UNIVERSITY OF COPENHAGEN

MIDL Lübeck 2021

Deep ensemble model for segmenting microscopy images in the presence of limited labeled data

Jan Mikolaj Kaminski[1,2], Ilary Allodi[1], Roser Montañana-Rosell[1], Raghavendra Selvan[1,2] , Ole Kiehn[1]
[1]Department of Neuroscience, [2]Department of Computer Science, University of Copenhagen

Introduction

- Supervised Learning assumption = data is ***i.i.d***
- Small dataset are mostly not ***i.i.d***
- Current ensemble models focus more on differences in the ***models*** than ***data***
- Class imbalance/various annotators → multiple data ***modes***

Objective:

Creating deep ensemble model which takes advantage of different modes in the data therefore being able to perform efficient 2D segmentation.

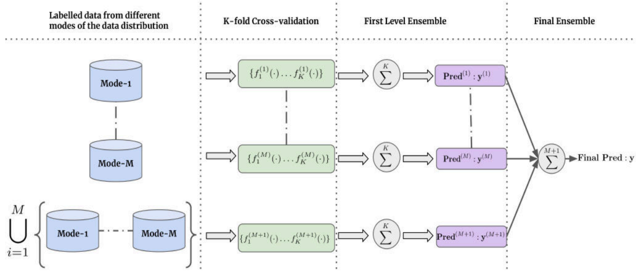
METHOD

Training data partitioning:

- **Partitioning of data into M modes performed based on expert knowledge or non-overlapping different raters**
- **Complete dataset as additional mode**

Deep ensemble model

- **k-fold cross-validation for each data mode**
- **(M+1)*K identically structured segmentation models**
- **First level ensemble is a combined segmentation for each mode**
- **Final ensemble is combined from all mode-level predictions with (M+1)*K discrete prediction levels**

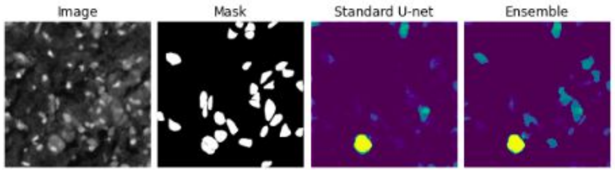


$$\mathbf{y}^{(m)} = \frac{1}{K}\sum_{k=1}^{K}[f_k^{(m)}(\mathbf{x}) > 0.5] \qquad \mathbf{y} = \frac{1}{(M+1)}\sum_{m=1}^{M+1}\mathbf{y}^{(m)}$$

Pipeline for the proposed Deep Ensemble model. The training dataset is split into M partitions; an additional combined dataset is obtained comprising all M partitions (last row). Each of these M + 1 partitions are trained in a K-fold cross validation setting. Predictions from each cross-validation are aggregated to obtain the first level ensemble prediction. Finally, the M + 1 predictions are combined to output the final model prediction.

Data & Experiments

- High-resolution, greyscale mice spinal cord microscopy images
  - acquired at 2 different timepoints; 1st- (P1) and 28th-(P28) postnatal day
  - images divided into patches of size 512x512;
  - 52 training patches with sparse labels; 2% of dataset
  - 4 test images fully annotated
- Standard U-net and Seeded Region Growing (SRG) for baseline
- 2 data modes and 15 separate U-nets trained in the ensemble
- DICE loss and F1 score as metrics



Example of a test image with the corresponding mask, standard U-net segmentation result and ensemble segmentation result.

Method	F1 score
Ensemble model	$\mathbf{0.64 \pm 0.10}$
Standard U-net	0.53 ± 0.02
SRG	0.59 ± 0.18

Average F1 score obtain over test data for the ensemble model, Standard U-net and SRG with threshold of t=0.06.

Conclusions

- The proposed ensemble improves the F1 score by ≈0.1 compared to standard U-net
- Best F1 score obtained with threshold of 0.06 implies that all models are usefull
- We hypothesise that proposed model is most useful under the non i.i.d data regime; the diversity is achieved by learning on different modes of data distribution

References
1. Yaser S Abu-Mostafa, Malik Magdon-Ismail, and Hsuan-Tien Lin. Learning from data, volume 4. AMLBook New York, NY, USA:, 2012.
2. Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. In Proceedings of Neural Information Processing Systems, 2017.
3. Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention, pages 234–241. Springer, 2015.