., incomplete or noisy annotations).

On one hand, the methods related to scarce data share the philosophy of increasing the information available to improve learning. They can rely on the exploitation of external data (e.g., transfer learning), on the increase in the variability of the data (e.g., data augmentation) or even on the incorporation of prior knowledge about the shape of target regions during the training [18].

On the other hand, some studies tried to exploit sparse annotation, easier to collect: point, bounding box (e.g., [3]), scribbles (e.g., [2]). To exploit these partially annotated data, part of research work attempts to automatically complete the annotation masks using various methods such as label propagation or mask dilatation (e.g., [5]). A last part of studies (including this paper) explore the segmentation without annotation reconstruction. For this purpose, the loss functions are modified to exploit this incomplete information. Bokhorst et al. [1] explore the use of mask of valid pixels (annotated pixels) and invalid pixels (not annotated pixels). This mask is used in such a way that invalid pixels are not taken into account in the loss computation. Çiçek et al. [20] explore a similar approach on segmentation of 3D images. They created a new class for unlabeled pixels. A weight of 0 was applied so that this class does not contribute to the loss computation.

In the current paper, we experiment several similar approaches for the segmentation of medical images using partial annotation and without annotation reconstruction: we also propose loss modifications. The novelty relies on the exploitation of non-expert annotations. We also explore the information fusion between classes.

3 PROPOSED FRAMEWORK

3.1 Network architecture

We first have to choose the neural network architecture. Numerous architectures [8] has already been proposed for image segmentation such as Mask R-CNN [6] or U-Net [16]. U-Net is particularly suitable for segmentation of medical images as it provides good performance even when the amount of data is limited [20].

Thereby, in this paper, a modified version of U-Net based on ResNet18 [7] and pretrained on ImageNet [17] is chosen. 1×1 convolutions are added before each increase in the number of filters. To improve the performance and generalization, a batch normalization [10] is also added after each convolution layer. In this paper, we explore modifications of the loss function, and the use of non-expert annotation, therefore the network architecture is fixed.

3.2 Loss function

The loss function expresses the learning goal of the network. It influences the accuracy of the learning. For segmentation models, pixel-wise binary cross-entropy (BCE) loss gives good results [11]. It is defined in equation (1):

$$L_{BCE}(y, \hat{y}) = -(y \cdot \log(\hat{y}) + (1 - y) \cdot \log(1 - \hat{y})) \quad (1)$$

where \hat{y} is the model predicted value for one pixel of the input image, and y is the expected value for the class mask at the same pixel position. It is a binary value and "1" means that the pixel belongs to the class of this mask. The loss function is computed and averaged over all pixels of the input image.

A partial annotation at pixel level should used a ternary value to express a positive or negative annotation, and also an absence of information for this pixel. Unfortunately, only binary mask are used. Hence, we make the hypothesis that the values "1" are positive samples, but values "0" are either negative or unknown. However, when using BCE, during training all "0" samples are considered as negative even if some in reality belong to the class. In practice, we have noticed that this partial annotation produces unsatisfactory results (see Fig. 4 (c)). Hence we propose the following solutions:

3.2.1 *Ignoring zero values.* This consists simply to ignore "0" value in class mask. Hence we modify the loss function in equation (2) so that only positive examples are considered during the training:

$$L_{PBCE}(y, \hat{y}) = -(y \cdot \log(\hat{y}) + (1 - y) \cdot \log(1 - \hat{y})) \cdot y \quad (2)$$

The pixel-wise Positive only BCE (PBCE) loss is the BCE value multiplied by the value of the class mask at this pixel. As a consequence, we remove from the loss computation all prediction from pixel corresponding to a "0" on the mask.

3.2.2 *Class masks fusion.* In this case, we consider annotations from all masks. We make the hypothesis that each annotated pixel can belong to only one class. Hence, for a given class mask, we propose to exploit information from other masks in this way: because of the previous hypothesis, for one "0" value in the class mask which is ambiguous, if this pixel is "1" in another mask, then it is no more ambiguous and can confidently be considered as a negative sample.

$$L_{MBCE}(y, \hat{y}, m) = -(y \cdot \log(\hat{y}) + (1 - y) \cdot \log(1 - \hat{y})) \cdot m \quad (3)$$

$$m = m_1 + m_2 + \dots + m_n \quad (4)$$

The Mask BCE (MBCE) have a new parameter m which is "1" if the annotation value for this pixel is "1" in one of the n class mask (see equation (4)).

