Imperial College London

INDIVIDUAL PROJECT INTERIM REPORT

IMPERIAL COLLEGE LONDON

DEPARTMENT OF COMPUTING

Tackling Crohn's disease using deep learning

Author:
Feifan Fan

Supervisor:
Dr. Bernhard Kainz

Second Marker:
Dr. Ben Glocker

March 6, 2023

Submitted in partial fulfillment of the requirements for the MEng Mathematics and Computer Science of Imperial College London

Contents

Chapter 1

Introduction

1.1 Crohn's Disease

Crohn's Disease [1, 2] is known as one of the two main types of Inflammatory Bowel Disease (IBD) [3]. Patients with IBD will suffer from severe and chronic inflammation of the gastrointestinal (GI) tract.

The main concern raised from such symptoms is that chronic and long-term inflammation can lead to damage in the GI tract, along with further possible complications, including bowel cancer. Recent research from the University of Nottingham [4] states that more than half a million people in the UK live with Crohn's Disease and Ulcerative colitis, another primary type of IBD.

However, unlike Ulcerative Colitis, which only affects the colon and rectum, Crohn's disease can develop lesions in any part of the GI tract, from the oesophagus to the anus. This fact poses two main problems. Since inflammation can occur in any part of the GI tract, the patient's digestive process can be affected. Also, due to the diversity of affected locations, the symptom of this disease will vary from patient to patient, including abdominal pain, diarrhoea, fatigue, and weight loss.

1.2 Clinical Challenges

Although the medical community has invested much research into Crohn's disease [5, 6], our clinical experts have not identified the exact cause of Crohn's disease yet. Therefore Crohn's disease is not curable today.

Fortunately enough, early diagnosis coupled with appropriate treatment from clinical professionals can significantly reduce the patient's suffering and improve the quality of their lives. Clinicians typically make a specific diagnosis through various approaches such as enteroclysis, endoscopy, colonoscopy, and radiographic diagnosis, including barium contrast X-rays, Computed Tomography (CT) and Magnetic Resonance Imaging (MRI). The MRI diagnosis has become popular since it is a non-invasive approach and produces more detailed images than its alternatives, such as computerised tomography (CT).

The examiners will look particularly into the terminal ileum, and right colon, where the areas are most involved in Crohn's disease, for symptoms such as bowel wall thickening, small bowel strictures, and mural hyperenhancement [7] to confirm the existence of Crohn's disease. Nevertheless, human experts still need to go through the MRI scans slice by slice, which is time-consuming and inefficient to diagnose one patient.

1.3 Motivation

The prevalence of Machine Learning and Deep Learning, especially convolutional neural networks (CNN), allows us to automatically learn the features from input imaging data and help the human expert with diagnosis. Given that the terminal ileum is essential in Crohn's Disease diagnosis, in 2019, Holland et al. [8] proposed a residual network focusing on the terminal ileum to detect Crohn's Disease from MRI scans automatically. They concluded that the framework's performance depends on the level of localisation in preprocessing. Consequently, they recommend that terminal ileal ground-truth segmentations also be collected for localising the terminal ileum and enhancing the performance of automated detection.

A follow-up study by Abidi et al. [9] in 2022 developed an enhanced deep learning tool using a nnU-Net architecture to automatically locate critical areas (particularly terminal ileum) where radiologists typically perform examinations to make a diagnosis. This study resolves the high dependence on the level of localisation in preprocessing in [8], and the model significantly outperforms the results of the prior work done by Xu et al. [10] in 2021 for terminal ileum segmentation. Furthermore, [9] establishes a strong foundation for a multi-class terminal ileum segmentation algorithm.

Inspired by studies of [8, 10, 9], this individual project aims to develop a multi-class segmentation model to differentiate normal and abnormal terminal ileum regions, and the clinicians will be significantly aided in the diagnosis of Crohn's disease with the success of this project.

1.4 Machine learning Challeges

The performance of a deep learning model is strongly related to the training data's quality, quantity and diversity. Unfortunately, our biggest challenge in this project is the scarcity of training data. Our training dataset is relatively small (only 233 patient cases available) compared with other industry-leading deep learning systems. Another fact we need to face is that developing gold-standard labels or point-wise centerlines for patient data requires manual segmentations by clinical experts, which is also a complex, time-consuming and inefficient process. Therefore, the acquisition of patient data with great quality and quantity is indeed a challenging process.

