Ministry of Education and Science of Ukraine

NATIONAL UNIVERSITY OF KYIV MOHYLA ACADEMY

Network Technologies Department of the Faculty of Computer Sciences

OPTIMIZING SEGMENTATION OF NEONATAL BRAIN MRI WITH PARTIALLY ANNOTATED MULTI-LABEL DATA

Bachelor thesis

in the specialty «Computer Sciences» 122

Co-supervisors:

Prof. Andriy M. HLYBOVETS

____

(signature)

Prof. Michael BRUDNO

____

(signature)

Authored by a 4th-year student

Dariia V. KUCHERUK

«____» __________ 2023yr.

Kyiv 2023

INDIVIDUAL TASK

Ministry of Education and Science of Ukraine
NATIONAL UNIVERSITY OF KYIV MOHYLA ACADEMY
Network Technologies Department of the Faculty of Computer Sciences

APPROVED
Head of the Department of
Network Technologies,
PhD, Professor
_____________ Gennadi I. Malaschonok
(signature)
«____»________________2023 yr.

INDIVIDUAL TASK

For Bachelor's Thesis

for the 4th year Faculty of Computer Sciences student Dariia V. Kucheruk

Topic: Optimizing Segmentation of Neonatal Brain MRI with Partially Annotated Multi-Label Data

Content of bachelor thesis:

Calendar plan
Contents
Abstract
Introduction
1. Automatic segmentation task overview.
2. Research of related work.
3. Development of a segmentation model trained on partially labeled data and results analysis.
Conclusion
References

Issue date «___» __________ 2023 yr. Supervisor ________________
(signature)
Task received___________
(signature)

CALENDAR PLAN

№	Stage name	Deadline
1.	Choosing the graduate thesis title	26.10.2022
2.	Receiving a task for thesis work	24.12.2022
3.	Researching the literature on solving the issue of partially annotated data in the training set for the model	26.10.2022-01.03.2023
4.	Developing a solution to use weakly annotated labels for training the automatic segmentation model	01.05.2023
5.	Running experiments comparing the original model to the proposed one	12.05.2023
6.	Working on the text part of the thesis	10.05.2023
7.	Final edits to the text and presentation preparation	20.05.2023
8.	Submitting the work to plagiarism check	22.05.2023

Contents

Abstract

This thesis proposes a multi-label segmentation model for optimizing the segmentation of neonatal brain MRI with partially annotated data. A multi-label segmentation model that addresses the challenges of limited annotated data by modifying the preprocessing, loss function, and postprocessing of the original multi-class label segmentation was developed. The proposed approach aims to improve the accuracy and efficiency of neonatal brain MRI segmentation by leveraging partially annotated data. We evaluate our method on a unique dataset of neonatal brain MRI and demonstrate its effectiveness compared to the models trained on fully annotated data.

Introduction

Clinicians all over the world use Magnetic Resonance Imaging (MRI) to perform medical examinations and conclude diagnoses. Having proved to be good at identifying soft tissue compared to other medical imaging techniques such as computed tomography (CT) **[1]**, MRIs provide high-precision imaging of relevant anatomy. Brain segmentation is crucial for providing accurate predictions to the patients due to the information which can be extracted from it, such as the shape and size of the tissue, underdeveloped or injured regions, and annual comparative results for brain growth.

Brain segmentation of infants' brains is more challenging than that of adults` brains. However, the peculiarity of the newborn brain is that the diversity of biological features changes drastically in the course of its development. Segmentation of infants' brains currently requires deep expertise and is a labor-intensive process, typically performed by an experienced radiologist, and typically only for research studies, and not for routine clinical care. Automatic segmentation methods would highly increase the number of infants with brain injuries for whom we would have segmentation data, enabling further studies of neurodevelopmental outcomes.

To address the issue an neonatal and child brain MRI segmentation pipeline utilizing U-Net-based automation was created by Pouria Mashouri et al **[2]**. In their work, they used a convolutional neural network architecture to segment 2D slices of MRIs and reached high accuracy in identifying the cerebrum, cerebellum, and brainstem. The U-NET-based model was chosen due to its diverse advantages, such as being capable of producing high-quality segmentation results whilst being trained on a small range of training samples. On top of that one of the very important considerations in building the segmentation model was to minimize the number of inputs. U-NET outdoes patch-based segmentation models in terms of preserving the full context of the input by having an end-to-end pipeline, which works through the whole MRI and outputs segmentation maps. This way it is only needed to provide the model with one input 3D MRI file to receive a 3D mask with the segmentation of various brain regions.

The model under consideration demonstrated exceptional performance on a uniform dataset; however, regrettably, it exhibited suboptimal outcomes when evaluated on heterogeneous test datasets from various healthcare facilities. One way to solve the problem of the model only being trained on a narrow set of data is to add more diverse data from different sources. However, this can be difficult in practice because many real-life datasets only have partial labels, and the original model is not able to handle these types of training conditions.

This work was supervised by Professor Andrii Hlybovets from the National University of Kyiv Mohyla Academy and Professor Michael Brudno from the University of Toronto. Additionally, thorough guidance was provided by Dr. Jessie Guo. Computational resources and training data were provided through collaboration with UHN and SickKids hospitals.

In this paper, we present a novel approach to improving the performance of the model by upgrading its architecture from a Multi-Class Segmentation Model to a Multi-Label Segmentation Model. The proposed upgrade involves modifications to the preprocessing and postprocessing steps, as well as the implementation of a new loss function in the original model. Our findings suggest that this upgraded model yields superior performance in the partially labeled training settings compared to its predecessor and can be applied effectively in a variety of applications.

1 AUTOMATIC SEGMENTATION TASK OVERVIEW

1.1 Introduction to Segmentation

Segmentation is a term that has many various definitions depending on the domains where it is used. In this work, our main interest falls under segmentation in the field of brain MRIs and the field of Semantic Segmentation in computer vision.

First of all, let`s introduce medical image segmentation. Medical image segmentation is a process that requires a medical professional to extract a particular organ from a medical image (MRI, CT, X-Ray, etc.) by hand or with the help of automatization tools **[3]**. There are two types of segmentation approaches: intensity-based and shape-based.

