Group B Progress Report

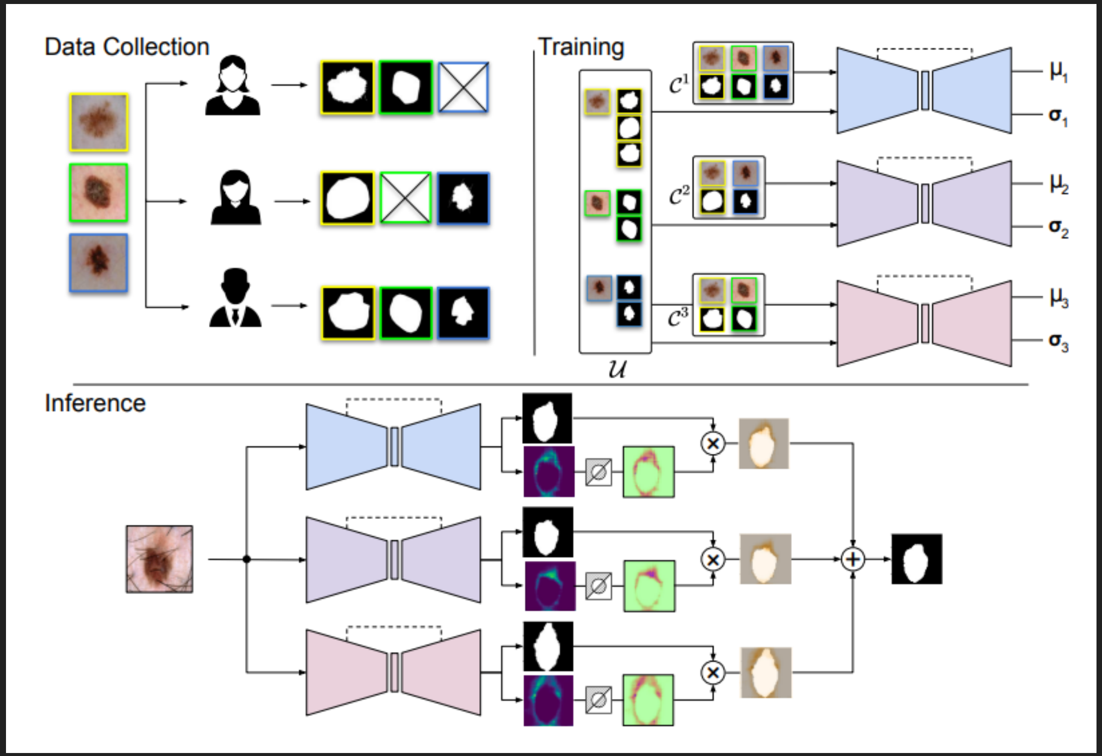
Part 1

Content

- Ensemble Tradeoff
- Model analysis

Ensemble on medical datasets

- In statistics and machine learning, ensemble methods use multiple learning algorithms to obtain better predictive performance than could be obtained from any of the constituent learning algorithms alone.



Why Ensemble

- Specifically, Medical image segmentation annotations suffer from inter- and intra-observer variations even among experts due to intrinsic differences in human annotators and ambiguous boundaries.
- Therefore, it will we hard to train a single model to cover all cases.
- While it is possible to train many of them to learn different tendencies.