To alleviate these problems, we introduce a proxy task with weak supervision before the target training task to get comparable results with learning a large number of data with weak labels that are less spatially informative. However, from prior work [9], it is suggested that training from scratch using the nnU-Net framework in the proxy task is inefficient since it takes a long time to converge and exhibits unstable performance. Applying transfer learning to proxy learning [11] or adding a related pre-training task may resolve this situation.

1.5 Objectives

The primary objectives of our work include the following:

1. Replicating the results of the prior work [9] for segmenting the terminal ileum from MRI scans.

2. Developing a multi-class segmentation model of the terminal ileum based on the prior work and employing the trained model as our baseline model.

3. Developing a sophisticated, related pre-training task for the multi-class terminal ileum segmentation that leverages the industrial leading transfer learning techniques to imporve the performance of the proxy model and associated stability.

4. Improving the results of the baseline segmentation model and developing a user-friendly GUI as an extension to this project.

Chapter 2

Related Work

2.1 Automatic Detection and Segmentation of Crohn's Disease Tissues from Abdominal MRI

We need to go back to a decade ago to see the initial accomplishments of applying deep learning techniques to medical image analysis. In 2013, Mahapara et al. [12] first proposed a machine learning-based method for segmenting parts of the bowel for detecting Crohn's disease tissues on MRI scans.

The pipeline firstly oversegmented the given MR image test volume into supervoxels and applied Random Forest (RF) classifiers to classify supervoxels containing diseased tissues, where these supervoxels define the Volume of Interest (VOI). Another set of RF classifiers is used on the test volume to generate a probability map, which indicates the probability of each voxel being diseased tissue, normal tissue or background.

The experimental results showed that the pipeline achieved a decent segmentation performance, with a dice metric value of 0.90 ± 0.04 and a Hausdorff distance of 7.3 ± 0.8 mm. However, this approach is computationally expensive and inefficient since each instance takes a long time for testing, each pipeline stage requires fine-tuning, and the architecture leaves lots of room for improvement.

2.2 Automatic Detection of Bowel Disease with Residual Networks

Holland et al. [8] proposed an end-to-end residual network [13] with a soft attention layer [14] to automate the detection of Crohn's disease. The detection is not based on any of the semantic segmentation work but only seeking for the terminal ileum. Consequently, they revealed the feasibility of using a deep learning algorithm for identifying terminal ileum Crohn's Disease from abdominal MRI.

From experimental results, the best result with localised data and a semi-automatic setting achieved a weighted-f1 score 0.83, which is strongly correlated with the MaRIA [15] score, the current clinical standard. Also, the performance is slightly inefficient when a fully automatic invariant is included.

They suggested that the performance of the model is highly dependent on the level of localisation in the training data and argued that gathering gold-standard segmentations of terminal ilium will be beneficial as a predecessor task if a better performance of the automatic detection needs to achieve.

2.3 Machine learning methods to predict the presence of intestine damage in patients with Crohn's disease

Enchakalody et al. [16] proposed two classifiers for aiding improvements of the reliability of Crohn's Disease in diagnosis and monitoring by looking into cross-sectional views of a small-intestine segment and detecting diseased issues. The classifiers are implemented based on Random Forest (RF) with ensembling and Convolutional Neural Network (CNN).

From the experimental result, both techniques can differentiate diseased and normal small bowel tissue with similar performance, even compared to the case with expert radiologists present. However, this study mainly focused on CT-enterography data, which slightly deviated from our cases with MRI data. And until the day of writing this report, there is still no fully automatic approach for diagnosing Crohn's disease on cross-sectional imaging that reaches the level of expert radiologists.

Chapter 3

Background

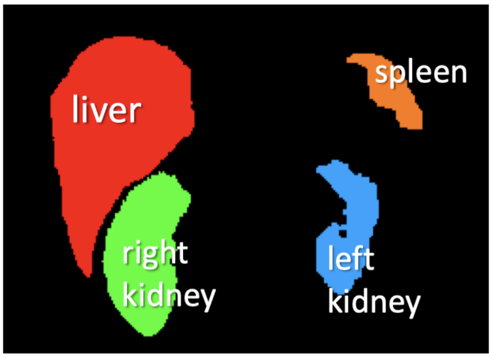
3.1 Image Segmentation

When we talk about "Image Segmentation", what exactly are we talking about? Suppose we have a picture of two birds; the segmentation task, or more precisely the **"semantic segmentation"** task, is to separate the whole image into different regions with different colour codes, where the regions represent the exact position of birds covered on the image and the background.

