Medical Segmentation Decathlon

A Project Report Submitted
in Partial Fulfillment of Requirements
for the Degree of

Bachelor of Technology

by

Abhinav Jindal(2016csb1026)
Mattela Nithish (2016csb1042)

Department of Computer Science & Engineering

Indian Institute of Technology Ropar

Rupnagar 140001, India

May 2020

Abstract

Image Segmentation is a task that involves pixel-wise classification in images without any human intervention. Most of the research conducted in medical image segmentation is tested on a small set of tasks, which limits our understanding of the generalisability. In the field of medical imaging, the data available is small and varies a lot such as unbalanced labels, multi-modal imaging, multi-class labels, etc. To tackle this problem, Grand challenges 2019 has come up with the Medical Segmentation Decathlon problem. This problem aims to come with a generalized segmentation model that works decently across various segmentation tasks. Our goal is to come with a simple model (preferably involving 2D image slices) without any complex ensembles. Most of the experiments have been done on the VGG-UNET model to come up with a generalized model. Experiments involve changing pre-processing, post-processing, and also some architecture changes to the VGG-UNET model. We made sure that architecture is constant for all the organs while testing.

Acknowledgements

And we would like to acknowledge our supervisor Dr. Deepti Bathula for her support and guidance during this project.

Honor Code

We certify that we have properly cited any material taken from other sources and have obtained permission for any copyrighted material included in this report. We take full responsibility for any code submitted as part of this project and the contents of this report.

Abhinav Jindal (2016CSB1026)

Mattela Nithish (2016CSB1042)

Certificate

It is certified that the B. Tech. project "Medical segmentation Decathlon" has been done by Abhinav Jindal (2016CSB1026), Mattela Nithish (2016CSB1042) under my supervision. This report has been submitted towards partial fulfillment of B. Tech. project requirements.

Dr. Deepti Bathula
Project Supervisor
Department of Computer Science & Engineering
Indian Institute of Technology Ropar
Rupnagar-140001

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Medical segmentation has become a common area of research for various computer vision and machine learning researchers. Medical segmentation decathlon is one of the problems from the grand challenge which aims at the generalization of the segmentation tasks. Most of the medical image segmentation researches aim at a particular organ and hence, has some procedures during pre-processing and post-processing that are specific to that organ. Hence the same method performs poorly on other organs. Medical Segmentation Decathlon aims at a generalized method that can perform segmentation optimally on any organ that is trained. Hence, the same procedure can be deployed to any other organ in the future and is useful if some organ doesn't have a dedicated procedure for its segmentation.

In this paper, we have mostly worked with bi-label data which can be easily extended to multi-label data. Various pre and post-processing experiments have been conducted and various modifications to the architecture were tested. While others have mostly used complex ensembles of various 3D and cascaded models, our model is a basic 2D UNET for feasible training. VGG UNET is our basic model for most of the experiments. Dataset used is from the challenge itself which consists of 10 organs of various modality, image size, dataset size, labels, adding extra generalization, and complexity to the project.

Chapter 2

Related Works

2.1 Multi-View Fully Convolutional Neural Network

A modified version of fully convolutional networks in Mathias Perslev and Pai [2018] uses a fusion model for slices along multiple axes to find the segmentation volume. The 3D features are extracted using a fusion of 2D fully convolutional neural models along with each view. The fusion model trains to learn the weights assigned to each view to maximizing the ensemble output performance. 6 isotropic axes are used for the fusion model and their weighted average is taken for the final output 3D image volume. We have used 3 major axes instead of 6 in our approach to extract the 3D properties using 2D models.