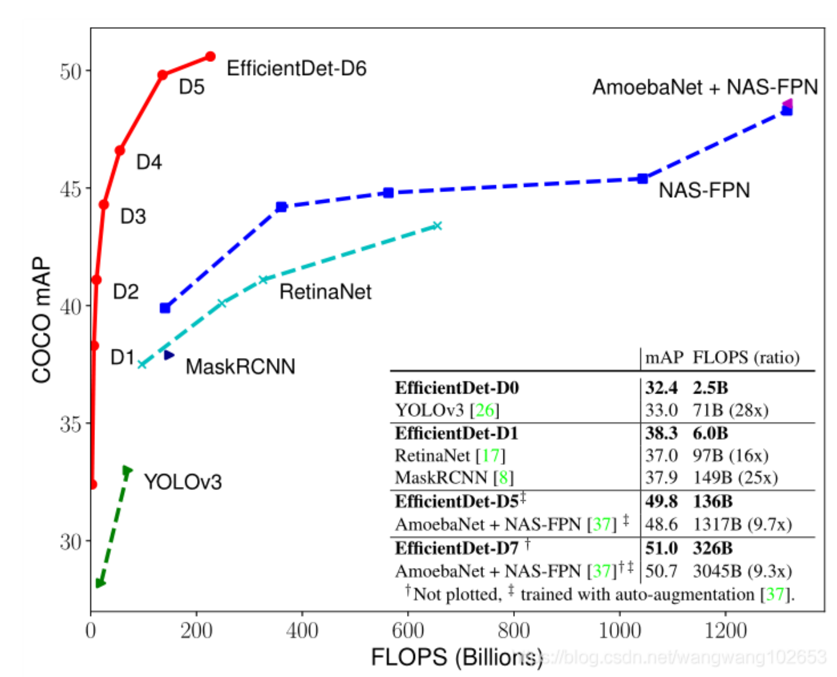
- https://openaccess.thecvf.com/content/CVPR2021W/ISIC/papers/Mirikharaji_D-LEMA_Deep_Learning_Ensembles_From_Multiple_Annotations_-_Application_to_CVPRW_2021_paper.pdf

Ensemble tradeoff

- Therefore, we introduced combining outputs of multiple models (currently, we have tried at most 20 models).
- While it is not always merits that are brought by adding new models.
- Sometimes, it evokes curse of democracy which slow down the prediction while worsen the general output.
  - Such problems are partly caused by training and model capacity.

Model analysis

- The general structure of the model is FPN
- Currently, we have these backbone models tried:
  - Efficientnet_b4
  - Efficientnet_b5
  - Efficientnet_b6
  - Efficientnet_b7
  - UneXt50
  - UneXt101 (Still training)



Architecture	Supervision	#Parameters	FLOPS	Top-1 Acc.	Top-5 Acc.
ResNet-18	semi-supervised	14M	2B	72.8	91.5
ResNet-50	semi-supervised	25M	4B	79.3	94.9
ResNeXt-50 32x4d	semi-supervised	25M	4B	80.3	95.4
ResNeXt-101 32x4d	semi-supervised	42M	8B	81.0	95.7
ResNeXt-101 32x8d	semi-supervised	88M	16B	81.7	96.1
ResNeXt-101 32x16d	semi-supervised	193M	36B	81.9	96.2
ResNet-18	semi-weakly supervised	14M	2B	**73.4**	91.9
ResNet-50	semi-weakly supervised	25M	4B	**81.2**	96.0
ResNeXt-50 32x4d	semi-weakly supervised	25M	4B	**82.2**	96.3
ResNeXt-101 32x4d	semi-weakly supervised	42M	8B	**83.4**	96.8
ResNeXt-101 32x8d	semi-weakly supervised	88M	16B	**84.3**	97.2
ResNeXt-101 32x16d	semi-weakly supervised	193M	36B	**84.8**	97.4

Model Analysis

- By comparison and experiment, we have got some rule-of-thumb to discern the performance of model individually without submit single model prediction.
  - 1. Best Dice_th on most folds should exceed 0.82
  - 2. Parameter size should be higher than UneXt50 (.pth size >= 100m approximately)

Submission and Description	Status	Public Score	Use for Final Score
[Inference]-HuBMAP fast.ai starter (EfficientNet) Version 15 (version 15/15) 6 hours ago by Rathgrith Notebook [Inference]-HuBMAP fast.ai starter (EfficientNet) | Version 15	Succeeded Efficientnet b4 added	0.69	
[Inference]-HuBMAP fast.ai starter (EfficientNet) Version 13 (version 13/15) 11 hours ago by Rathgrith Notebook [Inference]-HuBMAP fast.ai starter (EfficientNet) | Version 13	Succeeded UneXt50 dropped	0.69	
[Inference]-HuBMAP fast.ai starter (EfficientNet) 16model!!!!!!! (version 11/15) 12 hours ago by Rathgrith Notebook [Inference]-HuBMAP fast.ai starter (EfficientNet) | Version 11	Succeeded	0.70	

Future plans

- We are trying to replace all models we have by stronger models (SwinUnet) that are trained on adjusted dataset.
- This process would likely to be very painstaking, but still promising.

Discussion