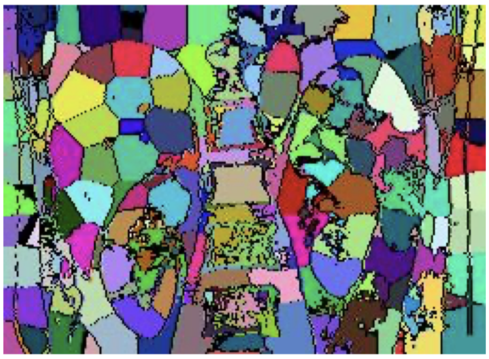
3.1.1 More on Semantic Segmentation

The word "semantic" means we always assign a word label (a **semantic meaning**) to the segmented region, representing what this region refers to. For example, for the example we mentioned above, the regions we segmented for the bird will be assigned with the label "bird", and the background region will be labelled as "background".

Note. Semantic segmentation is different from the tasks based on "pure" clustering of images to coherent regions. The regions we are looking to have their value or semantic meaning from the task we defined.



(a) Semantic Segementation



(b) Segmentation based on clustering

Figure 3.1: Different Image Segmentation Tasks [17]

After getting a glance at **Semantic Segmentation**, we will discuss the methods for segmentation tasks in the following sections.

3.2 Manual Segmentation

Manual Segmentation must be the most straightforward method we can come up to our mind. It refers to a process that lets human experts as an observer make gold standard labels based on their background knowledge. It is clear to see that the segmented labels are confirmed to be validated. However, it might take a long time for human observers to create such labels due to the time cost of training and the amount of data required to be labelled.

Furthermore, inter or intra-observer variability may bring a concern to the validity of the labelled data by manual segmentation since there might be disagreement between different human observers and/or disagreement from one observer but on different occasions. Also, the bias from human experts might cause problems. We still need the assistance of computers to utilise intelligence for performing the segmentation task.

3.3 Threshold-based Methods

Thresholding is one of the simplest and most popular segmentation techniques used in the industry. This method is based on the assumption that intensity distribution has multiple modes. That is to say, the grayscaled intensity of the pixels can be divided into two (or more) clusters, and we can pick thresholds from the "gaps" between the clusters to segment the background and foreground. The segmentation can be achieved by setting the pixels with an intensity lower than the threshold background and those with an intensity equal to or greater than the threshold foreground. It is also accepted to select multiple thresholds to select pixels within a specific range only for segmentation.

The algorithm is simple and fast, yet the algorithm can only be applied when regions (of interest) are homogeneous and distinct. That is to say, the pixels in the same region should have similar intensity values. Also, finding consistent threshold frequencies across different images is different, since the intensity of pixels will vary from image to image.

3.4 Region-based Methods

The region-growing method is a typical region-based method. The algorithm is based on the assumption that the segmented regions are homogeneous. The algorithm starts with a seed pixel, and then it will grow the region by adding pixels similar to the seed pixel. The algorithm will stop when the region reaches a specific size or when the region is homogeneous, where interpreted as the neighbouring pixels are too dissimilar with each other in comparison [18].

Applying the region-growing method is usually very efficient, yielding a connected region from the seed point. Unlike thresholding, region-growing methods achieve segmentation without the explicit use of image properties. However, the region-growing method is not robust to noise since the algorithm will keep growing the region even if the neighbouring pixels are not similar to the seed pixel.

Another major issue that the region-growing method has is that the initial seed point selection is significant. If the seed point is not selected correctly, the algorithm will not be able to grow the region properly. Consequently, this initial step is time-consuming and inaccurate to find the optimal seed point, and it also requires human intervention to evaluate the choice of seed point.