The intensity-based segmentation operates by making use of the fact that the gray value of voxels are quite similar within the borders of a particular organ. During manual segmentation, it can be difficult to identify different organs due to the small difference between the average value of voxels for every other organ. However, the use of computer vision approaches can help segment the regions of interest much more precisely.

There is also shape-based segmentation, which relies on the fact that the shape of the object of interest is known beforehand and therefore this information is used during segmentation. In the biomedical field, even the shapes of well-known organs can vary from patient to patient, which makes the use of shape-based segmentation not trivial.

In the field of computer vision, the term semantic segmentation stands for the task of categorizing each pixel or voxel into a class to create a segmentation map of a given input **[4]**. Nowadays, semantic segmentation is based on Artificial Intelligence approaches and is being used in many domains, such as healthcare, automatic driving systems, eCommerce, etc.

1.2 Automatic Segmentation in the Field of Healthcare

The field of healthcare is known throughout the world for its great responsibility to society and its large and constant workload. Unfortunately, because of the overload in hospitals sometimes situations when help arrives to patients later than it should occur. Every so often current approaches are not yet advanced enough to treat certain conditions.

Automatic segmentation's contribution to the healthcare industry relies on the improvement of the offered medical services in their speed and efficiency. For instance, modern healthcare workflows could become much more balanced with the use of AI tools for segmentation and outcome prediction models in a way that more people would get the best treatment possible whilst medical practitioners wouldn`t struggle with working extra shifts and hours due to the overload in hospitals that can be seen these days.

In addition to that, medical and AI researchers all over the world would highly benefit from the enlargement of the various brain imaging datasets, which would include MRIs with a vast range of brain ages, anatomical structures, levels and kinds of injuries, technical parameters, and many other characteristics to be explored.

Nowadays MRIs from neonates are usually being taken for research purposes only which limits the accessible data drastically. The wide usage of AI segmenting tools resulting in a major expansion of very needed data would help the worldwide society of scientists explore brain development as well as provide a much better understanding of brain development and thus the prediction of long-term outcomes.

2 RESEARCH OF RELATED WORK

2.1 Related work in the field of automating brain segmentation

The results of different approaches to automating brain segmentation were published previously **[5][6][7].** It is important to note the scientific breakthroughs which were made in the field with the help of machine learning techniques **[8]**. Despite their effectiveness in classifying adult brain tissue, classification accuracy for infant brain specimens needs improvement. There are also works where deep learning techniques were used to segment children's brains **[9][10][11]**, although they have limitations for certain age groups of patients and are not applicable to either pre-term or older (age 4+) children. Originating from **[12],** a segmentation model based on a convolutional neural network reached a competitive accuracy in labeling brain tissue, though the resulting technology was not tested on datasets of highly injured brains.

The task of semantic segmentation has been previously carried out within the context of multi-label training, employing convolutional neural networks in applications for map segmentation **[13]** and object detection in natural environments **[14]**. However, it is worth noting that the training sets employed in these experiments were fully labeled. Furthermore, the model's output was treated as Multi-Class in spite of this. There is a study **[15]** that integrated the previously mentioned methodologies, specifically the UNET architecture and multi-label approach. It is worth noting, however, that akin to the authors of the preceding paper, the authors of this study employed fully annotated datasets to train their model. Moreover, their training methodology diverged from our present investigation as it was conducted using 3D data for training, in contrast to our utilization of MRI 2D slices.

While the subject matter of multi-label semantic segmentation utilizing weakly annotated training sets remains largely unexplored, it is noteworthy that a significant amount of progress has been made in the realm of classification tasks under the partially labeled data paradigm. Despite the paucity of research on the former topic, it is vital to recognize the advancements that have been made in related fields, as they may provide invaluable insights and serve as a stepping stone toward the successful resolution of multi-label semantic segmentation challenges.

Regarding the matter at hand, there has been a scholarly paper **[16]** published on the subject of utilizing partially segmented data for training a model aimed at solving a classification task. The researchers addressed a significant challenge in the realm of single positive multi-label learning, which effectively mitigates the issue of false negative assumptions associated with unannotated labels.

A model **[17]** was developed to classify partially labeled data utilizing class-aware selective loss. This approach enabled the utilization of probabilistic distributions of specific labels across the data points, while also acknowledging the likelihood of particular labels for certain data inputs. Furthermore, a notable conference paper **[18]** was published, which provided a comprehensive overview of the current methodologies employed in addressing the challenge of Learning from Partially Labeled Data.

In summary, it is evident that the topic of training on partially labeled data has been the subject of numerous studies, spanning from semantic segmentation tasks to classification ones. This current work aims to amalgamate the strengths of both approaches and enhance the pre-existing segmentation U-NET model by empowering it to effectively handle weakly annotated training data.

2.2 Related work in the field of learning from partially labeled data.

During the research conducted in the development of the proposed model, it became apparent that there are a few existing approaches to learning from partially labeled data, such as:

a) Using separate models for each label;

b) Pseudo-labeling (Semi-supervised learning);

c) Multi-Label Learning from Single Positive Labels.

The "separate models" approach had been implemented previously **[19]**. However, the focus of the research was primarily on utilizing federated learning to learn from the data collected from various hospitals that maintain strict policies regarding sharing data with other healthcare facilities. Furthermore, the distributed architecture creates additional complexity for upgrading the model. For instance, the training labels employed in the proposed model for this thesis adhere to a hierarchical structure and employing a hierarchical constraint layer **[20]** could potentially enhance the predictions. That being said, this would not be a straightforward process in the case of federated learning.

We chose not to use the "separate models" approach in this work for another reason: overlapping label predictions. Effectively managing these overlaps would require consulting with medical specialists to establish a suitable threshold. It is necessary to ensure the accuracy and reliability of the model's output, but it can also add significant complexity to the development process. In this particular case, the process of determining an appropriate threshold was deemed too time-consuming and labor-intensive to pursue.

The pseudo-labeling approach is a widely used method for addressing weakly annotated training data in various domains, including audio classification **[21]** and environment detection on mobile phones **[22]**. This method allows the model to generate its training labels based on predictions. Although, relying solely on pseudo-labeling would mean disregarding valuable information from the partially labeled datasets. Additionally, implementing this method would require the assistance of medical specialists to establish thresholds for identifying valid and invalid pseudolabels. As a result, a more nuanced approach is necessary to balance the advantages and limitations of each level of access to the annotations.