3.3 Adding annotation from non-expert

Another way to disambiguate "0" value on masks is to ask a non-expert to manually localize areas where one can be sure no pixel belong to any class. In practice, in this work, there are only two classes: gum and teeth. Annotator have to set a box that visually include both gum and teeth. Hence, we are sure that outside this box, no pixel can belong to either gum or teeth class. In this way, we incorporate annotation from non-expert during the training.

Technically, the bounding box is used to automatically create negative samples for the loss calculation. Indeed, areas outside the bounding boxes can be considered as irrelevant areas and consequently as negative samples. This detector was used to create a new mask: areas outside of the bounding box are coded as "1" and areas inside as "0", noted as b in equation (5). So, the prior annotation masks m_i are fused with bounding box mask b . Then, the loss result is multiplied by the fused masked as in the equation (3) with this new m value:

$$m = m_1 + m_2 + \dots + m_n + b \quad (5)$$



Fig. 1. Example of bounding box annotation encompassing teeth and attached gingiva (gum)

4 EXPERIMENTS

4.1 Datasets

Two datasets freely available with partial segmentation annotations has been used: Oral and Dental Spectral Image Database (ODSI-DB) [9] and Labial Teeth and Gingiva Image Database (LTG-IDB) [4]. For the LTG-IDB, the annotation of gum and teeth is available for 45 images. For the ODSI-DB, only photos taken from the front with an annotation containing gum (Attached or Marginal gingiva) and teeth are used, corresponding to 50 images. The annotation is incomplete: some areas are completely ignored (e.g., areas at the back of the mouth), others are only partially annotated (e.g., the gum near the teeth is not annotated).

One example of image and related annotations from each database are presented in Fig. 2.

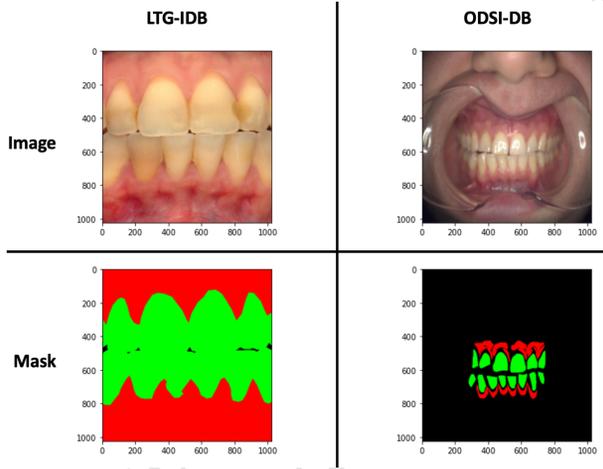


Fig. 2. Examples of images and annotations of ODSI-DB and LTG-IDB

4.2 Hyper-parameters

The following hyper-parameters are used to train the networks during the experiments:

- Image size: 512 pixels
- Batch size: 8
- Data augmentation:
 - Random rotation (+/- 30 %)
 - Random horizontal flip (50% of the time)
 - Random Color Jitter: Brightness (+/- 20 %), Contrast (+/- 20 %), Saturation (+/- 20 %)
 - Random Blur: $\sigma_{min}=0.01$, $\sigma_{max}=0.2$, kernel size=5
 - Normalization¹: $Mean_{Red,Green,Blue} = (0.57, 0.39, 0.25)$ $SD_{Red,Green,Blue} = (0.25, 0.21, 0.21)$.
- Data split: 80% train and 20% test
- Learning rate: .01
- Number of epochs: 50
- Number of repetitions: 5
- Optimizer: Stochastic gradient descent with learning rate=.01, momentum=0.9, and weight decay=0.0001

All the experiments were carried out on a NVIDIA A6000 48 GB.

¹Parameters calculated on ODSI-DB and LTG-IDB datasets.

4.3 Annotation from non-expert

To reduce the annotation labor for the non-expert, we fine-tuned a pretrained object detection model (Faster R-CNN [15]) to detect gum and teeth. For this purpose, 100 photos coming from an internal database have been annotated by a non-expert using label studio [19]. The bounding boxes have been drawn to encompass the teeth and attached gingiva. An example of annotation is presented in Fig. 1. The manual annotation of all images takes only 1h.

Then, the fine-tuned model was used to generate bounding boxes only on ODSI-DB (not used for LTG-IDB dataset as the images contains only teeth and gum).