To address this problem in medical imaging, Poonguzhali et al. proposed an automatic region growing method for ultrasound images [19]. They adopt an automatic seed point selection based on textural features such as co-occurrence features and run length features without human intervention. The experimental results showed the feasibility and effectiveness of the proposed region-growing algorithm in selecting seed points and segmenting the ROI without manual intervention.

3.5 Deep Learning Methods

U-Net [20] is a deep learning method for image segmentation. It is a fully convolutional network (FCN) [21] that uses a contracting path to capture image context and a symmetric expanding path that enables a precise level of localization. The network is trained end-to-end from raw pixels to pixel-wise segmentation masks. The network architecture is shown in Figure 3.2.

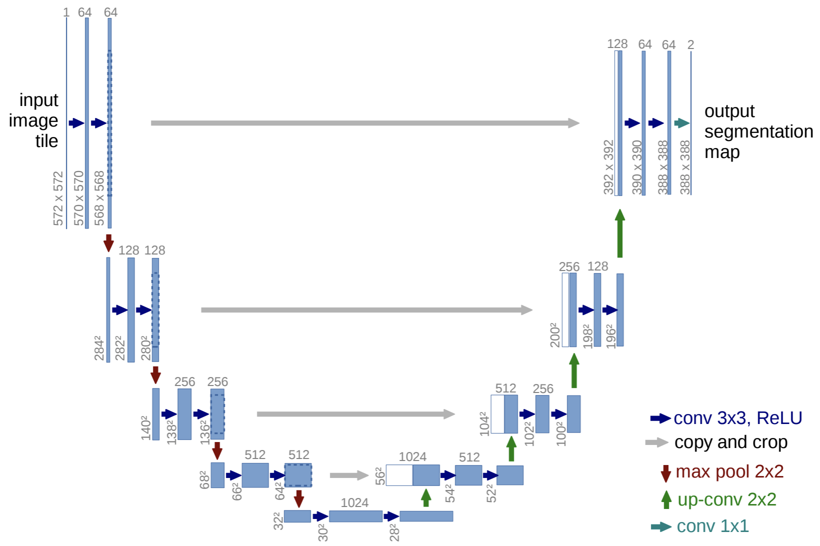


Figure 3.2: An example of the U-net architecture. Each blue box in the figure represents a multi-channel feature map, and on the top of each box indicates the number of channels. The x-y-size is shown at the lower left edge of the box. Each white box represents a copied feature maps. The arrows denote the different operations. [20]

In the contracting path of the network (on the left), the network consists of a series of repeated applications. Each step starts with two 3x3 convolutions, and each convolution doubles the number of feature maps. Then the network applies a rectified linear unit (ReLU) as activation, followed by a downsampling step via a 2x2 max pooling operation with a stride of 2.

The expansive path (on the right) starts with an upsampling of the feature map followed by a 2x2 convolution, halving the number of feature channels. The feature maps are then concatenated with the corresponding crops of the contracting path. After that, two 3x3 convolutions are applied to the image, and a ReLU activation follows each convolution. The reason for cropping the convolution results from the contraction path is the loss of border pixels in every convolution. At the final layer, the network used a 1x1 convolution to map each 64-component feature vector to the desired number of classes.

nnU-Net [22], a framework based on U-Nets, is a self-configuring segmentation framework that can automates the configuration of the preprocessing, network architecture, training, and post-processing of a segmentation pipeline. The selected configuration is depending on the medical data used for training. The framework provides a fully automated segmentation workflow that can be trained and inferred on any medical dataset.

For deterimining hyperparameters, nnu-Net uses "data-fingerprints" with heuristic rules to select the best hyperparameter configuration for a given dataset before processing the training data. The framework also uses data fingerprints to generate pipeline configurations including the inferred parameters (image resampling, normalisation, batch, and patch size), and the blueprint parameters (loss function, optimiser, architecture).

Combined with pre-selected hyperparameters and the pipeline configurations, the pipeline generates network training for 2D, 3D full-resolution, and 3D-Cascade U-Nets [23]. The optimal performance (e.g. average dice coefficient) is obtained by ensemling the configurations of the three networks. After that the optimal configuration will be deployed and evaluated on test dataset. The pipeline architecture is shown below.