On the contrary, Multi-Label Learning from Single Positive Labels **[23]** allows for the use of all datasets available during the study. This approach also eliminates the non-trivial task of finding thresholds to solve the overlapping issue, which was one of the main disadvantages of the pseudo-labeling approach. Additionally, it allows for upgrades in the form of additional layers on top of the monolithic architecture of the multi-label model.

Given the domain and settings provided for this thesis, a decision was made to favor a monolithic structure in the multi-label segmentation model. This approach was chosen to ensure the utilization of all available data through the use of a loss function that can effectively distinguish between annotated and unannotated labels. Such an approach holds tremendous potential for achieving optimal results while mitigating the risk of overlooking key insights in the data. Ultimately, this strategy enables the model to operate at peak performance, unlocking new possibilities for advancing the field and delivering significant value to the research community.

3 DEVELOPMENT OF A SEGMENTATION MODEL TRAINED ON PARTIALLY LABELED DATA AND RESULTS ANALYSIS

3.1 Problem Statement

The original model **[2]** was trained on 3D MRI scans of neonates. These scans were divided into sets of 2D slices so that the model could be trained on each 2D slice individually. Each pixel in a 2D scan was assigned a positive integer according to the Multi-Class paradigm, which indicated the presence of an annotated label and its corresponding class. However, this approach presented a few issues, which we aim to address by employing the Multi-Label paradigm.

One of the primary issues with the Multi-Class paradigm is that it assumes that the classes assigned to each pixel are mutually exclusive, which is not the case for labels that represent brain regions. For instance, the Brainstem label consists of a combination of labels, such as Midbrain, Pons, and Medulla. During the training process, the Multi-Class segmentation model used data where each pixel in a 2D scan slice was assigned an integer, representing the class that the pixel belonged to [**Figure 3.1.a**]. In contrast, the Multi-Label segmentation model would assign each pixel in the training data a corresponding binary vector, where each index would correspond to a class from the overall set of classes [**Figure 3.1.b**]. This approach helps to address the issue of mutually exclusive classes and allows for the representation of multiple classes for a single pixel.

Using vectors opens up the opportunity to assign multiple labels to a single pixel. For instance, if a pixel was originally labeled as 5 - "Midbrain", the vectorized value would be [0, 0, 0, 0, 0, 1, 0, 0, 1], where 1 is assigned to indexes 5 and 8. The index corresponds to the "Brainstem" label.

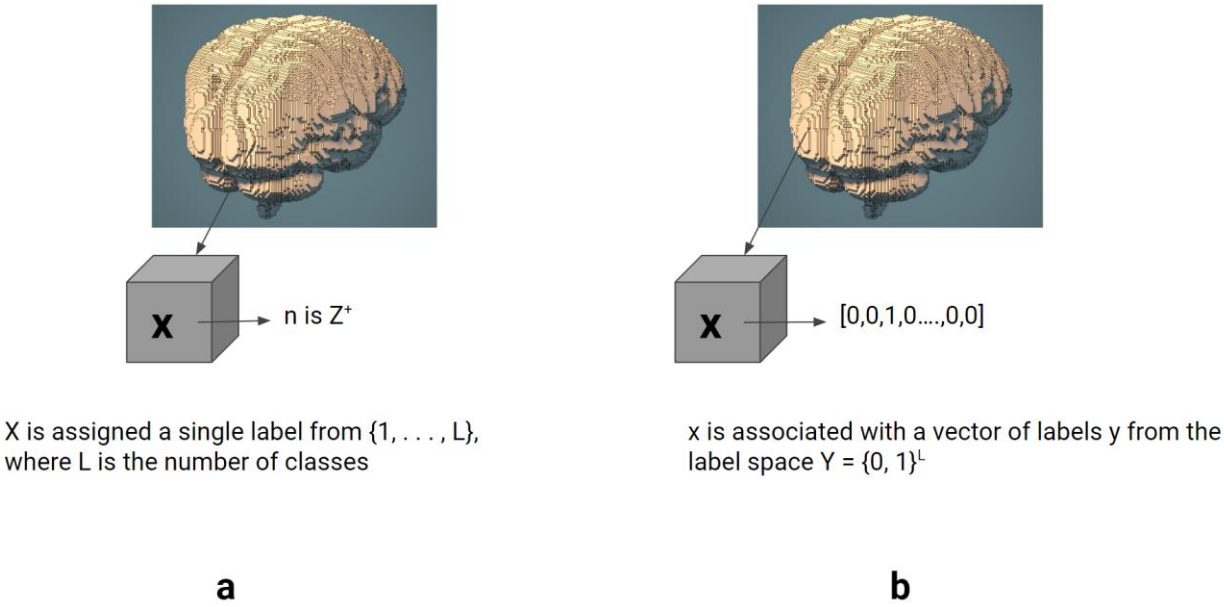


Figure 3.1: a) Visualization of pixel values in the training data for the Multi-Class Segmentation Model to be trained on; b) Visualization of pixel values in the training data for the Multi-Label Segmentation Model. **[27]**

When dealing with partially segmented data, some of the pixels in the 2D slice may not be assigned to a specific class, which raises the question of how to interpret these unannotated inputs in the corresponding vector. In this work, unannotated classes are marked with a value of 2 in the vector. For example, if the MRI annotation is incomplete and the Brainstem sub-labels are not annotated anywhere in the scan, we would assign values of 2 at the indexes of 5, 6, and 7 in the array:

[0, 0, 0, 0, 0, 2, 2, 2, 1].

The approach of introducing the "unknown" value to the vector of class presence in the input was inspired by a paper prepared with the support of Microsoft Research **[23]**. The authors of that paper analyzed existing loss functions used in multi-label classification and created their own approach to tackle the task of Multi-Label Learning from Single Positive Labels.