4.4 Illustration of the process

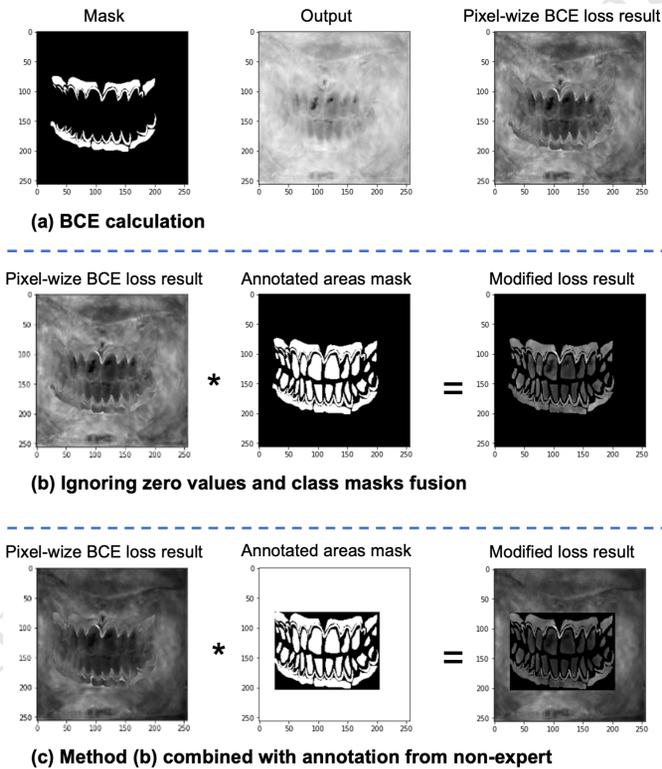


Fig. 3. Illustration of loss calculation for the "Gum" class

The proposed approaches on loss calculation are illustrated in the Fig. 3. The Fig. 3 (a) describes the standard BCE pixel-wise loss. The output is the visual result of the network predicting for each pixel its belonging to the class "Gum". Strong probability of a pixel to belong to this class is shown as a white value. The last image is the loss result.

In the Fig. 3 (b), the pixel-wise loss result is multiply by the annotation masks (i.e., fusion of gum and teeth masks, equation (4)) to take into account only annotated areas in the loss computation.

In the Fig. 3 (c), the pixel-wise loss result is multiply by the fusion of annotation masks (equation (5)) and the annotation from non-expert to remove areas with uncertain information in the loss computation.

5 RESULTS

As a baseline, we experiment a segmentation model of gum and teeth using a standard pixel-wise BCE.

For each experiment, the segmentation performance for the gum class and for the teeth class has been measured using a Dice score. The mean and standard deviation of the dice scores over the 5 repetitions are presented in the Table 1 (a) and an example of predicted masks for each experiment is presented in Fig. 4.

The combination of the modified loss with non-expert annotation showed similar results compared to the standard BCE. Nevertheless, these metrics cannot properly evaluate the segmentation performance because it can only be calculated from available annotations. This approach does not allow to calculate the performance on un-annotated areas.

Even if the results presented in Fig. 4 showed a qualitative improvement related to the proposed approaches, the experiments presented above do not allow them to be properly assessed.

To overcome this limit, new experiments were run where 25% of annotation masks have been randomly masked during the training in order to imitate partial annotation. For this purpose, the masks are divided in 5 areas of the same size (i.e., top left corner, top right corner, bottom left corner, bottom right corner, and the center). Then, one of the five areas is randomly masked (i.e., the selected values in the mask are set to 0). In this way, it is possible to assess quantitatively the difference between approaches on annotated areas. The results of these experiments are presented in Table 1 (b).

The combination of non-expert annotation and class masks fusion allows an improvement of 33% on Gum segmentation, an improvement of 18% on Teeth segmentation.

We also evaluate the influence of pre-training using ImageNet on the best approach. Initializing the network from weights pre-trained on ImageNet improves performance by 9% for the gum class and 6% for the teeth class compared to training from scratch.

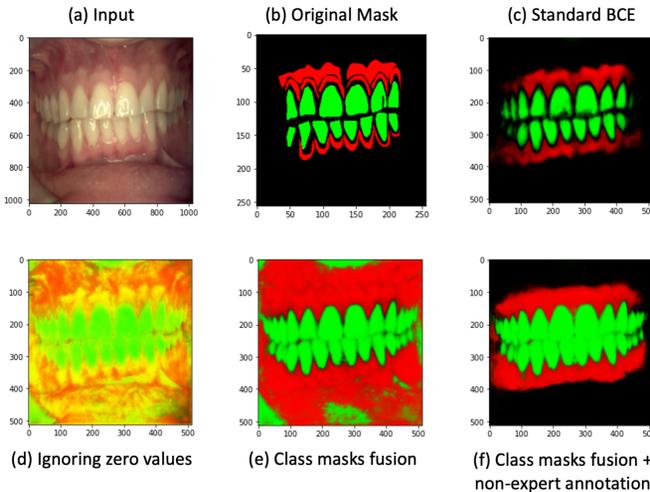


Fig. 4. Experiment results. Green values correspond to the teeth, Red values correspond to the gum and Black values correspond to neither

