Paper Reading No.9

A New Ensemble Learning Framework for 3D Biomedical Image Segmentation

Sheng Lian

July 2019

1 Brief Paper Intro

- ***Paper ref:*** AAAI 2019, http://arxiv.org/abs/1812.03945
- ***Authors:*** See Fig 1.

Hao Zheng*, Yizhe Zhang*, Lin Yang*,
Peixian Liang, Zhuo Zhao, Chaoli Wang, Danny Z. Chen
Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, IN 46556, USA
{hzheng3, yzhang29, lyang5, pliang, zzhao3, cwang11, dchen}@nd.edu

Figure 1: authors' brief intro.

- ***Paper summary:*** This paper propose a new ensemble learning framework for 3D biomedical image segmentation that combines the merits of 2D and 3D models. Authors set up several base-learners and a meta-learner. The latter is used for selecting the 'best' base-learners result.

- ***Reading motivation:*** Another paper that tries to combine the merits of 2D and 3D segmentation models. The setting of base-learners and meta-learner is interesting to me.

2 Backgrounds

3D medical image segmentation face the problem of trade-off between **the field of view** and **utilization of inter-slice information in 3D images**.

- 3D CNNs attempt to fully utilize 3D information, but due to storage and computing power limitations, this kind of model have to choose limited field of view (e.g., 64*64*64).
- 2D CNNs can have a much larger FOV (e.g. 512 * 512) but are not able to fully explore inter-slice information.

3 Methods

3.1 Framework overview

The overall framework of the proposed method is displayed in Fig 2. The model contains two main components:

- a group of 2D and 3D base-learners;
- a meta-learner to combine the results from the base learners.

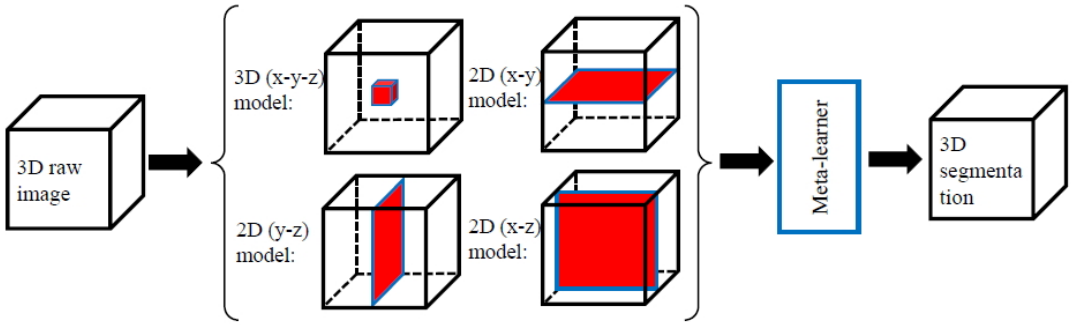

Figure 2: The overall framework of the proposed method.

3.2 2D and 3D base learners

The base learners for 2D and 3D segmentation don't have much to introduce, authors adopted other models that have been validated in the field of medical image analysis. The 2D model used in this paper follows the structure same as [?] do. This model is a FCN-based model with recent advances on deep learning network, such as batch normalization, residual networks and bottleneck design. For 3D model, authors adopted DenseVoxNet [?]. This model is a 3D FCN based model with dense connection and auxiliary side paths for deep supervision.

For boosting diversity, within the groupof 2D base-learners, authors leverage the 3D image data to create multiple 2D views (representations) of the 3D images (e.g., xy, xz, and yz views). So in this paper, authors use the following four base-learners: a 3D DenseVoxNet for utilizing full 3D information; three 2D FCNs for large fields of view in the xy, xz, and yz planes.

3.3 Deep meta-learner

After the base-learner introduced above, the more important issue in this paper is how to obtain the advantages of different base learners. So, authors design a meta learner that is capable of learning robust visual features for jointly utilizing the diverse prediction results (from the base-learners) as well as the raw image information.

So below I will briefly introduce the design of the meta learner.

Formula symbol list:

1) a set of image samples: $X = \{x_1, x_2, \ldots, x_n\}$

2) a set of base-learners: $F = \{f_1, \quad f_2, \ldots, f_m\}$

3) pseudo-label set for each x_i: $PL_i = \{f_1(x_i), f_2(x_i), \ldots, f_m(x_i)\}$

The goal of the meta-learner is to utilize the merits of base learners, while minimizing the risk of over-fitting. So, in this stage, the pseudo-labels, instead of the groundtruth, is used to fom the supervision signal. As introduced above, there are 4 base-learners (maybe more in different situation). So, how to choose the **"best"** base learner in each step for meta-learner to fit? The meta-learner training consists of two phases: (1)random-fit, and (2) nearest-neighbor-fit.

The first training phase aims to train the meta-learner $\mathcal{H}$ to reach a near-optimal solution.

Random-fit The random-fit step randomly chooses a pseudo-label from PL_i and sets it as the current "ground truth" for x_i, which is described in Fig 3. In this setting, the meta-learner will not impose any bias towards any base-learners, and thus reducing the opportunity of over-fitting.