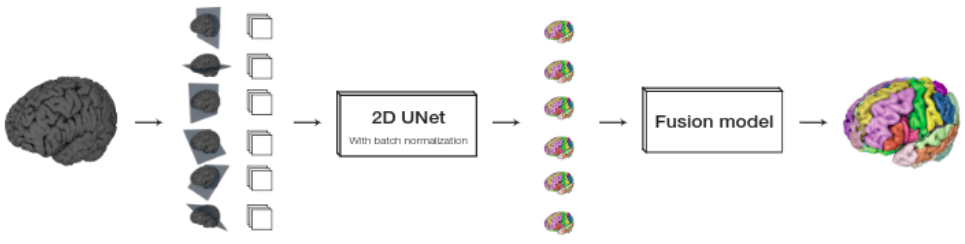

Figure 2.1: 6 axis for multi view model and their fusion (Mathias Perslev and Pai [2018]).

$$z(x)_k = \sum_{i=1}^{|V|} W_{i,k} \cdot p_{i,x,k} + \beta_k \quad \forall\, k \in \{1, ..., K\}$$

Figure 2.2: Weighted average is taken for the overall prediction where the weights are also learned during training(Mathias Perslev and Pai [2018]).

2.2 nnU-Net: Self-adapting Framework for U-Net-Based Medical Image Segmentation

nnU-Net by Isensee et al. [2018] is one of the good models to have a high dice score in medical decathlon tasks. This model adapts itself based on the dataset provided. The standard UNet model has ReLU as the activation function but nnU-Net uses Leaky-ReLU as the activation function. This model comprises of 3 variants of UNet model. They are a 2D-UNet model, a 3D-UNet model, UNet-Cascade model. The ensemble of Appropriate models is chosen to predict based on the task provided. The network architecture is dynamically chosen and depends on the median dimensions of the task dataset provided. The pre-processing involves Cropping, Re-sampling, and Normalization. The Normalization step depends on the modality type of task provided. While training, Data augmentation is used to improve results. In the post-processing step, A connected component analysis is performed where only the largest connected segmentation output is chosen. To further increase the segmentation performance and robustness all possible combinations of two out of three of these models are ensembled for each dataset. This model fine-tunes everything depending on the dataset given which is not too generalized. This model is an ensemble of 3 UNet models which is not feasible in a general sense. We have experimented with some of the pre-processing steps mentioned in this paper.

Chapter 3

Basic Information

3.1 Dataset

Dataset is directly taken from the Grand Challenge dataset provided along with the problem. It consists of 10 different organs each of which contains a different number of 3D or 4D images. The basic properties for images of each task(organ) in the dataset are given in Table 3.1 and the label information is given in Table 3.2.

3.2 Evaluation Metric

Dice score is taken as the evaluation metric for our project. It is basically the ratio of twice the intersection between actual and predicted regions to the sum of these 2 regions (3.1).

$$(2 * |A| \cap |B|)/(|A| + |B|) \tag{3.1}$$

,where A is the predicted image and B is the actual labeled image

TASK	MODALITY	DIMENSION (Approx.)	NUMBER OF IMAGES
Brain	Multimodal multisite MRI data (FLAIR, T1w, T1gd,T2w)	(240, 240, 155, 4)	750 (484 Training + 266 Testing)
Liver	Portal venous phase CT	(512, 512, 247)	201 (131 Training + 70 Testing)
Prostate	Multimodal MR (T2, ADC)	(320, 320, 20, 2)	48 (32 Training + 16 Testing)
Hippocampus	Mono-modal MRI	(34, 53, 37)	394 (263 Training + 131 Testing)
Spleen	CT	(512, 512, 164)	61 (41 Training + 20 Testing)
Heart	Mono-modal MRI	(320, 320, 120)	30 (20 Training + 10 Testing)
Lung	CT	(512, 512, 450)	96 (64 Training + 32 Testing)
Pancreas	Portal venous phase CT	(512, 512, 82)	420 (282 Training +139 Testing)
Colon	CT	(512, 512, 127)	190 (126 Training + 64 Testing)
Hepatic Vessel	CT	(512, 512, 55)	443 (303 Training + 140 Testing)