After introducing the "unknown" label to the vectors, the "ignore unobserved" (IU) loss function (1) was applied to manage the introduced data during training. The IU function was also introduced in the same paper **[23].**

$$\mathcal{L}_{IU}(x, y) = -\frac{1}{L}\sum_{i=1}^{L}(1_A(x_i)\log(y_i) + 1_B(x_i)\log(1 - y_i)) \quad \textbf{(1)}$$

Where

- **L** stands for the amount of classes;
- **x** is a vector of original labels for the current pixel; x ∈ $\{0, 1, 2\}^L$, where

  0 – class under index i ∈ L is not present in the current pixel;

  1 – class under index i ∈ L is present for the current pixel;

  2 – class under index i ∈ L is not observed (“unknown”);
- **y** is a vector of class probabilities predicted by the model for some input pixel.
- $\mathbf{1_A}$ and $\mathbf{1_B}$ are indicator functions, where A= {1} and B = {0}

3.2 Methods

In this section, we will discuss the segmentation model used in this thesis, which is a UNET-based model trained on unique data of neonatal children from Canada. The original version of this model was developed by Pouria et al **[2]** for multi-class segmentation of neonatal brain MRI images. Nevertheless, in the current work, several modifications to the model were made, including changing the preprocessing of the data and altering the model's architecture from being multi-class to multi-label. These modifications were made to improve the accuracy and robustness of the model and to adapt to the learning from partially labeled data task of this work. These modifications will be described in detail and the results of the experiments conducted will be presented as well to evaluate the performance of the new model.

3.2.1 The Nature of the Dataset

In 2006, a prospective longitudinal cohort study was initiated at the British Columbia's Women's Hospital in Vancouver, Canada. The study aimed to investigate the neurological development of preterm neonates and enrolled a total of 234 participants. The included neonates were born at a postmenstrual age (PA) of 24 to 32 weeks, making them particularly vulnerable to neurodevelopmental deficits. To assess the developmental trajectory of these preterm neonates, brain MRI scans were obtained at various timepoints throughout their early life.

Specifically, at a PA of 32 weeks, 208 patients from the full cohort underwent brain MRI scans. Subsequently, 188 patients received MRI scans at term-equivalent age, providing a comprehensive understanding of neurodevelopmental changes that occurred during the third trimester of gestation. Furthermore, to investigate the long-term impact of preterm birth on neurological development, 133 of the original patient cohort received MRI scans at 8 years old. By employing a comprehensive approach that involved scans at various timepoints throughout early life, the study aimed to obtain a thorough understanding of the neurodevelopmental trajectory of preterm neonates and to identify any associated risks for long-term neurodevelopmental deficits.

The model proposed in this work was trained utilizing both preterm and at-term scans. This selection was made purposefully, as a result of the experimental nature of the study. Furthermore, by focusing solely on preterm and at-term neonates, the training time was optimized. It is important to highlight that the validation and test sets were exclusively comprised of preterm and at-term scans. This approach allowed for a streamlined and efficient method for both training and evaluating the model's performance. Notably, this methodology maintains the possibility for future expansion, including the integration of scans from varying gestational ages.

3.2.2 Pre-processing

To make the dataset homogeneous and uniform, adjustments were made to the shape, orientation, dimensions, and intensities of each MRI scan. The first preprocessing step involved aligning all scans in one plane, with the proposed model trained on the coronal view of the scans. It is important to note that MRI scans can be taken in different directions, and in different medical facilities, this can result in the voxel size of the 3D scan being non-isotropic. Fortunately, we were able to work with isotropic MRIs, which allowed us to normalize the dataset to be viewed from one plane across all scans.

The next step was to ensure that all scans had the same shape. We chose a 3D scan shape of (256,256,256), and we used zero-padding to achieve this for all scans. Additionally, we normalized the grey values for every scan to be within the range of 0 to 1.

This work introduces some new preprocessing steps. In addition to the raw scans, we also have segmentation maps that were prepared by highly skilled medical practitioners using a manual segmentation method. The original model used the original version of the provided labels, where each voxel would have a number corresponding to the region of the brain from the next label map.:

0 – Background

1 – Total Cerebral volume (TCV)

2 – Cerebrospinal fluid (CSF)

3 – Left Cerebellum

4 – Right Cerebellum

5 – Midbrain

6 – Pons

7 – Medulla

8 – Brainstem

It is evident that the original annotation is most suitable for a multi-class segmentation task, where all classes are mutually exclusive. However, as previously mentioned, the classes are not exclusive, and a voxel can be annotated with more than one class.

Our main objective is to utilize partially labeled datasets, where some MRIs have been thoroughly annotated using labels such as Midbrain, Medulla, and Pons, while others have more general annotations using global labels like Brainstem.

To achieve this goal, we added another preprocessing step. Whenever we encounter a label that is a known part of another more general label, we create a vector where a value of 1 is assigned to both the index of the sub-label and the index of the global label (refer to **Figure 3.2**).

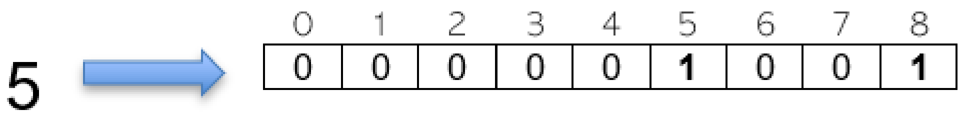


Figure 3.2: the original label 5 is shown on the left side of the arrow, which represents the class "Midbrain". On the right side of the arrow is a resulting vector where a value of 1 is assigned to indexes 5 and 8. The value at index 8 represents the annotation for class "Brainstem" for the given voxel.

For labels that are not part of more general labels, we perform simple one-hot encoding, as shown in **Figure 3.3**.

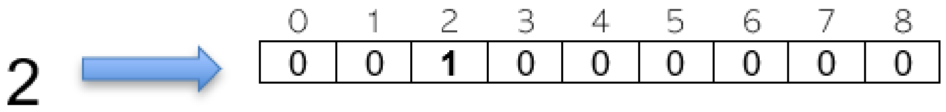


Figure 3.3: one-hot encoding example

3.2.3 Model Architecture

Convolutional Neural Networks (CNNs) have demonstrated their effectiveness in image segmentation tasks. Therefore, it was decided to use a CNN based on the U-NET architecture **[24]** to segment brain MRIs from neonates.

Although each scan has a shape of (256,256,256,9) after the preprocessing step, it was decided to break down each 3D image into 2D slices of shape (256,256,9) and train the model on each slice separately. There are several reasons why this approach was chosen.

