Learning Segmentation with Unlabeled and Noisy Labeled Examples

1. General Info

Project Title: Learning Segmentation with Unlabeled and Noisy Labeled Examples

Contact Person: Roger D. Soberanis-Mukul

Contact Email: roger.soberanis@tum.de

2. Project Abstract

In this project, we address the problem of limited training data for medical applications. For this, we rely on the mean teacher schemes that have been proposed for semi-supervised learning, and recently adapted for training with noisy (low quality) annotations. We use the organ segmentation problem as the study case for evaluating semi-supervised learning and learning with noisy labels strategies, discussing their performance, limitations, and opportunities for improvement.

3. Background and Motivation

Deep convolutional neural networks are the current state of the art in different image processing tasks. However, these models work under the assumption that high amount of data is available for training. In computer vision tasks, this image annotation process can be easily performed, since the objects of interest are easy to identify for the non-expert eye. However, the annotation of medical images requires of trained annotators, capable of distinguish between the different anatomical structures included in the image. It leads to the well-known problem of data availability for training deep models for medical applications.

To address this problem, different proposals investigate the use of unlabeled data to complement the annotated examples, leading to the definition of semi-supervised learning strategies. The mean teacher model (Fig. 1) is a learning strategy proposed to include the unlabeled data in the learning process by employing the consistency of two CNN (a teacher and a student model)[1,2]. For this, different perturbation can be added to the input image, and then evaluated by each model.

Similarly, it has been of interest to explore the use of non-expert annotations as a complementary source of training data for the models. Non-expert labels can contain errors and inconsistencies derived to the limited experience of the annotator. However, they can also be easy to obtain, compared with the high-quality labels of expert and experienced annotators. The main drawback of these noisy labels is that a naïve use of them can

[1] Wenhui Cui, Yanlin Liu, Yuxing Li, Menghao Guo, Yiming Li, Xiuli Li, Tianle Wang, Xiangzhu Zeng, Chuyang Ye. "Semi-Supervised Brain Lesion Segmentation with an Adapted Mean Teacher Model". IPMI 2019

[2] Zhihao Chen, Lei Zhu, Liang Wan, Song Wang, Wei Feng, and Pheng-Ann Heng. "A Multi-task Mean Teacher for Semi-supervised Shadow Detection" CVPR 2020

lead to a deteriorated model performance. Given the interest of this topic, recent publications[3] have explored the use of noisy annotations under the mean teacher learning model.

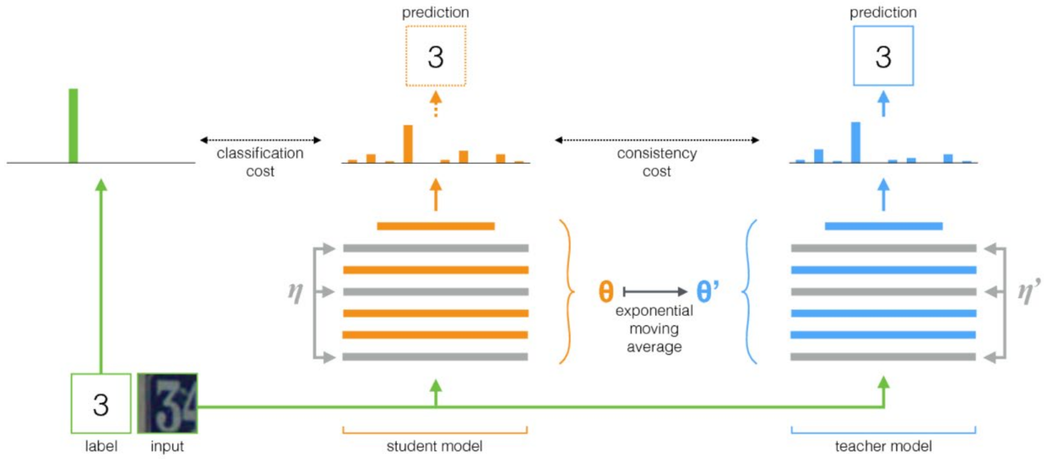

Figure 1 The general mean-teacher learning model (source: https://github.com/CuriousAI/mean-teacher)

In this project, we evaluate and compare semi-supervised learning with the use of noisy labels employing the mean teacher model. We use these strategies in the organ segmentation problem, evaluating the use of noisy labels over the use of unlabeled data, defining the best strategy to use, and considering the possibility of use both additional sources in the model training. For this, we rely on the use of publicly available CT datasets, like the medical image decathlon (http://medicaldecathlon.com), and the NIH Pancreas-CT dataset (https://wiki.cancerimagingarchive.net/display/Public/Pancreas-CT).

4. Technical Prerequisites

- Good skills in Python
- Familiar with PyTorch or related CNN packages (e.g., Tensorflow)
- Previous knowledge on machine learning or/and deep learning.

5. Benefits:

- Working on the open ML medical problem of label availability.
- Learning or practicing PyTorch programming skills.
- Exploring the use of mean teacher strategies for semi-supervised learning.
- Exploring strategies for training with noisy labels.

6. Students' Tasks Description

Students will work in two groups. Each group will implement and evaluate one approach:

Group 1: Semi-supervised learning
Group 2: Learning with Noisy labels

Both models will be inspired in the mean teacher strategy, allowing a fully collaborative project development.

[3] Zhe Xu, Donghuan Lu, Yixin Wang, Jie Luo, Jayender Jagadeesan, Kai Ma, Yefeng Zheng, Xiu Li. "Noisy Labels are Treasure: Mean-Teacher-Assisted Confident Learning for Hepatic Vessel Segmentation". MICCAI 2021.

Group 1: Semi-supervised Learning

- Get familiar with the state of art and the mean teacher model.
- Familiar with the dataset.
- Implement data loaders.
- Implement a baseline fully supervised FCN.
- Generate a noisy/partially labeled dataset.
- Implement the semi-supervised learning model.
- Evaluate the three models.
- Comparison and discussion.

Group 2: Learning with Noisy Labels

- Get familiar with the state of art and the mean teacher model.
- Familiar with the dataset.
- Implement data loaders.
- Implement a baseline fully supervised FCN.
- Generate a noisy/partially labeled dataset.
- Implement the learning with noisy labels model.
- Evaluate the three models.
- Comparison and discussion.

7. Work-packages and Time-plan:

	Description	#Students	From	To
WP1	Get familiar with the state of art and the mean teacher model.	4		
WP2	Familiar with the dataset.	4		
WP3	Implement and data loaders.	4		
WP4	Implement a baseline fully supervised FCN.	4		
WP5	Generate a noisy/partially labeled dataset	4		
M1	Intermediate Presentation II	4		
WP6	Implement the semi-supervised learning model.	2		
WP7	Implement the learning with noisy labels model.	2		
WP8	Evaluate the three models.	4		
WP9	Comparison and discussion.	4		
M2	Final Presentation	4		