Table 3.1: Basic properties for each task in dataset

TASK	NUMBER OF LABELS	LABEL DESCRIPTION
Brain	4	Background, Edema, Non enhancing tumor, Enhancing tumor
Heart	2	Background, left atrium
Liver	3	Background, Liver, Cancer
Hippocampus	3	Background, anterior, posterior
Prostate	3	Background,TZ,PZ
Lung	2	Background, Cancer
Pancreas	3	Background, Pancreas, Cancer
Heptic Vessel	3	Background, Vessel, Tumor
Spleen	2	Background, Spleen
Colon	2	Background, Colon cancer primaries

Table 3.2: Label information for each task in dataset

Chapter 4

Model

4.1 UNET

UNET by Ronneberger et al. [2015] is an architecture comprising of contracting path (left side) and an expansive path (right side) as shown in Figure 4.1. The contracting path resembles the architecture of a convolution network. It consists of repeated application of 3x3 convolutions operations followed by a rectified Linear Unit (ReLU). The expansive path is used for the construction of the segmentation output. In every step of the expansive path, a concatenation of corresponding feature map from contracting path and feature map from the level below followed by an up-sampling operation is performed. At the final layer, a 1x1 convolution is used to map pixels of the image into desired classes.

4.2 VGG-UNET

VGG-UNET is a model in which the contracting path(left side) of UNET is replaced by the VGG network. VGG (Simonyan and Zisserman [2014]) is one of the top-performing CNN's for the object classification task. The feature vector obtained at every step of the contracting path is improved by using the standard VGG weights rather than training from scratch. It is used as the base architecture for our project with several experiments conducted on top of it to check improvement. The 3D image available is sliced into 2D images along the x,y, and

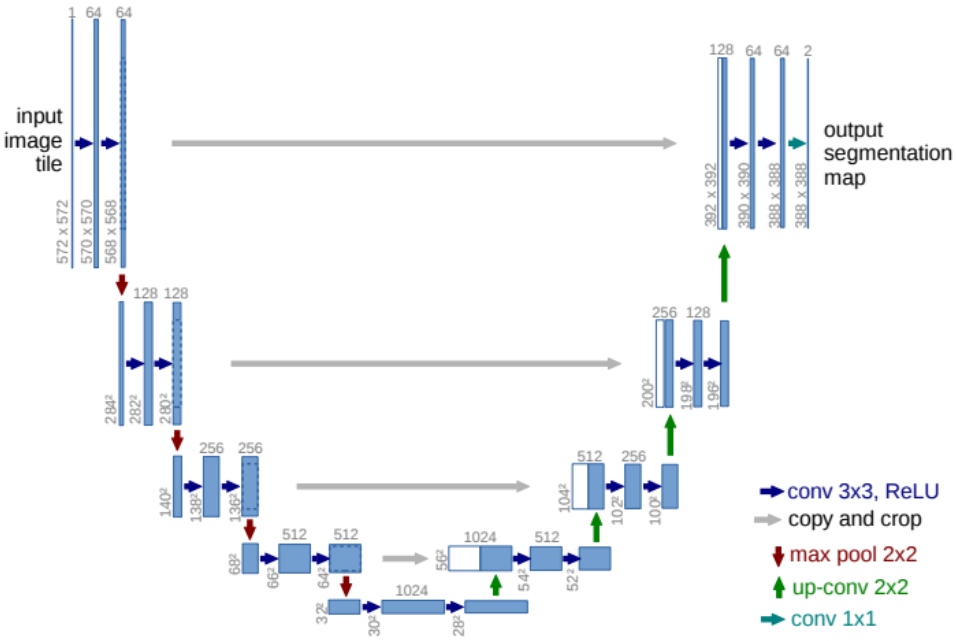

Figure 4.1: A sample UNET architecture taken from Ronneberger et al. [2015]. The blue boxes represent multi-channel feature map. The white boxes represent copied feature map.

z-axis, and these slices are passed to this network after resizing to (224,224,3) i.e. the input dimension for VGG for training.

Chapter 5

Experiments

5.1 Utilisation of different channels of input

The input dimension for the model is (224,224,3). In the base model, three copies of the input slice are given as input. In this experiment, Different forms of input are tested.