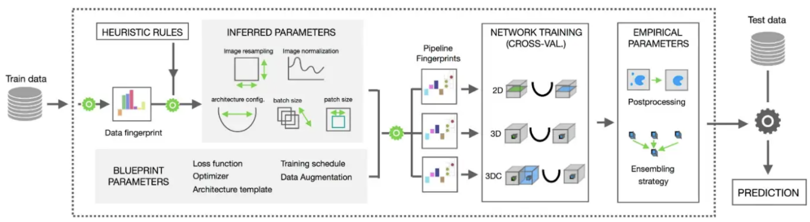


Figure 3.3: The pipeline representation of nn-UNet. [24]

The nnU-Net has achieved the state-of-the-art performance on a variety of medical imaging tasks, including brain tumour segmentation, prostate segmentation, and liver segmentation [25].

Chapter 4

Project Plan

For the project plan, I would like to divide the project into 4 phases:

- Phase 1: Exploratory Data Analysis of the Crohn's Data and Replicate the results for the prior work.
  - I will use MRI volumes, ground-truth segmentation masks and centerline coordinates for this project. Ground-truth labels for the multi-class segmentation need further discussion with the project supervisor.
  - For replicating the prior work, a further discussion about using available GPU resources is required.
  - In this phase, a checkpoint meeting/email is required every two weeks. The deadline for this phase is the end of February.
- Phase 2: Develop the multi-class segmentation model based on the prior work
  - In this phase, a checkpoint meeting/email is required every two weeks. A catch-up meeting could be rearranged for clashes with exams at the end of March.
  - The student is expected to have some deliverable progress by mid-April.
- Phase 3: Update the pre-training task for the proxy task
  - The student is expected to have some deliverable progress by mid-may.
- Phase 4: GUI implementation and final report

- The student is expected to have some deliverable progress by ten days of the project deadline.

Chapter 5

Evaluation Plan

For the model evaluation I would like to have the following steps:

1. Develop the multi-class segmentation model, and set the trained model as the initial baseline model.
2. Adapt the new pre-training task to the proxy task and compare the performance with the target model with new proxy task with the baseline model.
3. Compare the performance of the new proxy model and the proxy model implemented with methods in the prior work

Chapter 6

Ethical Discussion

In this project, ethics were carefully considered. The project involves further processing of the previously collected medical data, including image registration and data augmentation, and merging of the existing datasets as well.

However, all MRI scans were processed and approved by clinical radiologists at St Mark Hospital. All sensitive personal data, such as names, genders, ages and ID numbers, were removed. Consequently, patients' identities can not be traced back from the medical data.

For the model-training procedure, since we are using the nnU-Net framework with convolutional layers involved, the model learns from data by extracting features. The model will not store original training data but create feature maps in this case. In other words, the training data cannot be recovered from the model, so personal data privacy is preserved.

Bibliography

[1] Daniel C Baumgart and William J Sandborn. Crohn's disease. *The Lancet*, 380(9853):1590–1605, 2012.

[2] NHS. Crohn's disease - nhs. URL https://www.nhs.uk/conditions/crohns-disease/.

[3] Centers for Disease Control and Prevention. What is inflammatory bowel disease? (ibd) | ibd. URL https://www.cdc.gov/ibd/what-is-IBD.htm.

[4] University of Nottingham. Rates of crohn's and colitis have been vastly underestimated for decades, says new study. URL https://www.nottingham.ac.uk/news/rates-of-crohns-and-colitis-have-been-vastly-underestimated-for-decades-says-new-study.

[5] Gautier Hoarau, PK Mukherjee, C Gower-Rousseau, C Hager, J Chandra, MA Retuerto, Christel Neut, Séverine Vermeire, J Clemente, Jean-Frederic Colombel, et al. Bacteriome and mycobiome interactions underscore microbial dysbiosis in familial crohn's disease. *MBio*, 7(5):e01250–16, 2016.

[6] Joseph D Feuerstein, Edith Y Ho, Eugenia Shmidt, Harminder Singh, Yngve Falck-Ytter, Shanaz Sultan, Jonathan P Terdiman, Shahnaz Sultan, Benjamin L Cohen, Karen Chachu, et al. Aga clinical practice guidelines on the medical management of moderate to severe luminal and perianal fistulizing crohn's disease. *Gastroenterology*, 160(7): 2496–2508, 2021.