Algorithm 1: Random-fit

Input: $(x_i, PL_i = \{f_1(x_i), f_2(x_i), \ldots, f_m(x_i)\}, S(PL_i))$, $i = 1, 2, \ldots, n$

Output: A trained meta-learner $\mathcal{H}$

initialize a meta-learner $\mathcal{H}$ with random weights;

mini-batch = $\emptyset$

while *stopping condition not met* **do**

  for k = *1 to batch-size* **do**

    $p = rand\text{-}int(1, n)$

    $q = rand\text{-}int(1, m)$

    add training sample $\{(x_p, S(PL_p)), f_q(x_p)\}$ to the mini-batch

  update $\mathcal{H}$ using training samples in the mini-batch with forward and backward propagation

  mini-batch = $\emptyset$

Figure 3: The random-fit algorithm.

Nearest-neighbor-fit (NN-fit) After the former stage, authors believe that the meta-learner already has preliminary judgement and selection ability. For preventing the risk of producing new typeset of errors, authors choose to train the meta-learner to fit the nearest pseudo-label. The details of NN-fit algorithm are presented in Fig 4. Cross-entropy is used for measuring difference between a meta-learner's output and a pseudo-label.

Algorithm 2: Nearest-neighbor-fit (NN-fit)

Input: $(x_i, PL_i = \{f_1(x_i), f_2(x_i), \ldots, f_m(x_i)\}, S(PL_i))$,
$i = 1, 2, \ldots, n$,
meta-learner $\mathcal{H}$ (obtained from random-fit)
Output: A refined meta-learner $\mathcal{H}$
mini-batch = $\emptyset$
while *stopping condition not met* **do**
  for k = *1 to batch-size* **do**
    $p = rand\text{-}int(1, n)$
    $\hat{y} = \mathcal{H}(x_p, S(PL_p))$
    $\hat{q} = \arg\min_{q=1,2,\ldots,m} \mathcal{L}_{mce}(\hat{y}, f_q(x_p))$
    add training sample $\{(x_p, S(PL_p)), f_{\hat{q}}(x_p)\}$ to the mini-batch
  update $\mathcal{H}$ using training samples in the mini-batch with forward and backward propagation
  mini-batch = $\emptyset$

Figure 4: The NN-fit algorithm.

4 Experiments

Authors evaluate the approach using two public datasets: (1) the HVSMR 2016 Challenge dataset and (2) the mouse piriform cortex dataset.

Quantitative analysis on the HVSMR 2016 dataset are displayed below.

Method	Myocardium			Blood pool			Overall score
	Dice	ADB $[mm]$	Hausdorff $[mm]$	Dice	ADB $[mm]$	Hausdorff $[mm]$	
3D U-Net (Çiçek et al. 2016)	0.694 ± 0.076	1.461 ± 0.397	10.221 ± 4.339	0.926 ± 0.016	0.940 ± 0.192	8.628 ± 3.390	-0.419
VoxResNet (Chen et al. 2017)	0.774 ± 0.067	1.026 ± 0.400	6.572 ± 0.013	0.929 ± 0.013	0.981 ± 0.186	9.966 ± 3.021	-0.202
DenseVoxNet (Yu et al. 2017)	0.821 ± 0.041	0.964 ± 0.292	7.294 ± 3.340	0.931 ± 0.011	0.938 ± 0.224	9.533 ± 4.194	-0.161
Wolterink *et al.*(Wolterink et al. 2017)	0.802 ± 0.060	0.957 ± 0.302	6.126 ± 3.565	0.926 ± 0.018	0.885 ± 0.223	7.069 ± 2.857	-0.036
VFN* (Xia et al. 2018)	0.773 ± 0.098	0.877 ± 0.318	4.626 ± 2.319	0.935 ± 0.009	0.770 ± 0.098	5.420 ± 2.152	0.108
Base learner 2D (xy)	0.789 ± 0.076	0.852 ± 0.265	4.231 ± 1.908	0.930 ± 0.016	0.794 ± 0.153	5.295 ± 1.671	0.13
Base learner 2D (xz)	0.736 ± 0.093	1.000 ± 0.260	5.417 ± 1.604	0.924 ± 0.015	0.932 ± 0.113	7.951 ± 2.820	-0.098
Base learner 2D (yz)	0.756 ± 0.082	0.870 ± 0.181	4.169 ± 0.632	0.928 ± 0.012	0.812 ± 0.111	$\mathbf{5.229 \pm 1.721}$	0.108
Base learner 3D	0.809 ± 0.069	0.785 ± 0.235	4.121 ± 1.870	0.937 ± 0.008	0.799 ± 0.145	6.285 ± 3.108	0.13
Average ensemble	0.805 ± 0.073	0.708 ± 0.184	$\mathbf{3.211 \pm 0.923}$	0.936 ± 0.011	0.752 ± 0.119	5.960 ± 2.526	0.2
Our meta-learner (only training data)	0.823 ± 0.060	0.685 ± 0.164	3.224 ± 1.096	0.935 ± 0.010	0.763 ± 0.120	5.804 ± 2.670	0.21
Our meta-learner (transductive)	$\mathbf{0.833 \pm 0.054}$	$\mathbf{0.681 \pm 0.178}$	3.285 ± 1.370	$\mathbf{0.939 \pm 0.008}$	$\mathbf{0.733 \pm 0.143}$	5.670 ± 2.808	**0.234**

Figure 5: Quantitative analysis on the HVSMR 2016 dataset.

Here, the **'transductive'** means that we allow meta-learner to utilize test data. In this setting, authors use the full training data to train base learners, and use the training and testing data to train meta-learner. The results show the ability to refine the model after seeing the raw test data (no annotation for test data).

These 'transductive' issues shows the ability of the proposed method in semi-supervised scenario. The semi-supervised part and detailed ablation study can be referred in the paper.