5.1.1 Using Contrast enhancement images

In most of the test samples, the contrast of the images is not good. So having a contrast enhancement of the image as input can improve the overall result. In this experiment, The three channels of the input are image slice, Log transformation of the image slice, and Power transformation of image slice respectively. We have experimented with various values of gamma and none of them have improved any result.

5.1.2 Using adjacent slices

To increase the spatial information of the model, The three channels of the input corresponds to the previous slice of the image slice, the image slice, and the next slice of the image slice. This change has improved the overall result of the model

5.2 Pre-processing(Normalization)

For images in the dataset, the image values are not confined to a particular range. Some of the images have values between [0, 100] whereas some of them have values between [0,1000] in the same dataset. So to have the same range of values across all the input images, Min-Max Normalisation is performed where the data is rescaled into [0,1].

5.3 Post-Processing(Connected Component Analysis)

We observed that after segmentation using the network the predicted region was sometimes scattered and not connected. To improve this, we tried a post-processing method in the form of connected component analysis. Since the labeled region is connected in all the organs, we ran a connected component analysis and then took the largest connected component and discarded the rest.

5.4 Images of Major Axis

Initially, for base results, we had taken the image slices along all the 3 axes together as a single dataset for training. The idea was to compensate for the less training data using this technique. But with further analysis, we observed that the model was getting confused probably due to getting too much variable data as the slices along the 3 axes had very different dimensions and results were severely affected due to this. So, instead, we just use the images along the major axis (axis 0) for training, which proved to be quite an improvement.

5.5 Ensemble of models along 3 axes

Using the images along all the 3 axes was not a good idea, but using the images along a single axis was quite effective. But this leads to the loss of some information available from the other axis. So we train 3 different models, one for each

of the axes and use the OR operation to get the final predicted region and then apply connected component post-processing to get the ROI. This leads to better utilization of the 2D slices along all the 3 axes and better results.

5.6 Self Supervised Learning

In the base model, The encoder part uses VGG Image Net weights and are fine-tuned accordingly. Self Supervised learning as mentioned in Misra and van der Maaten [2019] is to construct image representations that are semantically meaningful via pretext tasks such as image reconstruction, jigsaw puzzle solver..etc, that do not need any semantic annotations. In this experiment, we have taken image reconstruction as the pretext task and trained a VGG-UNET model with slight modifications at the decoder side. Once the model has been trained, we train a VGG-UNET for image segmentation by loading the encoder weights of the image reconstruction model trained. The idea behind this experiment is that by training an image reconstruction model, The encoder side of the model learns a rich set of features than the default VGG network.

5.7 Attention Gated Network

Attention gated networks in Schlemper et al. [2018] are effective for segmentation in cases where ROI has great variation. Attention network helps in effective localization and reducing false positives. It can be included in a CNN with almost no computational overhead and better training effectiveness. It helps in learning salient features through training the attention network. The attention mechanism requires a context and feature vector. In machine translation, the input language feature vector as the features and the converted text so far acts as the context. Similarly in UNET, the skip connections act as the context and upsampled features on the decoder side act as features. We then apply additive attention to the skip connection and upsampled features. We use ReLU and sigmoid as the activation function. ReLU is proved to be effective in CNN's and sigmoid helps in scaling to [0,1] to scale the features appropriately. The features and then learned to scale effectively and important features are extracted.