Table 1. Dice score for each experiment

Approach	(a) Original masks		(b) Masked masks	
	Dice - Gum	Dice - Teeth	Dice - Gum	Dice - Teeth
(1) Standard BCE (baseline)	.75 (.01)	.79 (.02)	.60 (.02)	.58 (.02)
(2) Ignoring zero values	.44 (.01)	.46 (.01)	.44 (.00)	.46 (.00)
(3) Class masks fusion	.53 (.00)	.60 (.01)	.53 (.00)	.60 (.01)
(4) Non-expert annotation	.75 (.01)	.79 (.01)	.59 (.02)	.58 (.02)
(5) Class masks fusion + non-expert annotation	.71 (.01)	.78 (.01)	.71 (.01)	.77 (.01)
(6) Class masks fusion + non-expert annotation trained from scratch	.67 (.02)	.73 (.01)	.65 (.01)	.73 (.01)

6 DISCUSSION

In this paper, we explored the segmentation of dental images in context of partially annotated data. For this purpose, some changes in the loss function was explored as well as the exploitation of non-expert annotation. The combination of these approaches showed a large improvement of the segmentation performance. Nevertheless, exploiting only the annotated areas without annotation from non-expert does not improve the Dice score. Indeed, even if we tried to make a fair quantitative evaluation of the proposed approaches, the evaluation is limited to the quality of original masks. As presented in Fig. 4, the proposed approaches (e) and (f) allows richer annotation than original masks (e.g., the teeth are fully segmented compared to (b)). But, the current evaluation cannot account for these contributions.

Some limitations related to the data must be specified. The available datasets contain mainly patients with no dental disease or limited diseases. The proposed approaches need to be evaluated on new patients with potential more important dental problems (e.g., tooth loss, dental prosthesis). Moreover, in some ways, the quality of images in the datasets is limited. For examples, some pictures have brightness issues increasing the difficulty in the segmentation. Data augmentation was applied to improve the network capabilities on these problems.

In the future, the concept could be extended with active learning principle (e.g., [13]). In this context, the annotation will be iteratively performed: (1) the expert will annotated some areas of the images, (2) a network will be trained with our method to segment using the partially annotated images, (3) the expert will manually correct the predictions, (4) the process will be repeated until a satisfactory result is obtained.

7 CONCLUSION

Building a good medical segmentation tool using machine/deep learning is limited to the availability of annotated images. When these images exist they are very few and they are not fully annotated, mainly because annotation has to be done by medical specialist and is very time consuming. In this paper, we have proposed several solutions to overcome this lack of exhaustive annotation. The

best solution seems the integrating of external explicit information which is doable by non-medical specialists. Hence, this paper paves the way for facilitate the segmentation of medical images, especially in the dental field which has been rarely explored. Future works will aim to extend this approach from partially annotated data using active learning methods.

ACKNOWLEDGMENTS

This work has been partially supported by MIAI@Grenoble Alpes (ANR-19-P3IA-0003) and the SATT Linksiium.

