Refining multi-organ segmentation using partially annotated datasets

Haobo Song, Chang Liu

[1]Fakultät für Pattern Recognition, FAU Erlangen-Nürnberg
`haobo.song@fau.de`

Abstract. U-Net is a generic deep-learning solution for segmentation in biomedical image data. When training a new U-Net[1] model for a specific segmentation task, getting finely-annotated datasets is usually difficult. In comparison to the method of training with a all-annotated dataset, we have verified a hypothesis. The model was trained by a series of partially annotated datasets with particular training method.

1 Introduction

In the field of radiation therapy and CT dose optimization, segmentation of organs at risk in CT scanning is usually required to facilitate the treatment planning or dose estimation. The growing number of people choosing neural networks for segmentation tasks [2]. When developing deep learning multi-organ segmentation tools, dataset is usually a problem since getting manual segmentation of all organs from medical staff is usually very difficult. Meanwhile, there exist many publicly available datasets where not all required organs are annotated. The following table 1 illustrates some public partially annotated datasets on the internet.

Table 1. Partially annotated datasets source from the Internet

Source	Liver	Lung	Brain	Cardiac	Kidney
Medicaldecathlon	True	True	True	True	False
kits21	False	False	False	False	True
Cancerimagingarchive-Lung	False	True	False	False	False
Codalab	True	False	False	False	False

However, simply mixing fully annotated dataset and partially annotated dataset can 'spoil' the training, because the partially annotated dataset can be considered falsely annotated and the convergence of the training is more difficult.

We offer an approach is that the ground truth in the training step will be combined with the prediction from the model. Then the loss function is expected not to be biased by the missing annotations. Thus, the training can keep the

performance for organs missing in partially annotated dataset, while improve the performance for organs in both fully and partially annotated dataset, since more training data is included. In the following section we explain the network architecture and analyse the result of experiments.
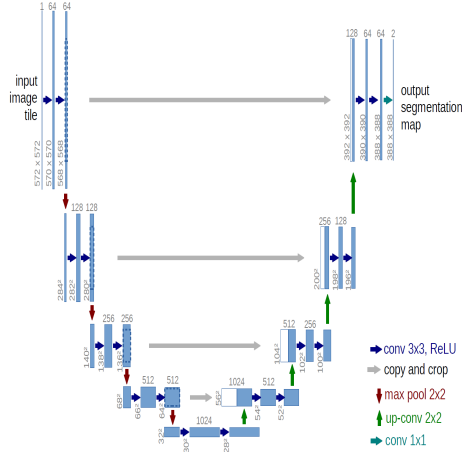
2 Material and methods

Our experiments is basis on U-Net and the experiment data are 3-D CT whole body volume. We need to modify the dataset and the training process in order to implement the algorithm. The dataset we need and the principles of the algorithm are described quickly below

2.1 U-Net

U-Net is named after its U-shaped architecture. Since it was introduced, it has become a popular deep network architecture to deal with semantic segmentation [5]. Due to the practicality of the network and the ability to learn from limited data. The following experiments is basis on U-Net.

Fig. 1. The architecture of U-Net[1]

The U-Net in the experiment was built according to the architecture diagram 1 above, and the final output was one layer which size is 512*512 pixels and include labels: background(0), liver(1), right lung(2), and left lung(3)

2.2 Dataset

The dataset is from public xxxx and consists of 34 patients. Each patient has 400-500 CT full body scan 2D images with same pixel-spacing. These 2D images have zero-paded into the same size 512 * 512. In the following experiment, the different experimental group consisted of the corresponding number of patients. Furthermore, the original dataset is a fully annotated dataset, we have utilized

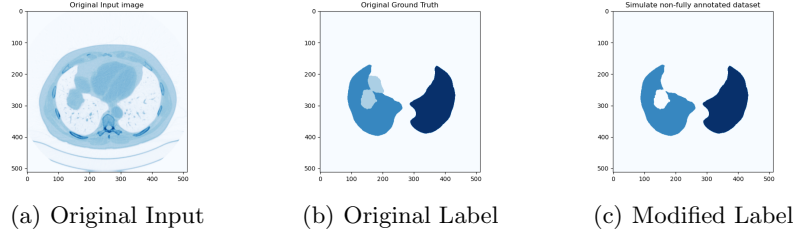masks to simulate the non-fully annotated dataset from fully annotated dataset.

(a) Original Input   (b) Original Label   (c) Modified Label

Fig. 2. Simulation of partially-labelled missing data sets by masking

As the image 2 shows, There are three organs in original Input image 2(a) that need to be segmented. At the beginning annotation data has three types of label, liver, right lung and left lung 2(b), after masking, the image has two types of annotation, the label of the right lung and left lung 2(c). The reason for this is that the annotation belonging to the liver has been changed to background. while for the lungs annotation, it was retained. In this way a dataset without liver annotations was simulated.