Firstly, the original authors of the model aimed to minimize the number of inputs. Therefore, the model only requires one 3D file as input and manages the slices of the image within the pipeline.

Secondly, training the model by processing 3D voxels of the input can be biased towards the voxel size and overall thickness of the slices. Instead of normalizing voxels and potentially causing data loss, it was decided to stick with the slicing approach.

Lastly, the work required many rounds of experimental training, which is a time and space-consuming process. If 3D data were used for training, it would take up much more space and training data due to the increased weight at the input stage caused by the large size of the 3D file.

The proposed model has 5 convolutional layers and is trained using 100 epochs with an option for early stopping. The model is written in Python and Tensorflow, with Batch Normalization used to regulate the training process using batches of size 32. Adam optimizer is used, and the Ignore Unobserved loss function is applied to learn from partially labeled data. Sigmoid activation layer is chosen due to the Multi-Label nature of the model. Once all slices were given a prediction mask, they are connected to form a resulting segmentation 3D map output of shape (256,256,256). The original Multi-Class model's architecture is shown in **Figure 3.4**.

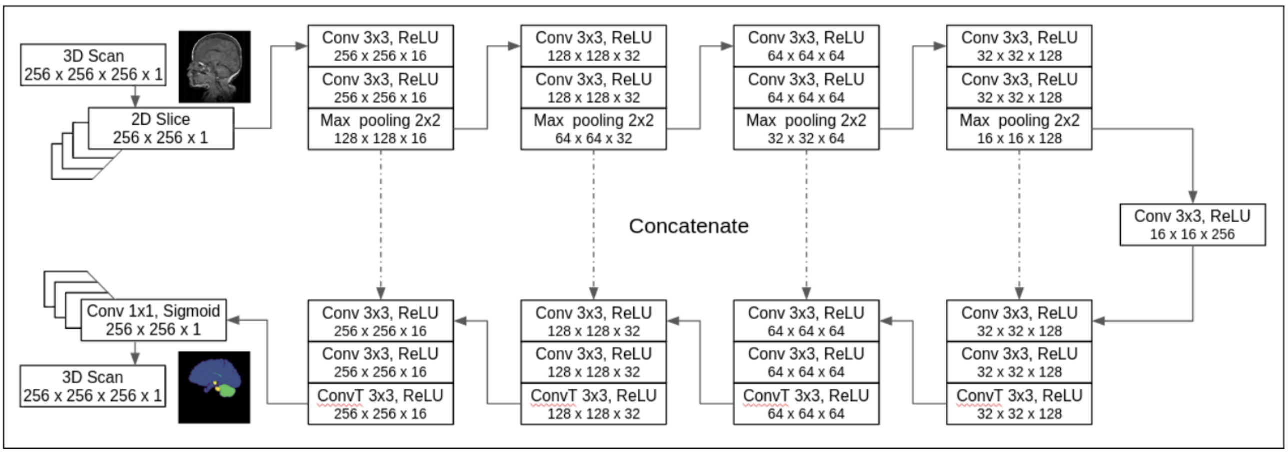


Figure 3.4: Original Neonatal Brain MRI Segmentation U-Net model architecture **[2]**

3.2.4 Post-Processing

The slice predictions of the model take the shape of (256, 256, 9), with the last axis representing a prediction vector created using a sigmoid activation layer. To enable the Multi-Label option, it was decided to choose the argmax between the nine indexes. If the argmax index corresponds to a class that is a combination of subclasses, the prediction values of the subclasses are considered, and the index with the maximum value is assigned a value of 1. An example illustrating this process is presented in **Figure 3.5**.

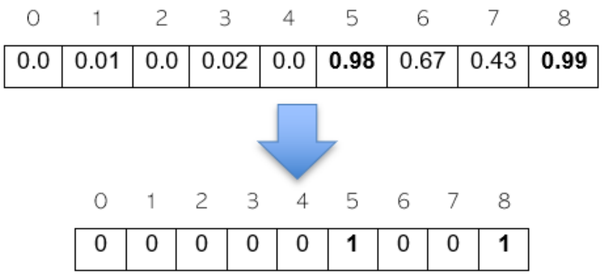


Figure 3.5: binarization of the output predictions generated by the model. In this figure, index 8 corresponds to the class "Brainstem", which comprises three sub-classes: "Midbrain" (index 5), "Pons" (index 6), and "Medulla" (index 7)

Once the binary form of the prediction map is generated, the prediction slices are reassembled into a 3D form. During the development and evaluation of the model's predictions, it was observed that small areas around the brain were occasionally misclassified as belonging to a specific class, when they should be identified as one-part areas. To address this issue, the Contour Approximation Method **[25]** from the OpenCV library was employed to detect and remove the outliers or "blobs" in the predictions.

3.2.5 Evaluation Metrics

To assess the accuracy of the segmentation produced by the trained model, a testing dataset was employed. The Dice Similarity Coefficient (DSC) **[26]** was used as the evaluation metric, as depicted in **Figure 3.6**. The DSC is a widely used metric in computer vision that evaluates accuracy based not only on the correct identification of positive labels but also takes into account the number of false positives.

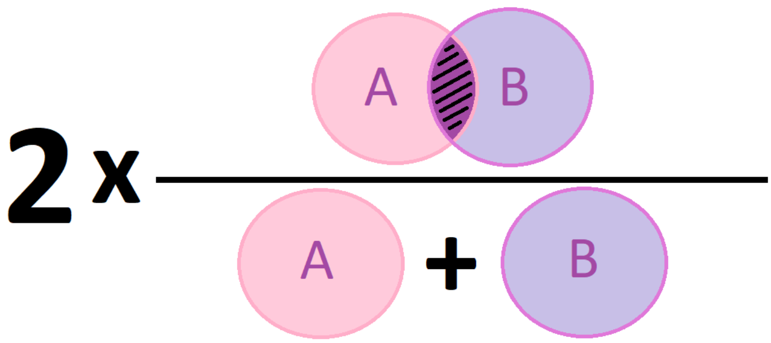


Figure 3.6: DCS formula, where A and B are sets of values(pixels). The result is doubled area of the overlap divided by the overall number of values in both sets.