REFERENCES

- [1] John-Melle Bokhorst, Hans Pinckaers, Peter van Zwam, Iris Nagtegaal, Jeroen van der Laak, and Francesco Ciompi. 2019. Learning from sparsely annotated data for semantic segmentation in histopathology images. In *Proceedings of The 2nd International Conference on Medical Imaging with Deep Learning*. PMLR, 84–91. <https://proceedings.mlr.press/v102/bokhorst19a.html> ISSN: 2640-3498.
- [2] Yigit B. Can, Krishna Chaitanya, Basil Mustafa, Lisa M. Koch, Ender Konukoglu, and Christian F. Baumgartner. 2018. Learning to Segment Medical Images with Scribble-Supervision Alone. (2018). <https://doi.org/10.48550/ARXIV.1807.04668> Publisher: arXiv Version Number: 1.
- [3] Jifeng Dai, Kaiming He, and Jian Sun. 2015. BoxSup: Exploiting Bounding Boxes to Supervise Convolutional Networks for Semantic Segmentation. *arXiv:1503.01640 [cs]* (May 2015). <http://arxiv.org/abs/1503.01640> arXiv: 1503.01640.
- [4] Timo Eckhard, E. Valero, and Juan Nieves. 2012. Labial teeth and gingiva color image segmentation for gingival health-state assessment. In *Conference on Colour in Graphics*.
- [5] Mingchen Gao, Ziyue Xu, Le Lu, Aaron Wu, Isabella Noguees, Ronald M. Summers, and Daniel J. Mollura. 2016. Segmentation label propagation using deep convolutional neural networks and dense conditional random field. In *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*. 1265–1268. <https://doi.org/10.1109/ISBI.2016.7493497> ISSN: 1945-8452.
- [6] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. 2018. Mask R-CNN. *arXiv:1703.06870 [cs]* (Jan. 2018). <http://arxiv.org/abs/1703.06870> arXiv: 1703.06870.
- [7] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2015. Deep Residual Learning for Image Recognition. *arXiv:1512.03385 [cs]* (Dec. 2015). arXiv: 1512.03385.
- [8] Mohammad Hesam Hesamian, Wenjing Jia, Xiangjian He, and Paul Kennedy. 2019. Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges. *Journal of Digital Imaging* 32, 4 (Aug. 2019), 582–596. <https://doi.org/10.1007/s10278-019-00227-x>
- [9] Joni Hyttinen, Pauli Fält, Heli Jäsberg, Arja Kullaa, and Markku Hauta-Kasari. 2020. Oral and Dental Spectral Image Database—ODSI-DB. *Applied Sciences* 10, 20 (Oct. 2020), 7246. <https://doi.org/10.3390/app10207246>
- [10] Sergey Ioffe and Christian Szegedy. 2015. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *arXiv:1502.03167 [cs]* (Feb. 2015). arXiv: 1502.03167.
- [11] Shruti Jadon. 2020. A survey of loss functions for semantic segmentation. In *2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*. IEEE, 1–7.
- [12] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen A. W. M. van der Laak, Bram van Ginneken, and Clara I. Sánchez. 2017. A Survey on Deep Learning in Medical Image Analysis. *Medical Image Analysis* 42 (Dec. 2017), 60–88. <https://doi.org/10.1016/j.media.2017.07.005> arXiv: 1702.05747.
- [13] Brendon Lutnick, Brandon Ginley, Darshana Govind, Sean D. McGarry, Peter S. LaViolette, Rabi Yacoub, Sanjay Jain, John E. Tomaszewski, Kuang-Yu Jen, and Pinaki Sarder. 2019. An integrated iterative annotation technique for easing neural network training in medical image analysis. *Nature Machine Intelligence* 1, 2 (Feb. 2019), 112–119. <https://doi.org/10.1038/s42256-019-0018-3> Number: 2 Publisher: Nature Publishing Group.
- [14] Shervin Minaee, Yuri Y. Boykov, Fatih Porikli, Antonio J Plaza, Nasser Kehtarnavaz, and Demetri Terzopoulos. 2021. Image Segmentation Using Deep Learning: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2021), 1–1. <https://doi.org/10.1109/TPAMI.2021.3059968> Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence.
- [15] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. 2016. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *arXiv:1506.01497 [cs]* (Jan. 2016). <http://arxiv.org/abs/1506.01497> arXiv: 1506.01497.
- [16] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv:1505.04597 [cs]* (May 2015). <http://arxiv.org/abs/1505.04597> arXiv: 1505.04597.

- [17] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. 2015. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)* 115, 3 (2015), 211–252. <https://doi.org/10.1007/s11263-015-0816-y>
- [18] Nima Tajbakhsh, Laura Jeyaseelan, Qian Li, Jeffrey N. Chiang, Zhihao Wu, and Xiaowei Ding. 2020. Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation. *Medical Image Analysis* 63 (July 2020), 101693. <https://doi.org/10.1016/j.media.2020.101693>
- [19] Maxim Tkachenko, Mikhail Malyuk, Nikita Shevchenko, Andrey Holmanyuk, and Nikolai Liubimov. 2020-2021. Label Studio: Data labeling software. Open source software available from <https://github.com/heartexlabs/label-studio>.
- [20] Özgün Çiçek, Ahmed Abdulkadir, Soeren S. Lienkamp, Thomas Brox, and Olaf Ronneberger. 2016. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. *arXiv:1606.06650 [cs]* (June 2016). <http://arxiv.org/abs/1606.06650> arXiv: 1606.06650.

Unpublished working draft.
Not for distribution.