2.3 Modified training method

Due to the dataset is composed of fully and partially annotated datasets. The training step should make some change to accommodate such special datasets. From the figure 3 you can see that for the partial annotated data set we have
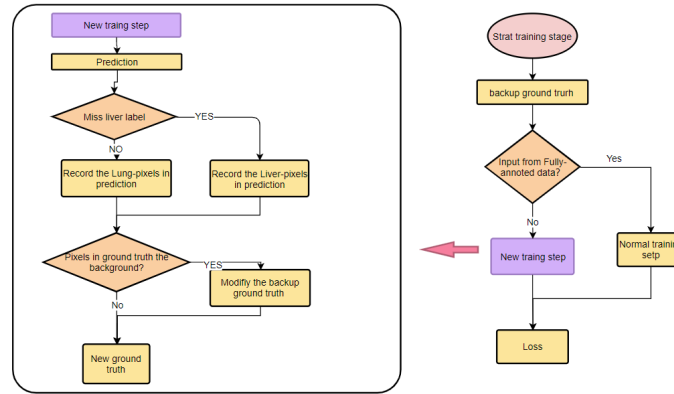

Fig. 3. Process diagram of the training stage

changed the training method.The main idea is that predicted results intervene on

the original label to reduce the impact of partially annotated data. The details of the algorithm for the new training step are as follows

---

Algorithm 1 New Train step

Input: train data $input$, ground truth $target$
Output: modified ground truth $targetCopy$
  $pred \leftarrow model(input)$
  $targetCopy \leftarrow target$
  if $input$ is from not all labeled dataset **then**
    for each $pixel$ in $pred$ **do**
      Record the $coordinates$ of $pixel$ which is classified as Missing label
    end for
    if $targetCopy[coordinates]$ belong to Background **then**
      $targetCopy[coordinates] \leftarrow pred[coordinates]$
    end if
  end if
  Update $loss(pred, targetCopy)$

The above pseudo-code1 explains the new training steps.
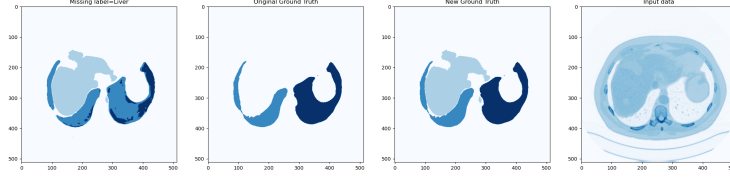

Fig. 4. Results of the new training steps

According to the figure 4, the input to the model is the dataset with the missing liver. Then as seen in the second figure on the left, there are no pixels within the ground truth image that are labelled as livers. After the algorithm 1, in the second figure on the right, which is the new ground truth image, there is a portion of the region where pixels are labelled as liver. However, the markers for the lungs in the predicted image did not migrate to the new ground truth image.

3   Result

In order to verify the validity of the algorithm and to control the experimental variables ,seven experiments groups were designed in this work.

Table 2. Assignment of experimental groups. The abbreviation FL, PAWOL and PAWOLI stands for Full labeled, partially-annotated without Lung and partially-annotated without Liver. The numbers below these columns represent the number of patients

	FL	PAWOL	PAWOLI	Patient Sum	Optimizier Method
Mode 1	24	0	0	24	False
Mode 2	8	8	0	0	False
Mode 3	8	8	8	24	False
Mode 4	8	8	8	24	True
Mode 5	0	8	8	16	False
Mode 6	0	8	8	16	True
Mode 7	16	0	0	16	False

From the above table 2, it can be seen that the differences between the different experimental groups are mainly in the implementation of the optimization algorithm , the amount of patient and the composition of the patient. For each experiment group we utilized the same optimizer and cross-entropy loss. Additonally, the validation sets are always consisting of the same four patients which are not included in train set. Besides that, the ground truth of the validation set data are not masked, regardless of whether the experimental group contains a partially-annotated data set. Purpose of evaluating the model we recorded the following metrics every two validation epochs

- Intersection over Union(IOU) on individual organs,
- Average IOU
- Validation Loss
- Dice score on individual organs
- Precision and Recall.

When all the experiment have been finished ,we according to the average IOU score saved the top 1 model and collected the following result.

wait result

If we focus only on the final performance of the model and in order to reduce deviations, we tested the model again with ten patient datasets. Then we get the following table

4   Discussion

wait Discussion

References

1. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation; 2015.
2. Havaei M, Davy A, Warde-Farley D, et al. Brain tumor segmentation with Deep Neural Networks. Medical Image Analysis. 2017;35:18–31. Available from: https://www.sciencedirect.com/science/article/pii/S1361841516300330.
3. Long J, Shelhamer E, Darrell T. Fully Convolutional Networks for Semantic Segmentation; 2015.
4. Badrinarayanan V, Kendall A, Cipolla R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation; 2016.
5. Baheti B, Innani S, Gajre S, et al. Eff-UNet: A Novel Architecture for Semantic Segmentation in Unstructured Environment. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops; 2020. .
6. Ruttimann UE, Groenhuis RAJ, Webber RL. Restoration of digital multiplane. IEEE Trans Med Imaging. 1984;3(3):141–8.
7. Lehmann T, Oberschelp W, Pelikan E, et al. Bildverarbeitung für die Medizin. Berlin: Springer; 1997.
8. Handels H. Medizinische Bildverarbeitung. Stuttgart: Teubner; 2000.