3.3 Results

To evaluate the proposed model, a series of experiments were conducted on various data distributions. The objective of this study was to train the model using partially labeled data while maintaining accuracy levels for fully labeled data and potentially improving predictions for partially annotated classes.

Prior to model training, a random selection of IDs from the training set was made, comprising 5%, 10%, and 20% of the full training set in each case. Two models were then trained for each percentage: one solely using the selected IDs with full labels, and the other trained using the selected IDs with full annotation, in addition to the remaining training set which was artificially made partially labeled.

The approach employed to simulate partial annotation involved assigning a value of 2 to classes "Midbrain", "Pons", and "Medulla" in every vector where the global label "Brainstem" was present. **Figure 3.7** illustrates an example of this transformation.

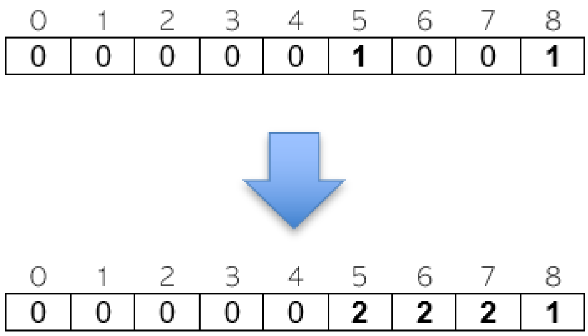
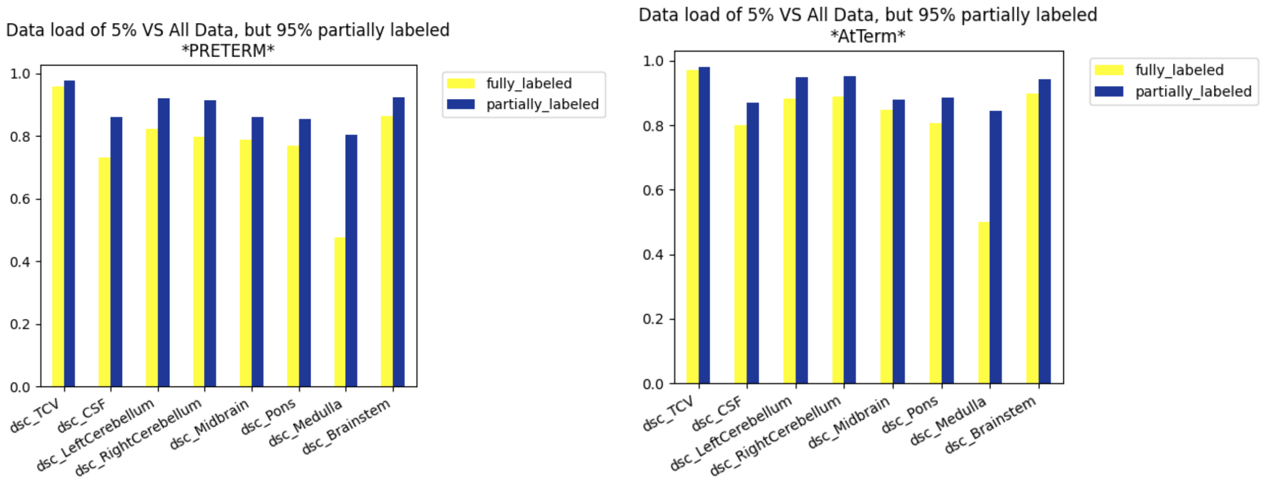


Figure 3.7: the creation of partially labeled data for training, where a value of 2 indicates that the class at the current index is unannotated.

Two models were initially trained for the study, one using 5% of fully labeled training data (FL), while the other was trained using 5% of fully labeled training data and 95% of partially labeled data (PL) created according to the scheme shown in **Figure 3.7**. This experiment was run 2 times on 2 different sets of ids with full annotation. The results presented in **Figure 3.8** demonstrate an improvement in the average DSC values for all labels, including those that were partially labeled in the second model ("Midbrain", "Pons", and "Medulla") for both preterm and at-term neonates' MRI scans.



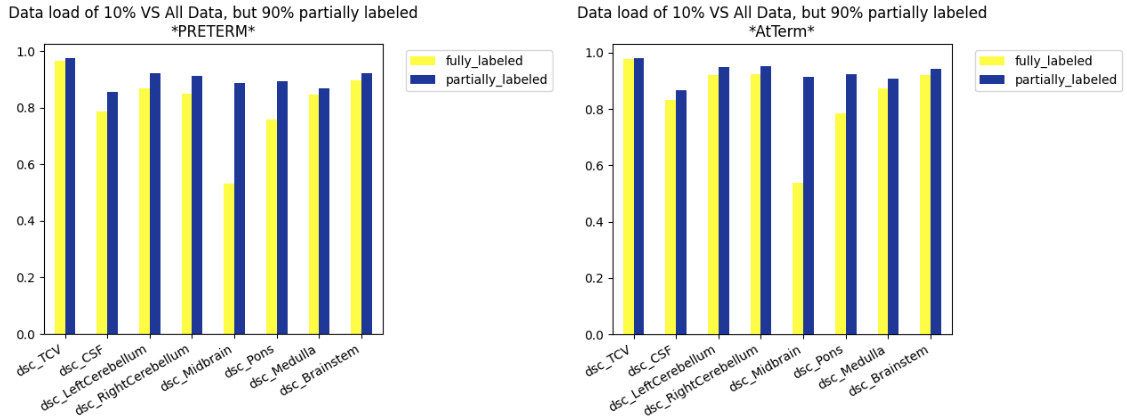
a) b)

Figure 3.8: a) plot of average DCS scores of each label for a model trained with 5% of fully labeled(FL) data and a model trained with 5% FL data and 95% partially labeled data, which were then evaluated on preterm infants' data; b) that for 2 models evaluated on at-term infants' data.

For preterm scans DCS value for "Midbrain" label improved by **~8%** (FL: 0.78895, PL: 0.8601); for "Pons" by **~8%** (FL: 0.77025,PL: 0.8534); for "Medulla" by **~33%** (FL: 0.4753,PL: 0.8046).

For at-term scans DCS value for "Midbrain" label improved by **~3%** (FL: 0.84845, PL: 0.87915); for "Pons" by **~8%** (FL: 0.8069, PL: 0.8851); for "Medulla" by **~35%** (FL: 0.4986,PL: 0.84575).