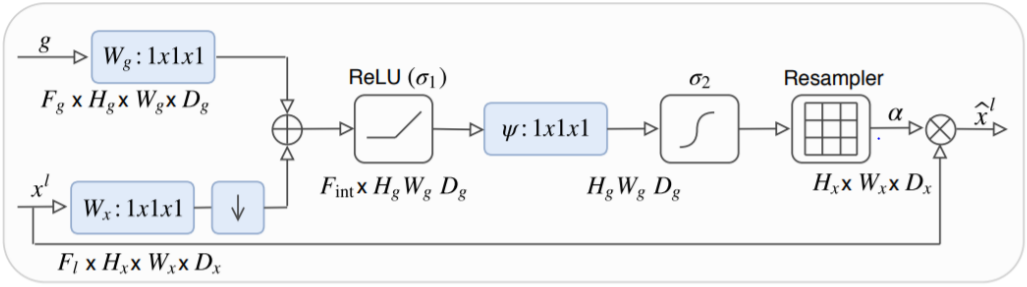

Figure 5.1: An attention module taken from Schlemper et al. [2018].

$$q^l_{att,i} = \boldsymbol{\psi}^T \left(\sigma_1 \left(\boldsymbol{W}_x^T \boldsymbol{x}_i^l + \boldsymbol{W}_g^T \boldsymbol{g} + \boldsymbol{b}_{xg} \right) \right) + b_\psi$$
$$\alpha^l = \sigma_2 \left(q^l_{att}(\boldsymbol{x}^l, \boldsymbol{g}; \boldsymbol{\Theta}_{att}) \right),$$

Figure 5.2: The attention formula that the module uses to train the network. Image taken from Schlemper et al. [2018].

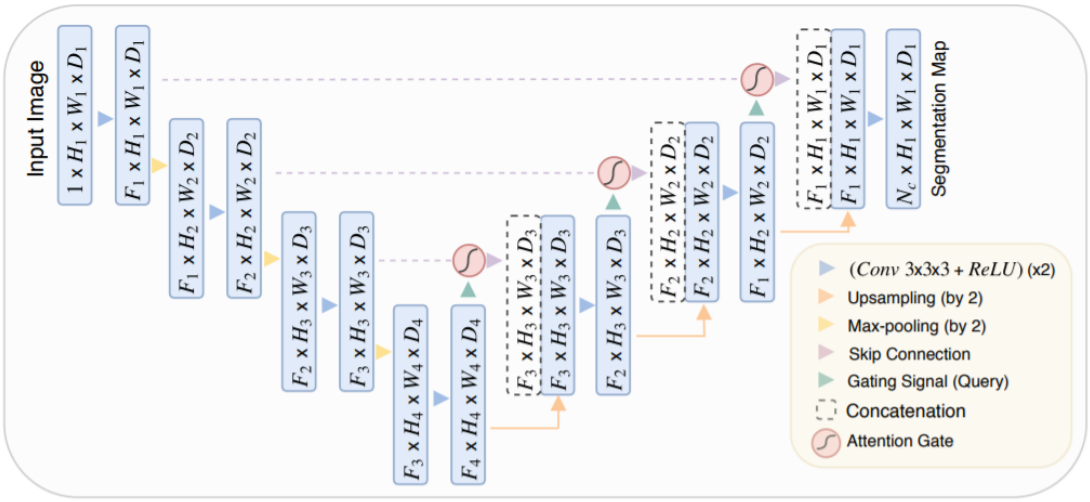

Figure 5.3: Attention UNET with attention modules in the decoder side of the unet. Image taken from Schlemper et al. [2018].

5.8 Ensemble of Attention Gated Network

This experiment is similar to the Ensemble of models along 3 axes, Here instead of a default VGG-UNET model, an Attention gated network is used.

Chapter 6

Results and Conclusions

The base results in Table 6.1 are for the basic VGG UNET with all the images from all the axis as a single dataset. We see that colon and lung have very poor score and others are still reasonable. We have tried to improve on these results in every experiment conducted. For each experiment, we have trained only some of the organs due to computational limitations. Our main purpose was to see the trend in the dice scores upon performing that experiment which can easily seen from the trained organs.

TASK	Dice Score(value+-variance)
Heart	0.69 +- 0.035
Hippocampus	0.720 +- 0.104
Prostate	0.715 +- 0.076
Lung	0.156 +- 0.246
Pancreas	0.381 +- 0.139
Heptic Vessel	0.171 +- 0.139
Spleen	0.74 +- 0.09
Colon	0.02 +- 0.01

Table 6.1: Base results

6.1 Utilisation of different channels of input

6.1.1 Using contrast enhancement images

The contrast enhancement almost did not affect the base results. This is probably because the network itself learns to count for contrast variations while training and adding these additional pieces of information does not matter to the final segmented output.