[7] David H Bruining, Ellen M Zimmermann, Edward V Loftus Jr, William J Sandborn, Cary G Sauer, Scott A Strong, Mahmoud Al-Hawary, Sudha Anupindi, Mark E Baker, David Bruining, et al. Consensus recommendations for evaluation, interpretation, and utilization of computed tomography and magnetic resonance enterography

in patients with small bowel crohn's disease. *Gastroenterology*, 154(4): 1172–1194, 2018.

[8] Robert Holland, Uday Patel, Phillip Lung, Elisa Chotzoglou, and Bernhard Kainz. Automatic detection of bowel disease with residual networks. In *International Workshop on PRedictive Intelligence In MEdicine*, pages 151–159. Springer, 2019.

[9] Ali Abidi. Tackling crohn's disease using deep learning. Master's thesis, Imperial College London, 2022.

[10] Ke Xu. Tackling crohn's disease using deep learning. Master's thesis, Imperial College London, 2021.

[11] Jong-Hwan Jang, Tae Young Kim, and Dukyong Yoon. Effectiveness of transfer learning for deep learning-based electrocardiogram analysis. *Healthcare informatics research*, 27(1):19–28, 2021.

[12] Dwarikanath Mahapatra, Peter J Schüffler, Jeroen AW Tielbeek, Jesica C Makanyanga, Jaap Stoker, Stuart A Taylor, Franciscus M Vos, and Joachim M Buhmann. Automatic detection and segmentation of crohn's disease tissues from abdominal mri. *IEEE transactions on medical imaging*, 32(12):2332–2347, 2013.

[13] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[14] Jo Schlemper, Ozan Oktay, Michiel Schaap, Mattias Heinrich, Bernhard Kainz, Ben Glocker, and Daniel Rueckert. Attention gated networks: Learning to leverage salient regions in medical images. *Medical image analysis*, 53:197–207, 2019.

[15] Jordi Rimola, Sonia Rodríguez, Orlando García-Bosch, Ingrid Ordás, Edgar Ayala, Montserrat Aceituno, Maria Pellisé, Carmen Ayuso, Elena Ricart, Lluís Donoso, et al. Magnetic resonance for assessment of disease activity and severity in ileocolonic crohn's disease. *Gut*, 58 (8):1113–1120, 2009.

[16] Binu E Enchakalody, Brianna Henderson, Stewart C Wang, Grace L Su, Ashish P Wasnik, Mahmoud M Al-Hawary, and Ryan W Stidham. Machine learning methods to predict presence of intestine damage in patients with crohn's disease. In *Medical Imaging 2020: Computer-Aided Diagnosis*, volume 11314, pages 742–753. SPIE, 2020.

[17] Ben Glocker. Image segmentation - machine learning for imaging course, 2023.

[18] Rolf Adams and Leanne Bischof. Seeded region growing. *IEEE Transactions on pattern analysis and machine intelligence*, 16(6):641–647, 1994.

[19] S Poonguzhali and G Ravindran. A complete automatic region growing method for segmentation of masses on ultrasound images. In *2006 International Conference on Biomedical and Pharmaceutical Engineering*, pages 88–92. IEEE, 2006.

[20] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[21] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

[22] Fabian Isensee, Paul F Jaeger, Simon AA Kohl, Jens Petersen, and Klaus H Maier-Hein. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods*, 18 (2):203–211, 2021.

[23] Fabian Isensee, Jens Petersen, Andre Klein, David Zimmerer, Paul F Jaeger, Simon Kohl, Jakob Wasserthal, Gregor Koehler, Tobias Norajitra, Sebastian Wirkert, et al. nnu-net: Self-adapting framework for u-net-based medical image segmentation. *arXiv preprint arXiv:1809.10486*, 2018.

[24] Prateek Gupta. nnu-net : The no-new-unet for automatic segmentation, 2022. URL https://medium.com/miccai-educational-initiative/nnu-net-the-no-new-unet-for-automatic-segmentation-8d655f3f6d2a.

[25] Fabian Isensee, Jens Petersen, Simon AA Kohl, Paul F Jäger, and Klaus H Maier-Hein. nnu-net: Breaking the spell on successful medical image segmentation. *arXiv preprint arXiv:1904.08128*, 1(1-8):2, 2019.