The next experiment was run 2 times on 2 different sets of ids with full annotation. The goal was to check results for another two models: one trained on 10% of fully labeled training data (FL), and another trained on 10% of fully labeled training data along with 90% of partially labeled data (PL). The results in **Figure 3.9** have shown, that the average DCS values improved as well for all of the labels, including "Midbrain", "Pons", and "Medulla" both for preterm and at-term neonates' MRI scans.



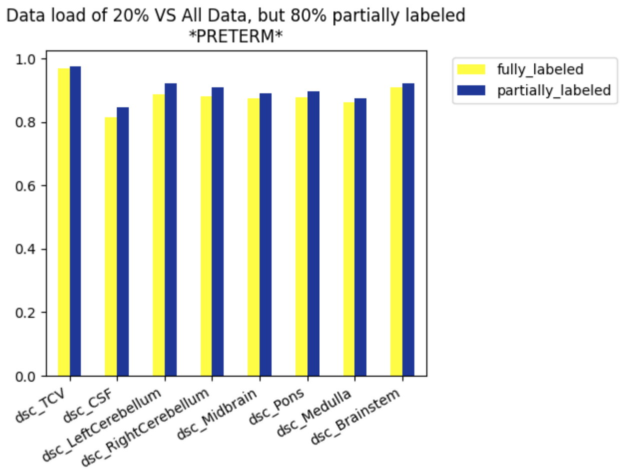
a) b)

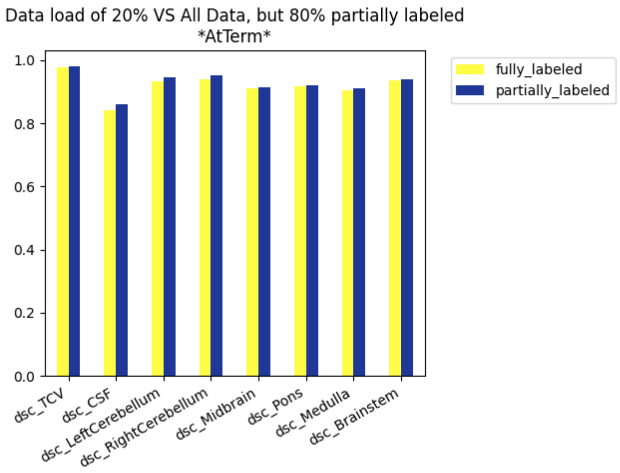
Figure 3.9: a) plot of average DCS scores of each label for a model trained with 10% of fully labeled(FL) data and a model trained with 10% FL data and 90% partially labeled data, which were then evaluated on preterm infants' data; b) that for 2 models evaluated on at-term infants' data.

For preterm scans DCS value for "Midbrain" label improved by **~35%** (FL: 0.5322, PL: 0.88655); for "Pons" by **~14%** (FL: 0.7586, PL: 0.8937); for "Medulla" by **~2%** (FL: 0.8457,PL: 0.8687).

For at-term scans DCS value for "Midbrain" label improved by **~38%** (FL: 0.53865, PL: 0.913); for "Pons" by **~14%** (FL: 0.7831,PL: 0.9228); for "Medulla" by **~3%** (FL: 0.8741,PL: 0.9089).

The final experiment was also run 2 times on 2 different sets of ids with full annotation. The goal was to check the results for the next two models: one trained on 20% of fully labeled training data (FL), and another trained on 20% of fully labeled training data along with 80% of partially labeled data (PL). The results in **Figure 3.10** have shown, that the average DCS values slightly improved for all of the labels, including "Midbrain", "Pons", and "Medulla".





a) b)

Figure 3.10: a) plot of average DCS scores of each label for a model trained with 20% of fully labeled(FL) data and a model trained with 20% FL data and 80% partially labeled data, which were then evaluated on preterm infants' data; b) that for 2 models evaluated on at-term infants' data.

For preterm scans DCS value for "Midbrain" label improved by **~2%** (FL: 0.8755, PL: 0.8913); for "Pons" by **~2%** (FL: 0.87865, PL: 0.89665); for "Medulla" by **~1%** (FL0.86295:,PL: 0.8766).

For at-term scans DCS value for "Midbrain" label improved by **~0.5%** (FL: 0.91035, PL: 0.9155); for "Pons" by **~0.3%** (FL: 0.91575,PL: 0.91895); for "Medulla" by **~1%** (FL: 0.90375,PL: 0.9102).

In summary, the experiments conducted in this study have shown that using partially labeled data can improve the performance of the model, especially when the ratio of fully annotated labels is small in the training set. This scenario is quite common in clinical settings, where medical practitioners typically briefly annotate MRI scans, resulting in a large proportion of partially labeled datasets.

3.4 Place for improvement

The work presented in this thesis represents a crucial first step toward developing the best possible approach for training a neonate’s brain MRI segmentation model on weakly annotated labels. However, there is still room for improvement in utilizing the information on the likelihood of labels appearing at specific locations. This can be achieved by implementing more advanced approaches to construct loss functions.

One such approach is the regularized online label estimation (ROLE) **[23]** loss function, which takes into account the expected number of positive labels per image and avoids the trivial “always predict positive” solution. The ROLE loss function also estimates labels under constraints imposed by domain knowledge, leading to better model performance. Therefore, the application of customized loss functions like ROLE has the potential to significantly improve the performance of the model.

In the field of brain image segmentation, labels are often linked by hierarchical and hard logical constraints. In such cases, published solutions in the form of additional layers to multi-label classification models **[20]** can be adapted for segmentation tasks. These solutions can help capture the intricate relationships among labels, thereby improving the model's performance.

In summary, incorporating domain knowledge into the model development process represents a promising next step. By developing more advanced loss functions and utilizing additional layers, we can better leverage the wealth of information contained in weakly annotated data and improve the accuracy of our models in various fields, including brain image segmentation.

Conclusion

In this thesis, extensive research was conducted on the cutting-edge technologies in the fields of automatic Medical Image Segmentation, Multi-Label Classification, and Learning from Partially labeled data. The goal was to develop a tool for the automatic segmentation of brain MRIs from preterm neonates. By making changes to the original model at the preprocessing, training, and postprocessing stages, a new Multi Label Segmentation model was developed. This model showed improved performance, particularly when the ratio of fully annotated labels is small in the training set, which is a common situation in the clinical domain.