6.1.2 Using adjacent slices

Using adjacent slices, prove to be an improvement as it provides some spatial context, if not all, to the model. Our model is made for 2D images and hence may miss some information available when we see the whole 3D image as one. But adding these slices helps in somewhat getting a minor part of this information and improves the results a bit as can been seen from Table 6.2.

TASK	Dice Score(value+-variance)
Lung	0.2057 +- 0.211
Pancreas	0.34 +- 0.1
Colon	0.064 +- 0.151

Table 6.2: Result obtained when using Adjacent slices as the input

6.2 Images of Major Axis

Using only the major axis slices instead of all the slices from all the axis led to a significant improvement. This is probably because giving too much variable data was confusing the model and hence, leading to poor results. Also while predicting we are using only the major axis and that might be the cause for improvement in the results (table 6.3).

TASK	Dice Score(value+-variance)
Pancreas	0.5 +- 0.19
Spleen	0.85 +- 0.04

Table 6.3: Result obtained when using images of major axis

6.3 Ensemble of models along 3 axes

Using the major axis only was an improvement as the model had now less variable data. But this leads to loss of some information available from other axes. Using the ensemble of models learned from all the axes leads to better utilization of this information. Spleen results especially have a very significant improvement with this experiment.

TASK	Dice Score(value+-variance)
Hippocampus	0.77 +- 0.104
Pancreas	0.42 +- 0.14
Spleen	0.90 +- 0.01

Table 6.4: Result obtained by using Ensemble of models along 3 axes

6.4 Attention Gated Network

Using attention in our UNET led to an improvement for most of the organs (Table 6.5). Attention leads to focusing on more important features and leads to better segmentation for small regions like in lungs whose score didn't improve much with other techniques.

6.5 Ensemble of Attention Gated Network

Using the ensemble of different axes models proved to be an improvement earlier. So we thought of trying it with attention but it was not that much of an improvement as we expected. This was probably due to OR operation being not

TASK	Dice Score(value+-variance)
Heart	0.75 +- 0.05
Hippocampus	0.720 +- 0.104
Lung	0.28 +- 0.28
Pancreas	0.52 +- 0.19
Spleen	0.86 +- 0.04

Table 6.5: Attention Gated Network results

that suitable in this case and we needed a better fusion model than just an OR operation. Spleen results are reduced quite significantly due to this.

TASK	Dice Score(value+-variance)
Heart	0.76 +- 0.035
Hippocampus	0.76 +- 0.06
Spleen	0.80 +- 0.1

Table 6.6: Ensemble of attention gated network results

Hence, we see the outcomes of the many experiments conducted for the project. Overall our results may not be to that level of the best available results in the challenge but keeping in mind that most of them have used complex models and ours is a simple 2D UNET, we can say that results are quite reasonable. Also, none of them seemed to used attention to improving their results but just a complex ensemble of models. So maybe if their models had a component of attention gated networks as well, the results might have been even better.

References

Fabian Isensee, Jens Petersen, Andre Klein, David Zimmerer, Paul F. Jaeger, Simon Kohl, Jakob Wasserthal, Gregor Koehler, Tobias Norajitra, Sebastian Wirkert, and Klaus H. Maier-Hein. nnu-net: Self-adapting framework for u-net-based medical image segmentation, 2018. 3

Erik B Dam Mathias Perslev, Christian Igel and Akshay Pai. A multi-view fully convolutional neural network for segmentation of medical image volumes. 2018. vii, 2, 3

Ishan Misra and Laurens van der Maaten. Self-supervised learning of pretext-invariant representations, 2019. 11

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, 2015. vii, 6, 7

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, 2018. vii, 11, 12

Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition, 2014. 6