Although the proposed model represents a significant step forward in the automatic segmentation of brain MRIs, further updates and enhancements are necessary to build a robust tool that can be used in various hospitals. Future work could include introducing a more advanced loss function during the training process, as well as adding an additional constraint layer after the base model is trained. Overall, this research highlights the potential for machine learning to revolutionize the field of medical imaging and improve patient outcomes.

Resources

1. Lauric, A., & Frisken, S.F. (2007). Soft Segmentation of CT Brain Data.
2. Mashouri, P., Haghpanah, F. S., Osia, S., Sheng, M., Guillot, M., Garfinkle, J., McAllister, D., Brudno, M., & Guo, T. (2022). Automatic U-Net based Segmentation Pipeline for Neonatal and Child Brain MRI. Retrieved from https://openreview.net/forum?id=tVZebHNiuJ
3. Li, J., Erdt, M., Janoos, F., Chang, T., & Egger, J. (2021). 1 - Medical image segmentation in oral-maxillofacial surgery. In J. Egger & X. Chen (Eds.), Computer-Aided Oral and Maxillofacial Surgery (pp. 1-27). Academic Press.
4. Papers with Code. (n.d.). Semantic Segmentation. Retrieved from https://paperswithcode.com/task/semantic-segmentation
5. Prastawa, M., Gilmore, J. H., Lin, W., & Gerig, G. (2005). Automatic segmentation of MR images of the developing newborn brain. Medical image analysis, 9(5), 457–466.
6. Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., & Smith, S. M. (2012). Fsl. Neuroimage, 62(2), 782-790.
7. Ashburner, J. (2012). SPM: a history. Neuroimage, 62(2), 791-800.
8. Fischl, B. (2012). FreeSurfer. Neuroimage, 62(2), 774-781.
9. Fetit, A. E., Cupitt, J., Kart, T., & Rueckert, D. (2020, September). Training deep segmentation networks on texture-encoded input: application to neuroimaging of the developing neonatal brain. In Medical Imaging with Deep Learning (pp. 230-240). PMLR.
10. Zöllei, L., Iglesias, J. E., Ou, Y., Grant, P. E., & Fischl, B. (2020). Infant FreeSurfer: An automated segmentation and surface extraction pipeline for T1-weighted neuroimaging data of infants 0–2 years. Neuroimage, 218, 116946.

11. Liu, M., Yan, X., Wang, C., & Wang, K. (2021). Segmentation mask-guided person image generation. Applied Intelligence, 51, 1161-1176.

12. Moeskops, P., Viergever, M. A., Mendrik, A. M., De Vries, L. S., Benders, M. J., & Išgum, I. (2016). Automatic segmentation of MR brain images with a convolutional neural network. IEEE transactions on medical imaging, 35(5), 1252-1261.

13. Davies, A. J. (2022). Semantic segmentation of aerial imagery using U-net in python. Retrieved from https://towardsdatascience.com/semantic-segmentation-of-aerial-imagery-using-u-net-in-python-552705238514

14. Jordan, J. (2018). An overview of semantic image segmentation. Retrieved from https://www.jeremyjordan.me/semantic-segmentation/#advanced_unet

15. Lempart, M., Nilsson, M. P., Scherman, J., Gustafsson, C. J., Nilsson, M., Alkner, S., ... & Olsson, L. E. (2022). Pelvic U-Net: multi-label semantic segmentation of pelvic organs at risk for radiation therapy anal cancer patients using a deeply supervised shuffle attention convolutional neural network. Radiation Oncology, 17(1), 114.

16. Zhou, D., Chen, P., Wang, Q., Chen, G., & Heng, P. A. (2022, November). Acknowledging the unknown for multi-label learning with single positive labels. In Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXIV (pp. 423-440). Cham: Springer Nature Switzerland.

17. Ben-Baruch, E., Ridnik, T., Friedman, I., Ben-Cohen, A., Zamir, N., Noy, A., & Zelnik-Manor, L. (2022). Multi-label classification with partial annotations using class-aware selective loss. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 4764-4772).

18. Mehrkanoon, S., Huang, X., & Suykens, J. A. (2020). Learning from partially labeled data. In ESANN (pp. 493-502).

19. Yan, K., Cai, J., Zheng, Y., Harrison, A. P., Jin, D., Tang, Y., ... & Lu, L. (2020). Learning from multiple datasets with heterogeneous and partial labels for universal lesion detection in CT. IEEE Transactions on Medical Imaging, 40(10), 2759-2770.

20. Giunchiglia, E., & Lukasiewicz, T. (2021). Multi-label classification neural networks with hard logical constraints. Journal of Artificial Intelligence Research, 72, 759-818.

21. Gururani, S., & Lerch, A. (2021, November). Semi-supervised audio classification with partially labeled data. In 2021 IEEE International Symposium on Multimedia (ISM) (pp. 111-114). IEEE.

22. Saffar, I., Morel, M. L. A., Singh, K. D., & Viho, C. (2019, January). Machine learning with partially labeled data for indoor outdoor detection. In 2019 16th IEEE Annual Consumer Communications & Networking Conference (CCNC) (pp. 1-8). IEEE.

23. Cole, E., Mac Aodha, O., Lorieul, T., Perona, P., Morris, D., & Jojic, N. (2021). Multi-label learning from single positive labels. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 933-942).

24. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18 (pp. 234-241). Springer International Publishing.

25. OpenCV. (n.d.). Contours: Getting Started. Retrieved from https://docs.opencv.org/4.x/d4/d73/tutorial_py_contours_begin.html

26. Zou, K. H., Warfield, S. K., Bharatha, A., Tempany, C. M., Kaus, M. R., Haker, S. J., ... & Kikinis, R. (2004). Statistical validation of image segmentation quality based on a spatial overlap index1: scientific reports. Academic radiology, 11(2), 178-189.

27. Backos, D. (Artist). (n.d.). Voxel Brain Edge [Digital image]. Retrieved from https://cdnb.artstation.com/p/assets/images/images/019/824/069/large/don-backos-voxelbrainedge.jpg?1565157850